Norbert Fürstenau   *Editor*

Virtual and Remote Control Tower

Research, Design, Development, Validation, and Implementation

Second Edition

Co-edited by   Anne Papenfuss and
Jörn Jakobi

DLR

Springer

Research Topics in Aerospace

Series Editor

Rolf Henke, Mitglied des Vorstands, DLR, Cologne, Nordrhein-Westfalen, Germany

DLR is Germany's national centre for space and aeronautics research. Its extensive research and development work in aeronautics, space, transportation and energy is integrated into national and international cooperative ventures. Within this series important findings in the different technical disciplines in the field of space and aeronautics, as well as interdisciplinary projects and more general topics are addressed. This demonstrates DLR's outstanding research competences and capabilities to the worldwide scientific community and supports and motivates international research activities in exploring Earth and the universe, while also focusing on environmentally-friendly technologies and promoting mobility, communication and security.

More information about this series at https://link.springer.com/bookseries/8625

Norbert Fürstenau
Editor

Virtual and Remote Control Tower

Research, Design, Development, Validation, and Implementation

Second Edition

Co-edited by Anne Papenfuss and Jörn Jakobi

🐴 Springer

DLR Deutsches Zentrum
für Luft- und Raumfahrt
German Aerospace Center

Editor
Norbert Fürstenau
Institute of Flight Guidance
German Aerospace Center
Braunschweig, Germany

Foreword: On the Origins of the Virtual Tower

It's a pleasure to write a personal account regarding the origins of the virtual air traffic control tower as reflected in our work at the NASA Ames Research Center. This type of air traffic display is now sometimes called the remote tower, but I think there is a significant difference between the two. The virtual tower is actually a much more radical proposal and is only in the last few years becoming clearly possible at a reasonable cost. But, as I discuss later, whether it provides any additional benefit beyond the remote tower depends strongly on the specific content and application.

The Ames work on the virtual tower can be traced to a meeting I had with my boss, Tom Wempe, to whom I first reported in the late 1970s. I was a National Research Council (NRC) postdoc working for him studying pilot's eye movements looking at a newly proposed Cockpit Display of Traffic Information. This display was an electronic moving map that was intended for use in commercial aircraft cockpits to aid air traffic avoidance and to help pilots accept automatic avoidance commands. When Tom not so subtly hinted that "It would be good for me to known around here as a displays person rather than an eye movement person," I got the point. This was the first time I had ever been explicitly directed to work on something specific. Even in grad school at McGill University, I never got a specific direction. Part of the education there was to be able to figure out for yourself what was important to work on.

So when Tom got even more specific and pointed out that "We were having trouble coming up with a good way to depict vertical separation on the 2D plan-view map" and that he would like me to work on this problem, I really began to worry. I didn't want to work on a display! So in some desperation, I suggested, "Well, why don't we make it look like a view out the window?" At the time I drew on his blackboard a sketch of what a pilot might see out the forward window. And Tom said, "OK, why don't you work on that." But I had absolutely no idea what I would do or how I would do it.

I proposed that I should try to find some interested colleagues for this project in Prof. Larry Stark's lab at Berkeley and the next week at his lab meeting suggested we find a student to work on the project. He had a new student named Michael McGreevy who was interested in the Bioelectronics Option for a graduate engineering program.

He turned out to be perfect. He was an engineer with a background in art who was also interested in computer graphics, which he was then studying in a class by Brian Barsky. We began a multiyear collaboration in which we worked on the design, implementation, and testing of a perspective format for a Cockpit Display of Traffic Information (CDTI). What interested me particularly were the perceptual phenomena associated with interpreting an accurate geometric projection of the relative position and direction of targets that might be presented on a pilot's display of surrounding aircraft. Mike was beginning to program the Evans and Sutherland Picture System 2 and we initiated a design collaboration to investigate the geometric and symbolic elements that would be needed to make a perspective CDTI suitable for a cockpit. The goal was to make a traffic display useable at a glance. Before our project, all CDTIs were plan-view. The perspective CDTI was eventually called VERT. It ultimately was evaluated with respect to a convention plan-view CDTI called INTRUD (Ellis & McGreevy, 1987).

From the design and testing of prototypes, we learned many things. For example, a "God's eye" view from behind and slightly offset was better than a forward, egocentric view as if directly out the cockpit. But most interesting was that we found from systematic testing of pilot's direction judgments an apparent perceptual distortion we called the "telephoto" bias. It was as if when spatially interpreting the display, the users were seeing through a telephoto lens and that their visual attention would, therefore, not be correctly directed out the window for visual contact with traffic. It turned out that theoretical models developed from working with Mike (McGreevy & Ellis, 1986), and later Arthur Grunwald (Grunwald, Ellis & Smith, 1988), and still later Gregory Tharp (Tharp & Ellis, 1990) provided several alternative but related techniques we could use to distort the display for better spatial interpretability.

It should be noted that considerable effort went into the initial design of the three-dimensional symbolic content of the perspective CDTI. In this design process, we learned that many of the difficulties of spatially interpreting perspective displays can be removed by the appropriate design of its geometry and symbology. Consequently, it became apparent that simple performance comparisons of perspective versus plan-view formats could be misleading. Symbology can be introduced to remove interpretive difficulties with the perspective format. For example, segmented vertical reference lines can remove spatial ambiguities due to the geometric projection.

Later in the early 1980s after being hired as a Civil Servant at Ames, Mike McGreevy became interested in jumping into the data space of the maneuvering aircraft as seen on at CDTI, as if it were a virtual environment. He began a series of projects to develop a head-mounted display for visualization of a variety of data spaces and environments. This was the birth of "VR" at NASA in 1985. The very first real-world digital content viewed in this was a complex pattern of interacting air traffic called the "Atlanta Incident." It was a series of worrisome close encounters of aircraft generally within the Atlanta TRACON. Despite the very poor visual and dynamic quality of the early NASA HMDs, which was not reflected in the contemporary accounts of the work in the press, the reincarnation of Ivan Sutherland's "Ultimate Display" was clearly demonstrated with these air traffic data.

I was generally not directly involved with the development of the virtual environment displays at Ames until the early 1990s when I began to work on the relationship of objective measures of system performance to virtual environment system usability. We studied, for example, full system latency and counter measures for it such as predictive filtering. My principal collaborator for this work was Bernard "Dov" Adelstein. The visual environments we studied at the time for our scientifically motivated design work were generally not particularly visually interesting so it became strategically and programmatically important to show realistic possible uses of the display format for applications that would interest NASA.

Since we were receiving support from both space and aeronautics programs at Headquarters, I felt we needed two separate demonstration environments. The "space" one was a fly-around of the Shuttle Orbiter with the task of identifying damaged tiles. The "aeronautics" one was a visualization of simulated aircraft landing at SFO. Initially, we used synthesized trajectories but later replaced them with recordings of live approach and landing data from DFW which was provided by Ronald Reisman. I called our display a virtual tower in that the head-mounted display user would appear to be immersed in the traffic pattern. I was surprised by how much attention this second demo attracted. One possible reason was the high visual and very high dynamic fidelity we achieved for the 1990s., attracting attention outside our agency. This time, however, the popular representations of our system's performance were more accurate.

However, I ultimately became concerned that advocacy for a virtual tower would involve way too much technological push so rather than pursuing a line of system development, I sought to back up and investigate the visual aspects of tower operation. I wanted to better understand the visual requirements for tower operations beyond the visual detection, recognition, and identification functions that seemed to circumscribe the visual concerns of the FAA when it came to visual tower operation. A better understanding of the visual features used by Tower controllers would help establish performance requirements for either virtual or remote towers. Two of our papers as well as six chapters in this volume (2, 3, 16, 9, 10 and 18, including the quasi-operational shadow-mode validation) address this concern.

The virtual tower history sketched above describes work leading to a virtual tower that could be essentially worn on a controller's head as a totally immersing virtual environment. Such a format isolates its users from their immediate physical environment and probably only makes operational sense when compactness, low power consumption, and portability are important. In fact, this head-worn display format might be appropriate for use by Forward Air Controllers on a battlefield. These soldiers have a job somewhat similar to an air traffic controller, though their goals may be different. In fact, a version of such an application called the Forward Air Controller Training Simulator (FACSIM) was developed at TNO, the Hague.

But now, as can be seen in the following volume, the time for a virtual, or more properly labeled, remote tower has come. The sensors, communications links,

rendering software, and aircraft electronics needed for the implementation of a practical system all seem to be in place. As will be evident from the following chapters much of the system integration work needed to complete such systems is afoot.

Moffett Field, CA, USA   Stephen R. Ellis

References

Ellis, S., McGreevy, M. W., and Hitchcock, Ro. (1987). Perspective traffic display format and airline pilot traffic avoidance. *Human Factors*, *29*, 371–382.

Grunwald, A., Ellis, S. R., and Smith, S. R. (1988). Spatial orientation in pictorial displays. *IEEE Trans. on Systems Man and Cybernetics* , *18*, 425–436.

McGreevy, M. W. and Ellis, S. R. (1986). The effect of perspective geometry on judged direction in spatial information instruments. *Human Factors*, *28*, 439–456.

McGreevy, M. W. and Ellis, S. R. (1991). *Format and basic geometry of a perspective display of air traffic for the cockpit. NASA TM 86680.* Ames Research Center, Moffett Field, CA.

Tharp, G. K. and Ellis, S. R. (1990) *The effects of training on errors of perceived direction in perspective displays NASA TM 102792.* NASA Ames Research Center, Moffett Field CA.

Preface to the Second Edition

Nineteen years after the proposal of the Virtual Tower project within DLR's Visionary Projects competition of 2002 (Fürstenau N., 2004), sixteen years after the grant of the first Virtual Tower patent ((Fürstenau, et al., 2008), application 2005 (German patent)), and six years after the first edition of the present volume, the worldwide implementation of Remote Tower Systems has gained speed. After licensing of the DLR-patent to industry, on December 4, 2018, the German Air Navigation Service Provider (ANSP) DFS had started its first Remote Tower Operation of Saarbrücken international airport with a Remote Tower Control Center (RTC) located 450 km to the east at Leipzig airport, to be followed by airports of Dresden and Erfurt. This event followed two years after the worldwide first start of RTO control by the Swedish ANSP LFV, with the RTC located at Sundsvall airport for air traffic management of Örnsköldsvik (2016). In Norway, the ANSP AVINOR opened a RTC on October 20, 2020, in Bodö north of the Arctic circle for remote control of two airports, and the plan is to extend it to a total of 15 by the end of 2022.

These RTC installations represent many others all over the world such as Stockholm, Budapest, London City, and they provided the motivation and sufficient material for an update to the first edition of the RTO book. In nine new chapters, it covers a number of additional topics which gained increasing attention during recent years: from Multiple Remote Tower (MRT) validations over workload measurement and analysis under MRT operation to advanced technologies and low-cost remote systems for non-certified air traffic services (ATS) like AFIS or UNICOM.

In the thirteen chapters of the 2016 Virtual Tower book edition, the focus was on basic preconditions for prototype development like visual features used by Air Traffic Controller Officers (ATCOs), technical aspects, and RTO design with integration of a high-resolution video panorama with broadband fiber optic data transmission as enablers, Augmented Vision functions based on real-time image processing, e.g. for automatic object detection and tracking, human factors questions for workplace design and workload issues, and field tests of RTO prototypes in Germany (DFS) and Sweden (LFV). Authors from NLR, LFV, and Saab described the first passive shadow-mode field testing of advanced video functions such as overlaid approach

radar information moving with tracked aircraft, taking place at the RTC test facility of Malmö Sturup airport observing Ängelholm 100 km to the north.

One major motivation of the Remote Tower Operation (RTO) research of course is cost effectiveness. Particularly with the transition from single to Multiple Airport Operation (MRTO), this is a key driving factor. It enables a flexible RTC work environment that would allow for a variable number of staff and flexible allocation of airports to a multiple remote tower module (MRTM) for the centralized control of several airports under large variations of traffic density. The corresponding MRTO research and development work had already started in the prototype phase described in the 1st edition of the book (Part II of the present one). It was focused, however, on Human-in-the-Loop (HITL) simulation experiments for workplace design and workload aspects, with field testing restricted to single airport RTO.

With nine new chapters and two new appendices, the content of the present 2nd edition has nearly doubled. Due to the large number of new chapters, we decided to re-structure the content so that most of the new chapters concentrate in Parts III and IV. Besides Introduction and the basic preconditions for RTO development like required visual information by ATCOs, Part I now includes two chapters with RTO aspects of historical interest: an extended review of the corresponding activities in the United States by V. Nene and a contribution by S. Inoue et al. (ENRI, Tokyo) on Aerodrome Flight Information System technology (AFIS) with remote visual surveillance of small uncontrolled airports in Japan as a kind of RTO predecessor. These introductory chapters are followed in Part II by the technical research and development chapters including augmented vision experiments using image processing for pan-tilt-zoom camera object tracking, and initial prototype field testing. It includes another new contribution by Inoue et al. (ENRI/Tokyo) on the integration of cooperative multilateration surveillance data with visual object identification and tracking via image processing.

Part III puts the focus on the important HITL simulation experiments, starting with the extended Remote Tower Laboratory environment at DLR-Braunschweig where several HITL experiments with domain experts provided new data on video frame-rate effects, object tracking issues, situation awareness, and workload. Quantification of workload effects under different work conditions represents one aspect of increasing interest (conventional tower vs. RTO vs. MRTO). Besides being addressed in different chapters, basic additional information is provided in two new Appendices (C and D).

Part IV addresses usability experiments with field testing of advanced technologies such as fusion of the video panorama with thermal camera information. Moreover, following the description of a controller friendly MRTO assistance tool by R. Leitner and A. Oehme, a large-scale validation experiment is described by Li et al. (Cranfield University) together with the Irish Aviation Authority (IAA), concerning the certification process for the new Dublin RTC MRT Center with remote control of Shannon and Cork international airports. The AFIS topic from Part I is taken up again describing a detailed validation study of an advanced although low-cost visual surveillance system for small low-traffic airports, based on a PTZ-camera remotely controlled by a VR-headset with head tracking.

Following the shadow-mode field tests of the initial DFS prototype at Erfurt airport in 2012 within a DFS–DLR cooperation, several European cooperations with EC-co-funding by the Single European Sky Air Traffic Research (SESAR) initiative were performed for developing and validating Remote Tower Operation. Within the European Organization for Civil Aviation Equipment (EUROCAE), a Remote Tower working group had been set up (WG100 "Remote and Virtual Tower") that published a document on Minimum Aviation System Specifications (EUROCAE, 2016). In parallel to an increasing number of industrial (M)RTO development activities, the breakthrough of this interdisciplinary research and development effort proved successful after more than 15 years, leading to a paradigm change in airport traffic control that will eventually lose its 100 years old symbol, the airport control tower.

Two of the editors (N.F., J.J.) were happy to be awarded the Manfred-Fuchs Innovation prize in 2019, for the "Successfully Realized Innovation 'Remote Tower'". This success would not have been possible without the fundamental contributions of the initial "Virtual Tower" technical team: Markus Schmidt as a chief engineer who realized (to our best knowledge) the worldwide first experimental system for field testing (The chapters "Remote Tower Experimental System with Augmented Vision Videopanorama" and "Remote Tower Prototype System and Automation Perspectives"), Bernd Werther as an analyst of the tower work procedures who designed and realized within his Ph.D. dissertation a colored Petri-net computer model of the controllers task and work network, in close cooperation with DFS (Werther, Cognitive modeling with Coloured Petri Nets for the analysis of human behaviour, 2005)(Werther & Uhlmann, Ansatz zur modellbasierten Entwicklung eines Lotsenarbeitsplatzes, 2005) (Werther, Airport control model for simulation and analysis of airport control processes, 2007), and Michael Rudolph as a software developer, who designed and wrote the basic RTO-software including augmented vision features, for field testing and human-in-the-loop simulation (The chapters "Remote Tower Experimental System with Augmented Vision Videopanorama" and "Remote Tower Prototype System and Automation Perspectives" and references therein).

As a editor of the 2016 volume, I am indebted to my co-editors of the present 2nd edition, Dr. Anne Papenfuss and Jörn Jakobi (Chairman, EUROCAE WG100) who were involved already in part of the research described in the 1st edition and in most of the research work described in the new chapters. They spent much of their time in motivating chapter (co-) authors for contributing to this volume and in reviewing the manuscripts. Together we express our sincere thanks to Satoru Inoue, Mark Brown, and Yasuyuki Kakubari (ENRI/Tokyo); Wen-Chin Li and Graham Braithwaite (Cranfield University, Bedfordshire, UK); and Peter Kearney (IAA, Dublin) for their contributions and their patience during the recent two (or so) years.

Special thanks of one of the editors (N.F.) are to Thea Radüntz (Unit Mental Health and Cognitive Capacity, Federal Inst. for Occupational Safety and Health (BAUA), Berlin) who together with Thorsten Mühlhausen (ATC Simulator Division, DLR) organized the first HitL-simulation experiment for validating the new EEG-based "Dual Frequency Headmap" method (DFHM, (Radüntz, 2017), see Appendix C) through quantifying workload under realistic ATC conditions. It provided basic data for the derivation of the logistic and power law WL-models (see Appendix D)

for prediction and statistical estimation of MRTO WL-parameters in the chapter "Model Based Analysis of Subjective Mental Workload during Multiple Remote Tower Human-In-The-Loop Simulations". Last but not least, I am indebted to Jürgen Rataj (Head of Controller Assistance Division) for his support of the DLR-BAUA cooperation and for enabling this book project through a consulting contract with one of the editors (N.F.) after his retirement in 2016.

Braunschweig, Germany   Norbert Fürstenau
November 2021

References

EUROCAE. (2016). *Minimum Aviation System Performance Specification for Remote Tower Optical Systems*. Internal Report, Malakoff, France. Retrieved 2018, from www.eurocae.net.

Fürstenau, N. (2004). *Virtual Tower. Wettbewerb der Visionen 2001–2004* (pp. 16-19). Köln: DLR.

Fürstenau, N., & Radüntz, T. (2021). Power law model for Subjective Mental Workload and validation through air-traffic control human-in-the-loop simulation. *Cognition, Technology, and Work*. Retrieved 05 30, 2021, from https://doi.org/10.1007/s101011-021-00681-0.

Fürstenau, N., Rudolph, M., Schmidt, M., Werther, B., Hetzheim, H., Halle, W., & Tuchscheerer, W. (2008, 12). Flugverkehr-Leiteinrichtung (Virtueller Tower). *European Patent EP1791364*.

Radüntz, T. (2017). Dual Frequency Head Maps: A New Method for Indexing Mental Workload Continuously during Execution of Cognitive Tasks. *Frontiers in Physiology, 8*, 1–15. https://doi.org/10.3389/fphys.2017.01019.

Werther, B. (2007). Airport control model for simulation and analysis of airport control processes. *Proc. HCII2007: Lecture Notes Computer Science. vol. 4558.* Beijing: Springer.

Werther, B., & Uhlmann, H. (2005). Ansatz zur modellbasierten Entwicklung eines Lotsenarbeitsplatzes. *VDI-Fortschrittberichte: Zustandserkennung und Systemgestaltung*, *22*, 291–294. Berlin: VDI.

Preface to the First Edition

The paradigmatic symbol in Air Traffic Control (ATC), essentially unchanged since the beginning of commercial air traffic early last century, is the characteristic control tower with its large tilted windows, situated at an exposed location, and rising high above the airport. Besides the impressive 360° panoramic far view out-of-windows it provides the tower controller an aura of competence and power. It actually hides the fact that tower controllers as employees of the Air Navigation Service Provider (ANSP) are members of a larger team of collaborating colleagues at different locations, including the apron, approach and sector controllers, not all of them enjoying the exciting view out of the tower windows (for more details see Chapter 1, Introduction, section 1). Only the apron controllers supervising the traffic on the movement area in front of the gates, mostly as employees of the airport operator, enjoy a similar panorama, although usually from a lower tower. The topic of this book, Virtual and Remote Control Tower, questions the necessity of the established direct out-of-windows view for aerodrome traffic control. It describes research towards an alternative work environment for tower and apron controllers, the Virtual Control Tower. It is probably no exaggeration to assert that this book is about a paradigm change in air traffic control, where paradigm in this context means a generally accepted way of thinking and acting in an established field of technology.

As explained already by Steve Ellis in the Foreword to this volume, Virtual and Remote Tower refers to the idea of replacing the traditional aerodrome traffic control tower by a sensor based control center which eliminates the need for a physical tower building. For small low-traffic airports, the main topic of this book, the out-of-windows view will be reconstructed by a high-resolution video-panorama which may be located anywhere on the airport or even hundreds of kilometers away at a different location. This concept quite naturally leads to a new type of aerodrome control center which allows for remote control of several airports from a single distant location. It is understandable that many tower controllers are not really happy with this revolutionary idea, viewing videos instead of enjoying the reality behind the windows. The detailed research towards the Virtual Tower presented in the following chapters will show that their scepticism is partly justified, and it is the responsibility of us researchers to take their critique serious and understand their requirements in

order to maintain and exceed the safety and performance level with the new system which the traditional one has achieved within nearly a hundred years of technical evolution.

After surfacing of the Virtual Tower idea several requirements for "Future ATM Concepts for the Provision of Aerodrome Control Service" were formulated by the International Federation of Air Traffic Controllers Associations (IFATCA), such as:

The controller shall be provided with at least the same level of surveillance as currently provided by visual observation

Controllers shall be involved in the development of aerodrome control service concepts

While the first condition relates to official regulations of ICAO (International Civil Aviation Organisation) concerning visual traffic surveillance on aerodromes, the second one addresses the methods for design, research and development, validation and implementation of the proposed new human–machine systems for aerodrome traffic controllers. It appears self evident that the introduction of a revolutionary new work environment in the safety-critical field of aeronautics which attempts to replace an established operationally optimized and validated existing one, requires intensive cooperation between developers and domain experts. In Germany most of them are employees of the Air Navigation Service Provider DFS (Deutsche Flugsicherung), cooperation partner in the recent Remote Tower projects.

While the development of any new human-machine system by definition is an interdisciplinary undertaking, nowadays involving at least experts from engineering, computer science/informatics, and engineering psychology/cognitive engineering, this book is about an especially challenging case. On the one hand a revolutionary concept based on latest technologies is suggested which promises a significant increase of efficiency and decrease of cost. On the other hand it attempts to replace a well established system with a hundred years of operational experience which has to satisfy two often competing goals: safety and efficiency.

One of the problems with this kind of interdisciplinary research and development is that the field of engineering psychology and cognitive ergonomics addressing the human operator side of the system has a much weaker scientific foundation concerning established and usable formal theories as compared to the technical-engineering side. The engineers and scientists on the technical side can usually rely on a well accepted and established basis of theoretical, mathematically founded knowledge (e.g. applied optics for the realisation of a high resolution videopanorama) and powerful software tools for simulating engineering problems and prediction of the technical system performance. The human factors experts/psychologists on the other side usually have to work with data derived from a huge amount of statistically quantified experimental results, backed up by only a relatively small number of generally accepted formal theories of human perception and behavior (e.g. Weber-Fechner Law/Steven's Function, and the Signal Detection Theory; see Appendices A, B). Moreover there are only very few if any usable quantitative approaches and simulation tools for addressing concepts like operators "mental model", "situational awareness" or "human performance" and decision making in a way which would allow for the numerical prediction of e.g. decision errors. System performance under

operationally relevant conditions is typically derived from human-in-the-loop simulations, with participant's responses derived from subjective questionaires (for cost reasons often only students instead of well trained domain experts, and not seldom with questionable statistical relevance). This situation makes it difficult to obtain reliable quantitative statements about the operators performance in the new environment. For specific questions regarding requirements and performance, experiments under more laboratory kind of conditions at the cost of reduced operational relevance can be designed which have a better chance to be comparable with theoretical predictions. Within the framework of the Remote Tower work system research the present truly interdisciplinary book contains chapters addressing, on different levels, both the technical system engineering, the human operator and (cognitive) ergonomics, and the human-system interface aspects.

At this point we would like to acknowledge several contributions and preconditions without which much of the research work described in the following chapters probably would not have been possible, probably it would not have started at all. Starting point within DLR was the first visionary projects competition launched in 2001 by the DLR board of directors under Walter Kröll. In this novel approach to generate and support innovative ideas the "Virtual Tower" proposal, submitted by one of the editors (N.F.) together with Markus Schmidt (one of the co-authors) and Bernd Werther (now with VW-Research) won a first prize. Well equipped with the prize money the core team was able to start the initial 2-years concept study and engage a software engineer (Michael Rudolph, co-author of Chapter 7) as fourth team member. In the years to come he designed and wrote all of DLR's Remote Tower related software code.

We acknowledge the contributions of the growing Remote Tower staff during the following two RTO projects (RApTOR: 2004–2007; RAiCE: 2008–2012): Maik Friedrich, Monika Mittendorf, Christoph Möhlenbrink, Anne Papenfuss and Tristan Schindler, some of them co- and chapter authors of the present book. They increasingly took over workshares of the RTO research, in articular addressing simulation trials and validation. The RTO-team furthermore was supported by colleagues from the DLR Institute of Optical Sensor Systems (Winfried Halle, Emanuel Schlüßler, Ines Ernst), who contributed to the image processing, movement and object detection (see Chapters 6, 7). RTO validation gained additional momentum with the start of an EC-funded validation-project together with DFS within the SESAR ATM-research joint undertaking, after finishing the RAiCe-shadow mode validation experiments.

The editor of this volume is particularly indebted to Steve Ellis (NASA-Ames/Moffett Field), author of the Foreword, of Chapter 2, and co-author of Chapter 16. As a kind of spiritus rector of the Virtual Tower idea he demonstrated in his Advanced Displays Lab. the initial concrete realisation, based on stereoscopic head mounted displays, which inspired us for submitting our initial proposal in 2001. Nearly ten years later, in 2010 he again advanced our research as host for the editor, spending a research semester as a guest scientist in his lab. In turn, during this period also Steve worked for two weeks as a guest researcher in the DLR Remote Tower Simulator where he introduced his profound psychophysics expertise into the

methodology repertoire of the RTO-research, supervising, performing and analyzing the video frame-rate experiments described in Chapter 16.

At the occasion of several international Remote Tower workshops and mutual visits and meetings at DLR's Braunschweig research facilities, with the Swedish air navigation service provider LFV in Malmö, with FAA/Washington, and with companies Searidge/Ottawa and Frequentis/Vienna we exchanged ideas and discussed problems and perspectives. I am very happy that besides Steve Ellis also several of the other colleagues and experts from external institutions and companies involved in the RTO research and development were able to contribute chapters to the present book. Specifically I would like to express my sincere thanks to the following colleagues who invested a considerable amount of work and time to help this book to provide the first overview on the worldwide endeavour towards the Virtual Control Tower: Rodney Leitner and Astrid Oehme from Human Factors Consult/Berlin for Chapter 20 on Multiple Airport Control, Dorion Liston from San José State University and NASA-Ames as co-author to Chapter 2 on the basics of visual cues used by controllers, Jan Joris Roessingh and Frans van Schaik from NLR/Netherlands who together with colleagues from LFV and Saab/Sweden contributed Chapters 3 and 18 on the basics of detection and recognition and on the Swedish RTO system, and Vilas Nene from MITRE/United States who provided an extensive overview on the US activities.

At this point one remark should be included concerning possible missing information and errors which may have been overlooked during the iteration of the manuscript to its final state. Most chapters are extended versions derived from previous publications, e.g. in conference proceedings volumes that underwent a selection process, usually including modest reviews, which typically however are less strict than journal contributions. All chapters were reviewed by the editor and all of them underwent at least one revision, some of them more. Nevertheless, we can not exclude that the critical reader and in particular the domain experts may detect unclear, maybe even false statements or missing information. Of course the editor and all Chapter authors will be happy about any feedback concerning errors and sugggestions for improvements that may be included in a followup edition of the present volume.

Mentioning the domain experts we certainly have to express our greates appretiation for long years of support and cooperation by active controllers and expert managers from Deutsche Flugsicherung (DFS), the German Air Navigation Service Provider. In particular in the early phase basic domain knowledge was provided during numerous discussions and meetings with Detlef Schulz-Rückert, Holger Uhlmann, Dieter Bensch and others which was used for a systematic work and task analysis. Later on a formal Remote Airport Cooperation (RAiCon) was started and many more experts and managers (we would like to mention Thorsten Heeb and Nina Becker) helped in defining requirements and setting up the experimental system at Erfurt airport for performing the initial validation experiment under quasi-operational conditions.

Special thanks are due to Dirk Kügler, director of the DLR Institute of Flight Guidance since 2008. One of his first tasks was a signature under the just finished RAiCe project plan. Since that time he showed continuous interest in the RTO activities and supported the project by intensifying the cooperation with DFS, resulting

in the formal RAiCon cooperation. Due to his engagement the Virtual Tower patent was successfully licensed to company Frequentis/Austria and a cooperation agreement signed in 5/2015. A month later Frequentis won the DFS contract for realizing the first commercial RTO system in Germany to be installed and validated on the airport of Saarbrücken. After successful validation DFS plans to set up two more RTO systems at airports Erfurt (location of the DLR-DFS validation trials of 2012, see Chapter 7, 9, 10) and Dresden (location of DLR's initial live Augmented Vision test, see Chapter 1), and start with a first Remote Tower Center operation from airport Halle/Leipzig for the three remote airports.

Last but not least we would like to express our thanks to Dr. Brigitte Brunner as the responsible science officer of the DLR program directorate. In an always supportive way she accompanied both DLR Remote Tower projects from the beginning. She provided extra ressources when there was urgent need, e.g. when the necessity of tower controller recruitment for human-in-the-loop simulations surfaced and it turned out that we had been kind of naïve with regard to the cost involved. She was tolerant and supportive also when things did not run as planned (as every active scientist and engineer knows, this is of course characteristic of any "real" research project), and when towards the planned project end it turned out that an extra half year was required for the shadow-mode trials, initial data evaluation, and for finishing the undertaking with an international final workshop. The proceedings booklet of this event, containing the extended abstracts of the presentations was the starting point for the present book.

Finally I would like to thank the team of Springer Publishers for their professional support, specifically Mrs. Silvia Schilgerius, Senior Editor Applied Sciences who encouraged me to start this endeavour nearly two years ago and Mrs. Kay Stoll, Project Coordinator who in a helpful way and patiently accompanied the gradual evolution from abstract collection through repeated manuscript iterations into the present thirteen chapters volume: thank you, it was fun!

Braunschweig, Germany   Norbert Fürstenau
11 October 2015

Contents

About the Editor and Co-Editors

About the Editor

Dr. Phil. nat. Norbert Fürstenau, Dipl. Phys. Since 1971 study of physics at the Universities of Braunschweig, Darmstadt, and Frankfurt. In 1981, Dr. Phil. nat. degree from Frankfurt University with a work on Laser Micro Mass Analysis in Biophysics. After postdoc research on Laser-induced cluster molecules, since 1981 research associate at DLR Inst. of Flight Guidance, and until 2000 group leader of photonic sensors research. In 2001, initialization of "Virtual Tower" research in the Human Factors division. Head of DLR's Remote Tower projects (2002–2012). Since 2016 retired and as scientific consultant research continued in cognitive load analysis. More than 100 journal and conference papers and 10 patents (including "Virtual Tower") in different research fields. Editor of the "Virtual and Remote Tower" book (1st edition 2016). Winner of DLR's 1st "Visionary Projects competition" ("Virtual Tower" proposal, 2001) and the Manfred-Fuchs Innovation award (2019, together with 2nd edition co-editor J. Jakobi).

Co-Editors

Jörn Jakobi, Dipl. Psych. received his Diploma in Psychology in 1999. Since 2000, he works as a human factors expert with DLR Institute of Flight Guidance in the domain of airport airside traffic management with a particular focus on concept design and validation of A-SMGCS and Remote Tower systems. Since 2010, he took over the position of a business developer with a main focus on the topic Remote Tower. In 2014, he became Chairman of the EUROCAE WG100 "Remote & Virtual Tower" and in 2016, the Single European Sky ATM Research program (SESAR2020) launched the biggest European Remote Tower Project with more than 30 partners, which since then is managed by Jörn Jakobi in the project coordinator role. In 2019,

he received the Manfred-Fuchs Innovation award for his achievements with Remote Tower innovations and implementation support (together with editor N. Fürstenau).

Dr. phil., Anne Papenfuss is a researcher at the Human Factors department of DLRs Institute of Flight Guidance since 2008. Her research field is teamwork in air traffic management, how it can be assessed and measured, with a focus on automating communication analysis. She received her Ph.D. from TU Berlin in 2019 in the field of performance-related communication behaviors of air traffic controllers. Within DLR's research on remote tower, she organized the first simulation studies for remote tower and remote tower center operations to assess the impact of these concepts on human performance.

Abbreviations

2-D	Two-Dimensional
3-D	Three-Dimensional
AC	Alternate Current
A/C	Aircraft
ACC	Area Control Center
ADD	Aircraft-Derived Data
ADS-B	Automatic Dependent Surveillance-Broadcast
AFIS	Aerodrome Flight Information Service
AFISO	Aeronautical Flight Information Services Operator
AGL	Above Ground Level
AIP	Aeronautical Information Publication (airport data)
AMC	Air Movement Control
AMS	Acquisition Management System
ANOVA	Analysis of Variance
ANOCOVA	Analysis of Covariance
ANSP	Air Navigation Service Provider
ANT	Automated NextGen Tower
AOI	Area of Interest
AP	Airport
APDU	Aircraft Position Display Unit
APREQ	Approval Request
AR	Augmented Reality
ARSR	Air Route Surveillance Radar
ART	Advanced Remote Tower (EC-FP6 project)
ARTCC	Area Route Traffic Control Center
ASD	Aeronautical Services Department (of SRD, Ireland)
ASDE	Airport Surface Detection Equipment
A-SMGCS	Advanced Surface Movement Guidance and Control System
ATC	Air Traffic Control
ATCO	Air Traffic Control Officer
ATCT	Air Traffic Control Tower

ATIS	Automatic Terminal Information Service
ATM	Air Traffic Management
ATMDC	Air Traffic Management Development Centre (of NATS)
ATS	Air Traffic Services
ATWIT	Air Traffic Workload Input Technique
AV	Augmented Vision
BWE	Braunschweig Wolfsburg Airport
CAMI	Civil Aerospace Medical Institute (US)
CAPAN	Capacity Planning for Networks
CAPEX	Capital Expenditure (airport capital investment)
CAT	Category
CAVOK	Clouds and Visibility OK
CDTI	Cockpit Display for Traffic Information
CERDEC	Communications-Electronics Research, Development, and Engineering Center (US Army)
CHI	Computer–Human Interface
CHMI	Collaboration Human–Machine Interface (Eurocontrol tool)
CI	Confidence Interval
CoDec	Compression–Decompression
CNN	Convolutional Neural Network
CTAF	Common Traffic Advisory Frequency
CTR	Control Range
CWA	Cognitive Work Analysis
CWP	Controller Working Position
DAA	Dublin Airport Authority
D.C.	District of Columbia
DFHM	Dual Frequency Head Maps (EEG)
DG-TREN	Directorate-General for Transport and Energy
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center)
DME	Distance Measuring Equipment
DoT	(Irish) Department of Transportation
DRRP	Detection and Recognition Range Performance
DST	Decision Support Tool
EASA	European Aviation Safety Agency
EEG	Electro Encephalogram
EFS	Electronic Flight Strip
ENRI	Electronic Navigation Research Institute (Japan)
E-OCVM	European Operational Concept Validation Methodology
EU	European Union
EUROCAE	European Organization for Civil Aviation Equipment
FAA	Federal Aviation Administration (US)
FACE	Flight object Administration Centre
FACSIM	Forward Air Controller Traffic Simulator
FFT	Fast Fourier Transform

FHA	Functional Hazard Assessment
FMS	Flight Management System
FOD	Foreign Object and Debris
Fps	(video) frames per second
FSC	Flight Service Centre
FSS	Flight Service Station
FOV	Field of View
GA	General Aviation
GLM	Generalized Linear Model
GMC	Ground Movement Control
GMU	George Mason University
GPS	Global Positioning System
GPU	Graphics Processing Unit
GPGPU	General-Purpose GPU
HCI	Human–Computer Interaction
HD	High Definition (video)
HF	Human Factors
HFOV	Horizontal Field-of-View
HITL	Human-in-the-Loop (Simulations)
HMI	Human–Machine Interface
HR	Heart Rate
HRV	Heart Rate Variation
IAA	Irish Aviation Authority
ICAO	International Civil Aviation Organization (UN body)
ID	Identification, (tracking) Identifier
IDP	Information Data Processing
IDVS	Information Data Handling System: System for displaying weather information
IEA	International Ergonomics Association
IFAC	International Federation of Automatic Control
IFATCA	International Federation of Air Traffic Controllers
IFIP	International Federation for Information Processing
IFR	Instrument Flight Rules
IFORS	International Federation of Operational Research Societies
IP	Internet Protocol, also: Information Processing
IPME	Integrated Performance Modeling Environment
IR	Infrared
ISA	Instantaneous Self Assessment
ITV	Industrial Television (Camera)
JCAB	Japan Civil Aviation Bureau
JND	Just Noticeable Difference (Webers Law)
JPDO	Joint Planning and Development Office
KATL	Hartsfield–Jackson Atlanta International Airport
KBBG	Branson Airport
KDCA	Ronald Reagan Washington National Airport

KDFW	Dallas–Fort Worth International Airport
Km	Kilometer
LAX	Los Angeles International Airport
LFV	Luftfartsverket, Swedish Air Navigation Service Provider
LM	Linear Model
LSD	Large-Scale Demonstration
MANTEA	Management of surface Traffic in European Airports (EC Project)
MASPS	Minimum Aviation System Performance Specification
METAR	Meteorological Aerodrome Report
MIT	Massachusetts Institute of Technology
MLAT	Multilateration (surveillance) System
MPID	Multi-Purpose Information Display
MRTM	Multiple Remote Tower Module
MRT(O)	Multiple Remote Tower (Operation)
MWL	Mental Work Load
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
NATCA	National Air Traffic Controllers Association (US)
NATS	National Air Traffic Services (UK)
NextGen	Next Generation Air Transportation Program (FAA, USA)
NIEC	NextGen Integration and Evaluation Capability
NL	Non-Linear
NLR	Nationaal Lucht- en Ruimtevaartlaboratorium
NM	Nautical Miles
NOTAM	Notices to Airmen
NRC	National Research Council (US)
NSA	National Supervisory Authority
NTA	Non-Towered Airport
NT	NextGen Tower
OCTPASS	Optically Connected Passive (MLAT) Surveillance System (Japan)
ODT	Objective Detection Time
OIS	Operational Improvement Step
OPEX	Operating Expenditure (airport operating cost)
OTW	Out-of-The-Window
PANS	Procedures for Air Navigation Services
PIP	Picture-In-Picture
PSR	Primary Surveillance Radar
PSSA	Preliminary System Safety Assessment
PTZ	Pan-Tilt-Zoom (Camera)
QNH	Atmospheric Pressure
RAG	Remote Air–Ground Communication
RAiCe	Remote Airport traffic Center (DLR project 2008–2012)
RApTOR	Remote Airport Tower Operation Research (DLR project 2004–2007)
RC	(rate or frequency of) Radio Calls
RCD	Radio Call Duration

RD	(cumulative) Radio (Call) Duration
RDP	Radar Data Processing
RJCR	Rishiri Airport (Japan)
RJEO	Okushiri Airport (Japan)
RNLAF	Royal Netherlands Airforce
RoF	Radio-over-Fiber
ROT	Remotely Operated Tower
RT	Recognition Time
RTC	Remote Tower Center/Remote Tower Control
RTM	Remote Tower Module
RTO	Remote Tower Operation
RVR	Runway Visual Range
SA	Situational Awareness
SAM	Safety Assessment Methodology
SATI	SHAPE Automation Trust Index (Eurocontrol)
SD	Standard Deviation
SDT	Signal Detection Theory
SES	Single European Sky
SESAR	Single European Sky ATM Research
SFO	San Francisco International Airport
SID	Standard Instrument Departures
SMC	Surface Movement Control
SMR	Surface Movement Radar
SMU	Safety Management Unit
SNT	Staffed NextGen Tower
SPECI	(aviation) Selected Special weather report
SRD	Safety Regulation Division (of IAA)
SSA	System Safety Assessment
SSR	Secondary Surveillance Radar
STAR	Standard Terminal Arrival Routes
STDEV	Standard Deviation
STER	Standard Error
TAR	Terminal Approach Radar
TC	Tower Controller
TCAS	Traffic Alert and Collision Avoidance System
TDU	Terminal Display Unit (weather)
TFDPS	Tower Flight Data Processing System
TIL	Traffic Information Log
TL	Task Load
TLX	Task Load Measure
TMA	Terminal Maneuvering Area
TMC	Traffic Management Coordinator
TMI	Traffic Management Initiative
TA	Time Available
TP	Time Pressure (:=TR/TA)

TR	Time Required
TRACON	Terminal Radar Approach Control
TS	Tower Supervisor
TWR	Tower
UN	United Nations
UNICOM	Universal Communications
USA	United States of America
VCS	Voice Communication System
VDOT	Virginia Department of Transportation
VET	Visibility Enhancement Technology
VFR	Visual Flight Rules
VHF	Very High Frequency
VIS	Visible Spectrum
ViTo	Virtual Tower (DLR project 2002–2004; winner of 1st Visionary Projects Competition)
VPA	Verbal Protocol Analysis
VPN	Virtual Private Network
VR	Virtual Reality
VSATS	Virginia Small Aircraft Transportation System (SATS) Laboratory
WAK	Workload Assessment Keypad
WAM	Wide Area Multilateration
WdV	Wettbewerb der Visionen (DLR visionary projects competition)
WJHTC	William J. Hughes Technical Center (FAA)
WL	Work Load
WP	Work Package
YOLO	You Only Look Once (CNN framework)

Preconditions

Introduction: Basics, History, and Overview

Norbert Fürstenau, Jörn Jakobi, and Anne Papenfuss

Abstract Since more than fifteen years an increasing interest is observed worldwide in remote control of low traffic airports by means of some kind of virtual control tower. As outlined in the Foreword by Steve Ellis and in the Preface to the 1st ed. of this book, "Virtual Tower" depicts the idea of replacing the out-of-the-window (OTW) view from a conventional control tower on airports by a Remote Tower Control Center (RTC) relying on advanced video, image processing, and virtual reality technologies. It eliminates the need for direct visual observation and consequently the requirement for a costly tower building at an exposed location in visual distance from the runway. The virtual/remote tower idea is connected with a paradigm change in air transportation due to the growth of low-cost carriers and the corresponding increased usage of small airports which nevertheless require air traffic services (ATS). Cost constraints require new ideas and concepts to meet these requirements, and the control of one or more small airports from a remote location without the need of a direct visual observation from a local tower is one of these visions. After providing in Sect. 1 of this introduction some basics of air traffic control in the airport vicinity, we continue in Sect. 2 with a personal account of Virtual and Remote Control Tower research from the DLR perspective. The section covers the time frame from initial ideas around 2001 to 2014, including advanced video and VR technologies, addressing first operational shadow mode tests of the HMI prototype, and finishing with patent licensing to an industrial partner. In Sect. 3 the focus is on the development and implementation phase between 2010 and 2020, covering the important aspect of multiple remote tower (MRT) operation and standardization within the growing international context. The concluding Sect. 4 contains an overview on the twenty-two Chapters and four technical Appendices of this second edition of the Virtual and Remote Tower book.

Keywords Aerodrome control tower · Remote tower operation concept · Multiple remote tower control · ICAO · RTO history · Video panorama · Augmented Vision · RTO prototype · Standardization · Implementation

N. Fürstenau (✉) · J. Jakobi · A. Papenfuss
German Aerospace Center (DLR), Institute of Flight Guidance, Lilienthalplatz 7, 38108 Braunschweig, Germany
e-mail: norbert.fuerstenau@dlr.de

1   Some Basics

The following brief overview puts the Virtual and Remote Control Tower research into the context of standard operating procedures for air traffic control at airports. It refers to procedures for IFR (instrument flight rules) traffic separately for arrival or departure phases of flight. For VFR traffic (visual flight rules, a large part of the general aviation) the procedures may be somewhat different in detail (for more information see e.g. (Mensen, 2003)(Tavanti, 2006). Classically, airport traffic control is performed via cooperation between a group of controllers at different locations as outlined in the workflow schematics of Fig. 1. Controllers of the area control center (ACC, en-route traffic, sector control) take over/hand over the traffic from/to the terminal or approach control (US terminology: TRACON, typically up to 30–50 nautical miles or 50–90 km from the airport). Approach control in turn hands over/takes over the traffic to/from the local or tower control for final approach or departure (airport environment, up to 5–10 nm from the airport).

The flight phases relevant for the remote tower operation (RTO) at aerodromes are approach, landing, taxi, parking for inbound and start-up, push-back, line-up, take-off, lift-off and climbing for outbound flights. Close to the approach area around the airport (upper part of Fig. 1) the flight is handed over from the area control center (ACC) controller to the approach controller and when established on final to the Tower or Runway Controller located at the airport control Tower. Under good visibility the OTW view from the tower cab allows for direct visual observation inside the control range (CTR) (i.e. <ca. 20 km). For inbound flights the Tower Controller
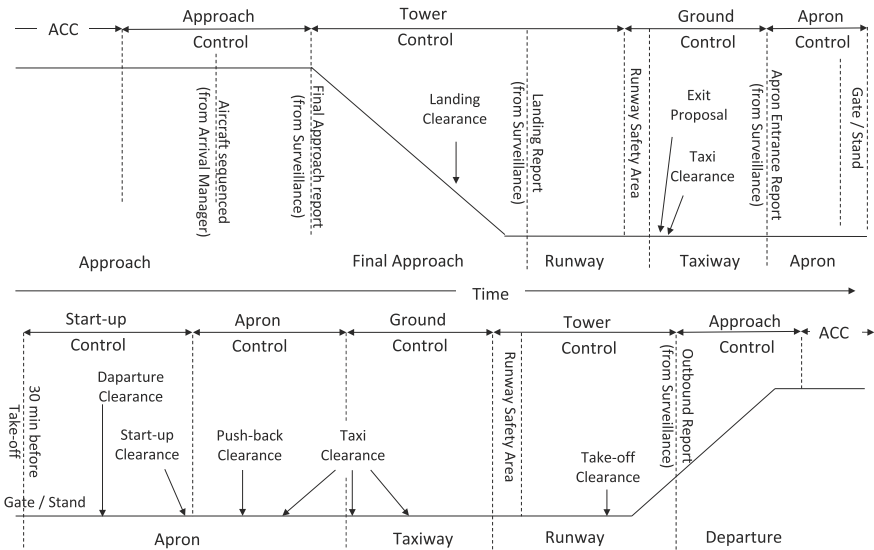
Fig. 1 Workflow schematic of the aerodrome air traffic control, separated in arrival traffic (top) and departure traffic (bottom)

is responsible for all flights from final approach until the aircraft vacates the runway and starts taxiing on the manoeuvring area. At this point the Ground Controller takes over the aircraft on her/his radio channel. When the aircraft enters the apron or ramp area the Ground Controller hands over the traffic to Apron/Ramp controller, who manages the final taxi to the parking position. The mirrored procedure for departure is depicted in the lower part of Fig. 1. An additional function here is granting the departure clearance and start-up clearance, which can be taken over by a separate role at bigger airports that is called clearance delivery controller.

At small airports, which are the main users of Remote Tower technology, all tower control operations may be in the hands of only one or two controllers. Beside small airports with ATC there are also many very small airports, which usually operate only little commercial traffic, mainly VFR traffic and are often uncontrolled and thus offer only lower Air Traffic Service (ATS) levels like Aerodrome Flight Information Services (AFIS) or just a Universal Communication Services (UNICOM). But even those airports could benefit from remote tower technology by sharing ATS effort with other airports or with existing remote tower centers. At request of the pilots, they could offer higher ATS levels like ATC remotely, they could serve as additional alternate airports, or they could prolong their opening hours. This would make them more attractive for their customers and increase their revenues with the same effort. In Chaps. 5 and 22 it is described how RTO can be adopted also by those very small airports.

In what follows we will continue in Sects. 2 (2001–2014) and 3 (2014–2021) with a historic survey of the development of the Virtual and Remote Control Tower idea. As a kind of continuation of Steve Ellis' personal account in the Foreword it covers the initial ideas and experiments since about 2001. The research started with (at that time very costly) high resolution video and augmented vision technology, followed by more advance Remote Tower concepts like Multiple Remote Control and low cost virtual reality systems for Remote AFIS 20 years later, and after prototype development and validation it ends with recent industrialization and standardization initiatives. Section 4 provides an overview of the 22 separate Chapters of this volume.

2 History of Virtual and Remote Tower Research and Development (2001–2014)

While many current towers on RADAR-equipped airports, even some at busy airports like London-Heathrow, can continue to operate (although with reduced capacity) totally without controllers ever seeing controlled aircraft under contingency conditions, it is clear from controller interviews that usually visual perception OTW plays an important role for control purposes (see Chaps. 2 and 3 of the present volume and references therein, and (Ellis & Liston, 2011; Werther & Uhlmann, 2005; Werther, 2007). In fact, these visual features go beyond those required for aircraft detection, recognition, and identification (e.g. Watson et al., 2009). Potentially important

additional visual information identified by controllers in interviews involve subtle aircraft motion (see Chaps. 2 and 3). The focus on a high-quality video-panorama reconstruction of the far view, one major technological aspect of RTO research, was also based on the ICAO regulations for aerodrome traffic control. Citing ICAO document 4444/Sect. 7, no. 7.1.1.2 (ICAO Int. Civil Aviation Organization, 2001) (ICAO, 2012):

> …Aerodrome controllers shall maintain a *continuous watch* on all flight operations on and in the vicinity of an aerodrome as well as vehicles and personnel on the manoeuvring area. *Watch shall be maintained by visual observation, augmented in low visibility conditions by radar when available.*

For large airports with even "Advanced Surface Movement Guidance and Control Systems (ASMGCS)" this requirement is somewhat relaxed. On small airports with lots of VFR traffic however, besides radio communication and possibly a direction finder the visual information is often the controllers only information source on the traffic situation, maybe supplemented by approach radar. For the initial RTO goal-application of small airports without expensive radar coverage the task would be to create a remote tower work environment without direct OTW view which nevertheless should provide at least the same information and safety level, i.e. for the controller the same if not better mental traffic picture as the conventional tower work environment.

In 2006 Brinton & Atkins of Mosaic ATM Company (Brinton & Atkins, 2006) had concluded that.

> *"Requirements for RTO are beyond capabilities of today's electronic airport surveillance systems", however:*
>
> *"a combination of electronic surveillance, optical surveillance and advanced decision support tools may satisfy the Remote Airport Traffic Service requirements."*

The following overview on RTO R&D is a personal account of the editor (N.F.) of the 1st edition of this volume. It represents the perspective of DLR's Virtual and Remote Tower research and development, starting with initial ideas, covering first prototype tests within a DLR—DFS cooperation (2012), and ending with patent licensing to an industrial partner in 2014.

2.1 Previous Work, Vision and Initial Steps

One very early proposal for a revolutionary new Virtual Control Tower work environment was put forward by Kraiss and Kuhlen (1996) within a scientific colloquium of the DLR Institute of Flight Guidance, organized by one of the editors (Fürstenau, 1996). In their contribution on "Virtual Reality—Technology and Applications" they proposed a VR concept for ATC, based on what they called "Virtual Holography". One proposed solution was the so-called virtual workbench, a table-like stereoscopic projection of the aerodrome traffic, allowing for viewing of 3-D trajectories with free

choice of perspective for the controller. VR projection systems of this type are nowadays commercially available but the actual research towards remote tower operation (RTO) went a more conservative way.

A couple of years after this event the preconditions emerged for the research and development work described in the present book. The initial research environment began to take shape at the DLR Institute of Flight Guidance when the editor proposed a research topic in advanced display systems which built on fifteen years of research in optical sensing technologies for aerospace applications. The idea of investigating the potential of the emerging VR-technologies for aerospace applications had been presented at an internal meeting already back in 1989 after a visit of one of the editors (N.F.) at NASA-Ames (Scott Fisher's VR Lab.) and at Jaron Larnier's famous VR-company VPL Research in Redwood City (Silicon Valley) where the so called data glove had been invented as advanced interaction device for virtual environments. In 1999 the author together with co-workers of the optical sensors group (Markus Schmidt, co-author in this volume, and Bernd Werther, now with VW-research) initiated the research on advanced VR-based human–machine interfaces and interaction systems as first step towards the Virtual Tower idea.

Two years later it was a lucky incident which pushed the realization of Virtual and Remote Tower ideas at DLR a large step forward: the Advanced Displays team had submitted the "Virtual Tower (ViTo)" research proposal to DLR's first Visionary Projects competition in 2001 ("Wettbewerb der Visionen", WdV), initialized under the former chairperson of DLR's board of directors, Walter Kröll. Somewhat unexpected, it actually won a first prize, well endowed with 200,000 € for two years of initial studies and concept development. So in 2002 DLR's Virtual Tower research took off, and remembering the Kraiss & Kuhlen presentation of 1996 the team started with a basic survey on the state of the art of VR-technology in Europe and the US and the shaping of an initial concept (Fürstenau, 2004). The most inspiring Virtual Tower ideas however (because based on well founded psychophysics experiments and theories (see the Foreword to this volume and e.g. (Ellis, 1991)), were imported in in the same year after a visit of the author at Stephen R. Ellis' Advanced Displays Laboratory at NASA Ames Research Center. Steve at that time performed research in fundamental problems and applications of head-mounted stereoscopic displays (HMD), including virtual and augmented reality applications in aerodrome traffic control. One problem was the latency problem involved in updating high-resolution virtual environments such as an aerodrome with synthetic aircraft driven by real data in a fixed laboratory frame of reference. The operators' movements have to be tracked and time-varying HMD-coordinates synchronized with the room-fixed aerodrome coordinates and aircraft positions in real-time in order to generate a 3D-VR environment, a problem that was solved with the help of predictive Kalman-filtering of the movement data.

2.2 Basic Research (2002–2005)

An important step towards initial experimental systems during the two years of the WdV-study was the engagement of a software engineer (Michael Rudolph, co-author of Chap. 7) who in the years to come realized all of DLR's Remote Tower software. The first realized code supported augmented vision experiments using self-made head tracking devices. Later on the complex software environment for videopanorama reconstruction of the tower out-of-windows view and the pan-tilt zoom camera control and augmented vision functions was realized (Chaps. 6 and 7).

This made it possible to start the initial experimental RTO research, beginning with a focus on Augmented Vision aspects for support of tower controllers (Tavanti, 2007) using wearable computing and (at that time) futuristic techniques such as the head mounted Nomad Retinal Laser Scanning Display (HMD, see also (Wickens, 1998)). One motivation for the investigation in this so called optical see-through technology (Barfield & Caudell, 2001) was the perspective to reduce head-down times in the tower so that controllers can read display information without loosing visual contact to the traffic situation on the movement areas (Pinska, 2006) (Tavanti, 2006). Figure 2 shows the first practical testing of a retinal scanning HMD at Frankfurt tower.

Another example was the transparent head-up display in the form of the holographic projection screen which was investigated by means of laboratory experiments (Fürstenau et al., 2004) and tested under operational conditions at Dresden tower as shown in Fig. 3. Here the idea was investigated to augment the air traffic controller's direct view out of the Control Tower windows, e.g. by weather data, approach radar, and flight data information superimposed on the far view, without additional head worn gear.

The DLR team during that time decided to turn away from the original idea of augmenting the controller's view out of the real-tower windows by means of the optical see-through technology and instead to follow the video see-through paradigm, i.e. using the video-reconstruction of the environment as background for super-posed additional information (Barfield & Caudell, 2001). This eliminates the latency problem, i.e. the real world superimposed information delay. The augmented vision research for tower controller support using the holographic projection system was continued for a couple of years through several Ph.D. theses at Eurocontrol Experimental Center in Bretigny/France and NASA-Ames Advanced Displays Lab. under the guidance of Steve Ellis. The focus there was research in stereoscopic systems (Peterson & Pinska, 2006).

One reason for this change of research direction at DLR were contacts to the Tower Section of the German air navigation service provider DFS (Deutsche Flugsicherung) which were initiated right from the beginning of the Virtual Tower research and later-on evolved into formal collaborations. Many discussions with domain experts during this time lead to the question if the Virtual Tower idea could provide a solution for a rather urgent requirement: cost reduction in providing aerodrome control service to small low traffic airports. The reason was the paradigm change in air transport mentioned above: small low traffic airports without electronic surveillance (usually

Fig. 2 Demonstration of a laser retinal scanning display, tested by controllers at Frankfurt tower under operational conditions (2/2003). Inset: Superimposed text depicts augmented vision information displayed by HMD via direct image projection onto the retina by means of a laser scanner. Wearable HMD-computing device at the back of DLR team member Markus Schmidt

surface movement radar SME) were increasingly used by low-cost carriers which nevertheless request controlled airspace, although often only for a few flights or a couple of hours per day. Previous "Dark Tower" experiments of DFS aiming at remote control of a low-traffic airport during night time (with nearly zero traffic) from the tower of a large airport, however without transmission of visual information, had provided initial experience on the potential feasibility of this concept. This requirement for cost reduction and more efficient use of staff resources lead to the main topic of this book: the Remote Tower as paradigm change, for low-traffic airport surveillance from a distant location and the perspective of a single remote tower center (RTC) for aerodrome traffic management of several small airports. The original Virtual Tower idea with synthetic vision displays and VR-technologies for large hub-airports would remain the long term goal, "Remote Tower" the more realistic intermediate step with relaxed technological problems and as little as possible changes of operational procedures for a single RTO working position.
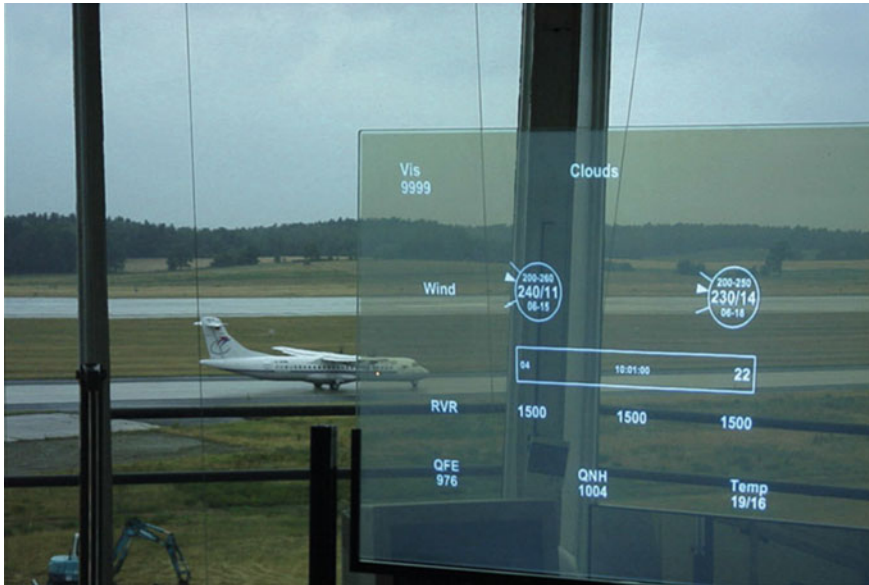
Fig. 3 Demonstration of head-up display based augmented tower vision using a holographic projection display for superimposing live weather information on the out-of-windows view (non-collimated view: image at display distance, tower at Dresden airport, 7/2003, (Schmidt et al., 2006))

At this point the idea of reconstructing the "far view" out of the tower windows by means of a suitable assembly of high-resolution digital video cameras emerged—a "down-to-earth" solution compared with the original "virtual holography" ideas and the VR-HMD display as developed at NASA Ames Research Center. Variants of the latter nevertheless remain a perspective for the future as low cost systems for AFIS (see Chaps. 5 and 22), and as completely sensor driven synthetic vision solution for contingency centers and eventually for the actual Virtual Tower on large airports. Figure 4 depicts the initial experiments during the ViTo concept study with available standard video technology of the late 90's for reconstructing the far view out-of-tower windows. These tests demonstrated the limits of this technology with regard to resolution and contrast, and lead to the requirement for the emerging high resolution cameras (UXGA; HD) based on latest CMOS or CCD chip technology. At that time the cost for a camera of this type was typically >15,000 €, without optics.

2.3   Proof-Of-Concept Project (2005–2008)

The corresponding high quality video-reconstruction of the "far view" became the main technical research topic of the Remote Tower team of the DLR Institute of Flight

Fig. 4 Initial tests (2003) of video based far view reconstruction with standard video technology. Camera position on DLR telemetry antenna tower, ca. 25 m above ground. Camera aiming at Braunschweig airport tower on the dark roof top. White building to the right is location of initial experimental videopanorama camera system (Chap. 6). Runway visible above the camera, extending in west direction

Guidance for the next 8 years (2005–2012), with resources provided by two internally funded projects including a budget of more than 6 M€. The first one (RApTOR: Remote Airport Traffic Operation Research, 2005–2007) as follow-up of the initial ViTo concept study started with intensive contacts between DLR's RTO team and DFS domain experts. Detailed work and task analysis by numerous structured interviews with domain experts were performed by one of the initial core-team members who finished the first doctoral dissertation related to this field (Werther, 2005). At the same time the worldwide first digital 180°-high-resolution live-video panorama as reconstruction of the tower OTW view was realized at the Braunschweig Research Airport, the location of DLR's major aeronautics research facilities (Chap. 6, and Fürstenau et al., 2008a), based on a RTO-patent filed in 2005 and granted in 2008 (Fürstenau et al., 2008b).

In parallel to DLR's research and development of RTO-systems, related activities continued in the US. An experimental system for single camera based remote weather information for small airports using internet-based data transmission had been set up in a NASA-FAA collaboration (Papasin et al., 2001). Clearly such a system could not fulfill requirements comparable to the high resolution low-latency videopanorama system of the DLR approach. Within the US the ATC-modernization

initiative NEXTGen (an analogue to the European SESAR joint undertaking) another direction of research aimed at the so called Staffed NextGen Tower (SNT), addressing the integration of advanced automation into conventional tower equipment with the same long term goal as DLR's WdV-proposal: a completely sensor based work environment without the need for the physical tower building (Hannon et al., 2008). An overview on the US activities is presented by Vilas Nene (MITRE Company) in Chap. 4.

After realization of DLR's experimental system it turned out that meanwhile also the Swedish ANSP (LFV) together with company Saab had started the same kind of development (see Chaps. 3 and 18), also targeting low traffic airports and using more or less the same videopanorama concept. A demonstrator facility was realized in Malmö for initial verification and validation of remote control of a distant airport. This development was continued within the 6th Framework EC project ART (Advanced Remote Tower; see Chap. 18). Since 2010, under the Single European Sky SESAR Joint Undertaking (project 6.9.3) the NORACON consortium with Saab, LFV, and other partners continued the Swedish RTO development and validation. In 2006 the DLR and Saab/LFV teams met for the first time for discussing the remote tower topic at the occasion of the international mid-term assessment workshop of DLR's RApTOr project in Braunschweig.

Meanwhile DLR's Virtual Tower team kept on growing and besides submitting a second RTO-patent application they published first results obtained with the experimental RTO system and initial human-in-the-loop simulations. The most relevant achievements are reviewed and/or referenced in the subsequent chapters of the present volume. Besides technical details on the setup of the experimental system, results of work analysis, realization of a simulation environment for initial human-in-the loop simulations and initial field testing with participation of professional tower controllers are described. The initial field tests described in Chap. 6 were a preparation for so called passive shadow mode testing under realistic operational conditions during the follow-up Remote Tower project "RAiCe" (Remote Airport Traffic Control Center).

2.4 From Prototype Development to Technology Transfer (2008–2014)

This second DLR-internal RTO project started in 2008 and it aimed at realizing a second generation prototype RTO system, investigating RT-Center aspects and testing long distance live high-resolution video panorama transmission. For this goal an advanced RTO system was to be set up at a second airport. The Remote Tower Center (RTC) idea with centralized remote control of ≥ 2 airports was pursued in parallel to the experimental testbed by means of human-in-the-loop simulations in an extended simulation environment (see Part III, Chaps. 11, 12, 13). For this purpose right
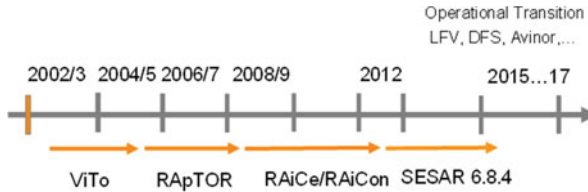
Fig. 5 Timeline of DLR RTO projects until 2015, partly in cooperation with DFS and transition into international cooperations for standardization and implementation (partly funded under the SESAR initiative, see the following section). Not shown are parallel RTO development activities since around 2006 in Sweden (LFV/Saab cooperation) and Canada (NavCanada/Searidge cooperation)

from the beginning of the new project contacts and cooperation between the RTO-team and DFS was intensified. The RTO-topic was selected as one of the strategic goals of DFS, and a DFS RTO-team was formed. A Remote Airport Cooperation agreement was signed for realising the second system at a DFS-controlled airport (project RAiCon, on DLR side headed by Markus Schmidt, co-author of the technical Chaps. 6 and 7). Based on the patented design of the DLR video panorama system and the combination with a professional DFS controller console, the prototype served as passive shadow mode platform for operational validation of the RTO-system with DFS ATCOs at the airport Erfurt during the final year (2012) of the RAiCe-project (see Part II, Chaps. 6, 7, 9, 10). The following sketch summarizes DLR's Remote Tower research since 2002 until the initial shadow mode testing together with DFS and technology transfer to an industrial partner (Fig. 5).

Besides the technical and engineering achievements and the advancement of human-in-the-loop (HITL) simulations the remote tower research also generated methodological progress in experimentation and data analysis within the human–machine interaction research. Again it was the spiritus rector of the virtual tower topic, Steve Ellis (NASA Advanced Displays Lab.; see also his "Foreword") who, based on his psychophysics expertise proposed specific two-alternative decision experiments for quantifying by means of signal detection theory (SDT) the effect of subtle visual cues used by tower controllers for their decision making (see Chaps. 2, 10, 16; Ellis et al., 2011a, b; Fürstenau et al., 2012, 2014). During a research visit of one of the editors of the present volume (N.F.) at the NASA Advanced Displays lab. in 2010, details for corresponding psychophysics experiments were worked out for quantifying video panorama frame-rate requirements for high angular speed (relevant for takeoff and landing). Steve in turn supervised and analyzed the actual experiments as part of a corresponding two-weeks RTO-HITL-simulation campaign, organized by RTO-team members Christoph Möhlenbrink and Anne Papenfuss (Chaps. 12, 13 and 16). Another experiment for determining frame rate requirements by Jakobi and Hagl, following preparations at the DLR tower simulator facility (see Chap. 11) is described in Chap. 17. The same successful methods for quantifying decision making in Chap. 16 were applied later on also to the analysis of results of the shadow-mode validation experiments under quasi-operational conditions (Chap. 10).

The results of the shadow-mode trials and the international final RAiCe workshop in December 2012 marked the beginning of an extended validation project in close cooperation between DLR and DFS, since 2012 funded by SESAR (Single European Sky ATM Research) Joint Undertaking. In close contact with the Swedish group it focused on Human-in-the-Loop simulations and field trials under operational conditions and was expected to help paving the way towards RTO-industrialisation and standardization.

In 2014 after about 10 years of successful Remote Tower research and development at DLR the Remote Tower patent was licensed to company Frequentis/Austria for product development and commercialisation of the RTO concept. In 2015 the Swedish ANSP LFV received its official approval from the Swedish Transport Agency to control the airport of Örnsköldsvik remotely from Sundsval RTC (a system developed by Saab/LFV in parallel and very similar to the DLR system, see Chaps. 3 and 18) (LFV, 2014).

3 Towards MRTO Standardization and Implementation (2010–2021)

The following timeline marks specific events of the (M)RTO research and development including the phase since 2014. It continues until the writing of the present 2nd edition of this volume that contains reports on the important research steps towards worldwide implementation of advanced MRTO (Fig. 6).

After "single" Remote Tower went operational, "Multiple" Remote Tower moved more into the focus of research. Several national and international project initiatives were launched in the 2nd decade of the new millennium. In 2010 within the DLR project RAiCe, (see Sect. 2.3) control of multiple airports was addressed by using human-in-the-loop simulations (Chaps. 11 and 13). For this part of research, general technology feasibility of the core RTO concept was taken as granted and research activities focused on questions more concerned with human performance in "multiple" setting. There was a big question mark whether controlling two or more airports by two or even a single controller at the same time was feasible or if these simultaneous tasks would cause cognitive switching costs that could result in loss of situation
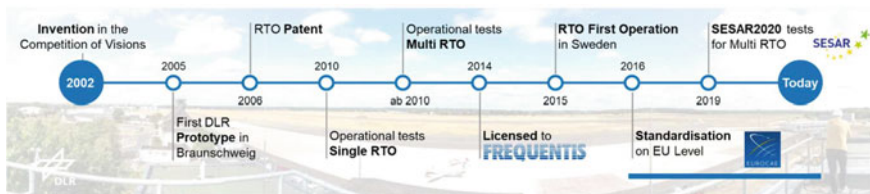
Fig. 6 Extended timeline of MRTO research and development including some important events, until 2021

awareness. By means of HitL simulation experiments, this work paved the way to focus on the crucial points of the multiple remote tower concept—workload, traffic numbers and work organization (Chap. 13).

Other pioneer projects dealing with the MRTO were a workload study within a DLR/DFS cooperation and initial research on EU project level as part of the Single European Sky ATM Research Programme (SESAR), with HITL real time simulations conducted by LFV and SAAB in Malmö/Sweden and DFS and DLR in Braunschweig/Germany. These studies focused on the feasibility of multiple remote tower operations from an operational perspective. In the beginning there were many doubts and concerns in the community, that one air traffic controller is able to handle two traffic situations at two aerodromes in parallel. The mid-air collision that happened 2002 near Überlingen was often cited as an example, where handling two tasks in parallel was identified as one factor contributing to the tragic accident. Consequently, in these studies workload experienced by the ATCOs when handling traffic at two airports was assessed, as well as the controllability of different situations and the influence on situation awareness. At the same time, ANSPs were also interested to understand if certain traffic numbers could be handled in a safe and efficient manner, as the business case of remote tower got more importance. Furthermore, special events typical for small and medium sized airports needed to be considered, for instance medical flights, flight school traffic and traffic under visual flight rules (VFR) in general. Overall, none of these studies' results were a show stopper for the planned implementation of remote tower operations, even so it became clear that introducing remote tower technology would not only replace the tower with video technology but would also change the working style of air traffic controllers.

In 2016 SESAR run into its successor program SESAR2020 (2016–2023) with another huge Remote Tower research project "PJ.05—Remote Tower". Core of the project was the research and implementation question how to allocate flexibly and most efficiently ATCO's/AFISO's productivity to various airports spread over various controller working positions in an RTC with a large number of connected airports. The project also included the development of RTC supervisor support systems and advanced automation functions, the integration of approach for airports connected to the RTC and connections between RTC and flow management units. Results are not directly published in the present volume since the project is still ongoing (August 2021) but first results can be found on the project's homepage (www.remote-tower.eu).

The new research focus on "multiple" and "RTC" settings also generated a need to advance the DLR simulation infrastructure which led to the development of the Remote Tower Lab (Chap. 11), a flexible and modular validation environment for addressing "multiple" remote tower and "center" concepts which was the basis for several validation experiments described in this volume (see Parts III and IV).

From an operational Remote Tower implementation perspective, the first milestone was set on 21st of April 2015: On this day the Swedish Air Navigation Service Provider (LFV) went operational with Airport Örnsköldsvik, which became the first airport in the world to be remotely controlled (LFV, 2014). From this time, local Tower control switched to the RTC at Airport Sundsvall, more than 100 km

south of Airport Örnsköldsvik. Three further airports were connected over the recent years: Sundsvall Timrå Airport, Linköping-Saab Airport and Scandinavian Mountains Airport, the first airport in the world designed from the beginning to be operated remotely rather than having a local air traffic control tower. On June 1st 2021 a second RTC in Stockholm was commissioned with Kiruna Airport as the first airport to be connected. In Germany the Deutsche Flugsicherung (DFS) opened RTC Leipzig in December 2018 with International Airport Saarbrücken as the first airport to be connected and Airports Erfurt and Dresden to be followed short-term. Airport Budapest in Hungary; Cork and Shannon airports controlled from RTC Dublin/Ireland; Bodø/Norway; Schiphol/The Netherlands; Orly/France; Cranfield, Jersey, London City or Heathrow Airport in the UK; Leesburg or Northern Colorado Regional Airport in the US, Singapore or Hong Kong in Asia or Sydney/Australia are other prominent pioneer examples using Remote Tower technology (without claiming to be an exhaustive list). And if one looks at the large number of current tenders, there will be many more implementations in the near future. Also tenders for "multiple" remote towers already exist, so that MRTO will become reality as well in mid-term.

Implementation activities benefit greatly from regulatory and standardized guidelines. Beginning in 2014 standardization bodies like EUROCAE became aware of Remote Tower. EUROCAE is not an official body of the European Governments but is an international non-profit organisation in Europe, which has set itself the task of creating performance specifications and guidance documents for civil aviation equipment, for adoption and use at European and world-wide levels. In June 2014 EUROCAE Working Group 100 "Remote and Virtual Tower" was founded. It consists of active contributors from more than 30 companies worldwide and acts in close coordination with EASA, ICAO, SESAR and EUROCAE WG41 A-SMGCS. In September 2016 the first Minimum Aviation System Performance Specification (MASPS) for "Remote Tower Optical Systems", ED-240, was released. This MASPS is applicable to all optical sensor configurations (including infrared as well as visible spectrum) to be used for the implementation of the remote provision of ATS to an aerodrome, encompassing the whole chain from sensor to display and also optional technologies such as 'visual tracking' and automatic 'PTZ object following'. An extension of the current MASPS (revision B) to cover the processing and integration of information produced by non-optical surveillance systems/sensors, such as PSR, SSR, SMR, WAM/MLAT and ADS-B, is planned with a target publication date in the year 2022. EUROCAE MASPSs are based on state-of-the-art technologies, operational requirements, best implementation practices but also on research findings like Chaps. 15, 16, and 17 in Part III of the present volume, which address dedicated research questions and provide scientific findings for defining sound minimum performance specifications.

To further support the harmonisation and deployment of RTO on the highest level, the International Civil Aviation Organization (ICAO) has updated the Procedures for

Air Navigation Services—Air Traffic Management (PANS-ATM).[1] Amendments to facilitate its use in the provision of aerodrome control service, including a definition of a visual surveillance system and related procedures, were introduced into PANS-ATM by Amendment No. 8 (of 8 November 2018). ICAO has also founded a working group to find gaps or any further changes required to existing ICAO provisions to support a harmonised implementation.[2] In the EU context, the European Union Aviation Safety Agency (EASA) initiated a rulemaking task (EASA RMT.0624 "Technical requirements for the provision of remote aerodrome air traffic services") to develop an appropriate and proportionate regulatory framework including necessary guidelines for "remote aerodrome ATS" implementations.[3,4]

4 Chapter Overview

The 22 separate chapters of the present volume are structured into four parts: I. Preconditions, II. Development and Field Testing of Remote Tower Prototype, III. Human-in-the Loop Simulation for RTO Workload and Design, and IV. Advanced and Multiple RTO: Development, Validation and Implementation. Four appendices address Basic Optics for Video Panorama Design (A), Detection Theory and Bayes Inference for Data Analysis (B), basics of Mental Workload and Measures for (quasi) Real-Time Applications (C), and Psychophysics of Mental Workload: Derivation of Model Equations (D).

Most of the chapters are reviewed, revised and extended versions of previous publications of the DLR RTO-team and of colleagues from other institutions involved in the international endeavour towards the Virtual / Remote Control Tower. Details can be found in the respective chapter's reference lists. Eleven chapters are taken more or less unchanged (apart from minor corrections) from the 1st edition. They describe basic preconditions and the research and development details towards the RTO prototype and are partly based on two RTO special sessions organized by one of the editors as part of the IFAC Human Factors conference in Valenciennes 2010 (Fürstenau, 2010) and the Berlin Workshop Human–Machine-Systems 2011 (Fürstenau, 2011). The framework for the 1st edition originated from the collection of extended abstracts of the international final RAiCe-project workshop which took place in November 2012 (Fürstenau, 2013), as a satellite event of the second EUROCONTROL SESAR Innovation days (SID 2012). The eleven new chapters (including this extended Introduction Chap. 1) address R&D work that followed the publication of the 1st edition with focus mainly on Multiple RTO/RTC, its specific

---

[1] ICAO Document 4444 "Procedures for Air Navigation Services - Air Traffic Management", 16th Edition, 2016.

[2] ICAO Working Paper, ATMOPSP/6-WP/6.

[3] EASA, Terms of Reference for rulemaking task RMT.0624, Issue 2.

[4] EASA, Annex I to ED Decision 2019/004/R, Guidance Material on remote aerodrome air traffic services, Issue 2, February 2019.

simulation and workload issues, and advanced technologies such as implementation of augmented vision functions. It includes contributions from the Electronic Navigation Research Institute in Tokyo (ENRI) and from Cranfield University (UK) in cooperation with the Irish Aviation Authority (IAA).

Part I of the book addresses fundamental aspects and pre-conditions of remote control tower operation and besides this updated Introduction Chap. 1, puts its focus on the visual cues relevant for object detection, recognition and operators decision making, in contributions by Steven R. Ellis with Dorion B. Liston (NASA / Ames Res. Center, Moffett Field) and Frans van Schaik[5] & Jan Joris Roessingh, both NLR (Chaps. 2, 3). Part I also includes a detailed overview by V. Nene (MITRE) on the remote tower research in the US conducted by FAA since 2006 until about 2015, addressing the Staffed Next Gen Tower (SNT) concept and investigations on Wide Area Multilateration (WAM) for improving services at small non towered airports (Chap. 4). In Chap. 5 an overview on remote AFIS in Japan as predecessor of the ongoing RTO research in Japan is provided by Satoru Inoue and Mark Brown from ENRI (Tokyo). The latter is described in more detail in Part II, Chap. 8 (see below).

Part II of the book covers in five chapters the core engineering part of the Remote Tower research and development: the technical Remote Tower design, development, and the initial field testing on the Braunschweig research airport. The basic features of the experimental high resolution video panorama system according to the main Virtual Tower patent (Fürstenau et al., 2008a, b) are described in Chap. 6, including initial verification of system performance with field tests using the DLR DO-228 test aircraft. Included in this chapter are the initial development phase and tests of advanced features: PTZ camera control with automatic A/C tracking, augmented vision using superimposed (video-see- through) information. Design and development of the second generation RTO-prototype system and work environment is described in Chap. 7 by M. Schmidt, M. Rudolph and one of the editors (N.F.). Besides the optical design it addresses basic features of the RTO-software system for live video-panorama construction with image processing for raw-data conversion and compression, the potential of thermal imaging, and aspects of technical verification including electromagnetic compatibility. It also contains a section based on work by Winfried Halle and his team from the Optical Information Systems department, DLR-Berlin, with details on automatic classification, movement, and object detection. In the new Chap. 8 Satoru Inoue, Mark Brown and Yasuyuki Kakubari from ENRI/Tokyo describe in detail the integration of a multilateration system into the high resolution augmented vision video panorama with data fusion of MLAT position information, corresponding to similar initial experiments with the DLR system using A/C transponder data as described briefly in Chap. 6.

In Chap. 9 Maik Friedrich describes the RTO validation experiment of 2012 within the DLR-DFS "Remote Airport Cooperation" (RAiCon, headed by Markus Schmidt), realized as final part of the RAiCe project with focus on shadow mode testing of the

---

[5] We are sad about the fact that Frans passed away during the writing of this 2nd edition. He joined the RTO research from the beginning with well recognized contributions and together with J.J. Roessingh was corresponding author for the Chapters from NLR with LFV and Saab.

RTO prototype. Like in the 2006 initial field testing, again the DLR test-aircraft DO-228 (DCODE) was used to generate a statistically relevant number of reproducible operational scenarios and aircraft maneuvers during aerodrome circling within the Erfurt-airport control zone. This allowed for direct comparison of controller performance under (conventional) tower and remote conditions and for quantitative data analysis using subjective and objective metrics. A detailed analysis of a subset of ATCO's decision tasks in Chap. 10 was based on advanced data analysis methods (Bayes inference and signal detection theory, see Appendix B) for quantification of the decision errors and visual discriminability differences under TWR versus RTO conditions.

The final validation trial with the RTO prototype in 2012 marked the step into the next phase of cooperation between DLR and DFS, now within the European "Single European Sky ATM Research" (SESAR), project 6.8.4. The cooperation within SESAR also supported the international Remote Tower harmonization through close contact with the ongoing Swedish effort towards an operational RTO-system within the LFV-SAAB cooperation.

Part III is concerned with human-in-the-loop (HITL) workload and validation experiments using the DLR tower simulator environment. The focus in the seven chapters is on centralized multiple remote airport control (MRTO) and minimum requirements for usage of advanced augmented vision functions. The specific MRTO simulation environment at the DLR Inst. of Flight Guidance is described by Sebastian Schier (DLR) in Chap. 11. Anne Papenfuss & Christoph Möhlenbrink (DLR) describe in Chap. 12 RTO simulaton studies with 12 professional ATCOs for assessing the operational validity of RTO work organisation including basic augmented vision functions. In the present 2nd edition these simulations experiments are extended in Chap. 13 by two of the editors (N.F. and A.P.) who provide a detailed analysis of workload data from a MRTO simulation experiment with another sample of 12 domain experts. As improvement of the standard (linear) statistical variance analysis (ANOVA) of work and taskload data (originally published in an internal DLR report Lange et al., 2011) the experimental results were analyzed by means of a recently published (nonlinear) psychophysics model based method (Fürstenau & Radüntz, 2021 and Appendix D). In Chap. 16 Ellis & Fürstenau describe a specific two-alternative decision experiment with the same sample of controllers that by means of SDT-discriminability and decision criterion parameters provided an initial estimate of minimum video frame rate requirements for minimizing prediction errors with the dynamic situation of fast moving landing aircraft. These results are complemented by recent investigations of one of the editors (J.J.) together with Maria Hagl in Chap. 17 that address the frame rate requirements for the dominating situations of low angular speed values of vehicles on the movement areas. In two more Chapters of Part III one of the editors (J.J.) together with Anneke Haman reports on HitL experiments targeting the impact of handover procedures on performance in MRT operations (Chap. 14), and in Chap. 15 together with Kim L. Meixner on the results of shadow mode experiments using live video recordings for determination of minimum visual object tracking performance (based on image processing with automatic object detection).

In the five final chapters of Part IV we address Advanced and Multiple RTO development, validation and certification. Roessingh and Schaik[5] from NLR/Netherlands, together with colleagues from the Swedish ANSP LFV and company Saab as RTO-development partners and participants in the European SESAR-funded RTO-consortium "NORACON", report in Chap. 18 details of an RTO approach with focus on advanced image processing and augmented vision technologies including its validation. Fabian Reuschling together with two of the editors (A.P., J.J.) and colleagues from the Fraunhofer-IOSB (Dptm. of Object Recognition, E. Michaelsen, N. Scherer-Negenborn) report in Chap. 19 on the fusion of infrared with visible video streams. In Chap. 20 Rodney Leitner & Astrid Oehme (Human Factors Consult, Berlin) describe the development of a specific planning tool for Multiple Airport Control. Wen-Chin Li from Cranfield University together with Peter Kearney and Graham Braithwaite (Irish Aviation Administration) describe in Chap. 21 the certification process of MRTO with the example of extensive operational MRTO testing of remote control of Shannon and Cork from Dublin RTC. In the final Chap. 22 Fabian Reuschling together with one of the editors (J.J.) takes up the Remote AFIS topic from Chap. 5, describing development and validation of a low-cost RTO concept based on a virtual reality system using pan-tilt zoom camera tracking.

The 22 chapters are complemented by four technical Appendices (A, B, C, D) which are thought to support the readability of this interdisciplinary book. They provide in Appendix A for the technical-optics non-expert some basics of applied optics required for deriving the design and limitations of the videopanorama reconstruction of the tower out-of-windows view (Chaps. 6, 7, 8), and in B for the non-expert of psychophysics methods some basics of two (related) theories that were employed for the data analysis of the visual discrimination/decision experiments (Bayes inference and detection theory SDT, Chaps. 10, 16). Two new Appendices provide in C an overview on methods and measures for determining in (quasi) real time operator's (subjective) workload (used in Chapters of Part III), and in D mathematical details on the derivation of the resource limitation (logistic) model for psychophysics (Stevens law) based prediction and estimation of work load parameters (used in Chap. 13).

References

Barfield, W., & Caudell, T. (2001). *Fundamentals of wearable computers and augmented reality.* (W. Barfield, & T. Caudell, Hrsg.) Lawrence Erlbaum.

Brinton, & Atkins. (2006). Remote airport traffic services concept. In *Proceedings, I-CNS conference, Baltimore 5/2006.* Baltimore.

Ellis, S. R. (1991). *Pictorial communication in virtual and real environments.* Taylor & Francis.

Ellis, S. R., Fürstenau, N., & Mittendorf, M. (2011). Determination of frame rate requirements of video-panorama-based virtual towers using visual discrimination of landing aircraft deceleration during simulated aircraft landing. *Fortschritt-Berichte VDI, 22*(33), 519–524.

Ellis, S. R., & Liston, D. B. (2011). Visual features used by airport tower controllers: some implications for the design of remote or virtual towers. In *Proceedings of the 11th IFAC/IFIP/IFORS/IEA*

symposium on analysis,design.and evaluation of human machine systems, France, 31/8/2010–3/9/2010. IFAC.

Ellis, S. R., Fürstenau, N., & Mittendorf, M. (2011). Frame rate effects on visual discrimination of landing aircraft deceleration: implications for virtual tower design and speed perception. In A. Human Factors and Ergonomics Society (Hrsg.)., (S. 71–75). Las Vegas/USA.

Fürstenau, N., & Radüntz, T. (2021). Power law model for subjective mental workload and validation through air-traffic control human-in-the-loop simulation. *Cognition, Technology, and Work.* Retrieved 30 05 2021, from https://doi.org/10.1007/s101011-021-00681-0

Fürstenau, N., Rudolph, M., Schmidt, M., & Lorenz, B. (2004). On the use of transparent rear projection screens to reduce head-down time in the air-traffic control tower. In *Proceedings of the human performance, situation awareness and automation technology (HAPSA II)* (S. 195–200). Lawrence Erlbaum.

Fürstenau, N., Rudolph, M., Schmidt, M., Werther, B., Hetzheim, H., Halle, W., & Tuchscheerer, W. (2008). Flugverkehr-Leiteinrichtung (Virtueller Tower). *Europäisches Patent EP1791364.*

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., & Halle, W. (2008). Augmented vision videopano-rama system for remote airport tower operation. In I. Grant (Hrsg.), *Proceedings of the ICAS, 26th international congress of the aeronautical sciences.* Anchorage, Sept. 14–19 2008: ICAS.

Fürstenau, N., Mittendorf, M., & Ellis, S. R. (2012). Remote towers: videopanorama frame rate requirements derived from visual discrimination of deceleration during simulated aircraft landing. In *Proceedings: 2nd SESAR innovation days 27th–29th November.* Retrieved 09 08 2021, from https://www.sesarju.eu/sites/default/files/documents/sid/2012/SID%202012-02.pdf

Fürstenau, N., Mittendorf, M., & Friedrich, M. (2014). Discriminability of flight maneuvers and risk of false decisions derived from dual choice decision errors in a videopanorama-based remote tower work position. *Lecture Notes Artificial Intelligence (LNAI)* 8020, 105–114.

Fürstenau, N. (1996). From sensors to situation awareness. *DLR Mitteilung 96–02.* DLR.

Fürstenau, N. (2004). *Virtual Tower. Wettbewerb der Visionen 2001–2004* (pp. 16–19). DLR.

Fürstenau, N. (2010). Virtual tower—special sessions 1,2. In F. Vanderhaegen (Ed.), *Proceedings of the 11th IFAC-HMI conference.* Online: www.ifac-papersonline.net/detailed/47051.html. IFAC.

Fürstenau, N. (2011). Steps towards the remote tower center—special sessions 3a, 3b. In *BWMMS 2011. Published: Fortschritt-Berichte VDI* (vol. 22, no. 33). VDI.

Fürstenau, N. (2013). Remote airport traffic control center (RAiCe). *DLR IB-112–2013/20.* DLR.

Hannon, D., Lee, J., Geyer, M., Mackey, S., Sheridan, T., Francis, M., . . . Malonson, M. (2008). Feasibility evaluation of a staffed virtual tower. *The Journal of Air Traffic Control*, 27–39.

ICAO. (2012). *Twelfth air navigation conference, working paper: Procedures at remote towers.* ICAO Int. Civil Aviation Organisation.

ICAO Int. Civil Aviation Organization. (2001). *Document 444 ATM/501; Procedures for air navigation services.* ICAO.

Kraiss, K., & Kuhlen, T. (1996). Virtual reality—principles and applications. In N. Fürstenau (Ed.), *From sensors to situation awareness. DLR-Mitteilung 112–96–02*, S. 187–208.

Lange, M., Möhlenbrink, C., & Papenfuss, A. (2011). Analyse des Zusammenhangs zwischen dem Workload von Towerlotsen und objektiven Arbeitsparametern. *Internal Report IB 112 2011/46*, 45. German Aerospace Center (DLR), Inst. of Flight Guidance.

LFV. (2014). LFV first in the world to have an operating licence for remote towers. *LFV Press release.*

Mensen, H. (2003). *Handbuch der Luftfahrt.* Springer Verlag.

Papasin, R., Gawdiak, Y., Maluf, D., Leidich, C., & Tran, P. B. (2001). Airport remote tower sensor system. In *Proceedings of the conference advances in aviation safety.*

Peterson, S., & Pinska, E. (2006). Human performance with simulated collimation in transparent projection screens. In *Proceedings of the 2nd international conference on research in air transportation* (S. 231–237). Eurocontrol.

Pinska, E. (2006). An investigation of the head-up time at tower and ground control positions. In *Proceedings of the 5th eurocontrol innovative research workshop* (S. 81–86). Eurocontrol.

Schmidt, M., Rudolph, M., Werther, B., & Fürstenau, N. (2006). Remote airport tower operation with augmented vision video panorama HMI. In *Proceedings 2nd international conference in air transportation* (S. 221–230). Eurocontrol.

Tavanti, M. (2006). Control tower operations: a literature review of task analysis studies. *Eurocontrol Experimental Center EEC Note 05*.

Tavanti, M. (2007). Augmented reality for tower: using scenarios for describing tower activities. In *Proceedings of the 28th digital avionics systems conference (DASC'07)* (S. 5.A.4–1 - 5.A.4–12). IEEE.

Watson, A., Ramirez, C., & Salud, E. (2009). Predicting visibility of aircraft. *PLoS ONE, 4*(5), e5594. Published online 2009 May 20. https://doi.org/10.1371/journal.pone.0005594

Werther, B., & Uhlmann, H. (2005). Ansatz zur modellbasierten Entwicklung eines Lotsenarbeitsplatzes. *VDI-Fortschrittberichte: Zustandserkennung und Systemgestaltung* (vol. 22, pp. 291–294). VDI.

Werther, B. (2005). *Kognitive Modellierung mit farbigen Petrinetzen zue Analyse menschlichen Verhaltens.* DLR FB 2005–26.

Werther, B. (2007). Airport control model for simulation and analysis of airport control processes. In *Proceedings of the HCII2007: Lecture notes computer science* (vol. 4558). Springer.

Wickens, C. D., Mavor, A. S., Parasuraman, R., & McGee, J. P. (1998). *The future of air traffic control—human operators and automation.* National Academy Press.

Visual Features Used by Airport Tower Controllers: Some Implications for the Design of Remote or Virtual Towers

Stephen R. Ellis and Dorion B. Liston

Abstract Visual motion and other visual cues are used by tower controllers to provide important support for their control tasks at and near airports. These cues are particularly important for *anticipated separation.* Some of them, which we call visual features, have been identified from structured interviews and discussions with 24 active air traffic controllers or supervisors. The visual information that these features provide has been analyzed with respect to possible ways it could be presented at a remote tower that does not allow a direct view of the airport. Two types of remote towers are possible. One could be based on a plan-view, map-like computer-generated display of the airport and its immediate surroundings. An alternative would present a composited perspective view of the airport and its surroundings, possibly provided by an array of radially mounted cameras positioned at the airport in lieu of a tower. An initial more detailed analyses of one of the specific landing cues identified by the controllers, landing deceleration, is provided as a basis for evaluating how controllers might detect and use it. Understanding other such cues will help identify the information that may be degraded or lost in a remote or virtual tower not located at the airport. Some initial suggestions how some of the lost visual information may be presented in displays are mentioned. Many of the cues considered involve visual motion, though some important static cues are also discussed.

1 Introduction

The visual cues necessary to fly and land an aircraft have been well studied over many decades (e.g. Gibson et al, 1955; Grunwald & Kohn, 1994). In particular, the degradation in piloting performance and the consequent need to reduce airport capacity due to bad weather is fairly well understood (FAA 71,010.65R, 2006). The

S. R. Ellis (✉)
Human Systems Integration Division, NASA Ames Research Center, Moffett Field, CA, USA

D. B. Liston
NeuroFit, Inc, Mountain View, CA, USA

present report outlines a complementary side of the airport capacity-safety trade-off. It identifies and quantifies some of the visual features and properties used by tower controllers to monitor and enable safe landing and maneuvering on or near airports. These features are especially interesting now due to recent proposals for technology and procedures in which controllers work in towers without a direct view of their controlled space. Such towers are described alternatively as a remote or "virtual tower" (JPDO, 2007). Work in these towers would be supported by controller displays of information about aircraft and the airport environment.

In general two types of displays can be considered: One would present a plan-view, map-like computer-generated display of the airport and its immediate surroundings (JPDO, 2007) somewhat like existing ASDE-x displays (Fig. 1). An alternative could present a composited perspective view, possibly provided by an array of radially mounted cameras positioned at the airport in lieu of a tower (Fürstenau et al., 2008) (Fig. 2). In either case, procedures and display techniques need to be developed which are cognizant of the current visual information used by controllers, which may be lost.

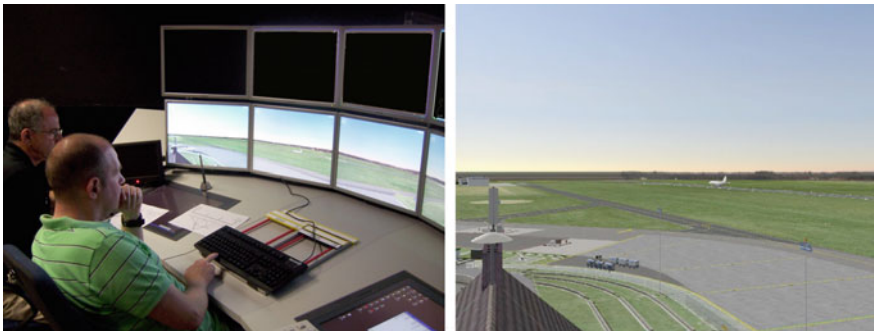**Fig. 1** ASDE-x airport map display

Fig. 2 Out-the-window camera or synthetic vision display format

The following discussion initially points out visual elements of the control task facing the tower evident in previous task analyses of tower operations (Paul et al., 2000; Werther, 2006). However, this earlier work appears to only provide very general descriptions of the specific visual features to which that the controllers attend. To the extent the visual functions that are important to the controllers are considered, they are generally limited to questions of detection, recognition and identification. The following discussion will consider other visual features that go beyond these basic three elements and relate in specific ways to the individual decision processes tower controllers develop to do their job; in particular, we discuss the motion of the controlled aircraft. The preliminary conclusion of the discussion is that tower controllers use visual features to provide predictive position information allowing them to use *anticipated separation* to effectively and safely merge and space aircraft, maximizing airport capacity.

The visual cues used by controllers are important for several reasons. In the first place, there is FAA interest in increasing airport capacity so that current operations under non-visual flight rules with reduced capacity may be modified to allow higher visual flight rules capacity during non-visual operations. For this purpose the currently used visual information needs to be provided by alternative means. Such "Equivalent Visual Operations" described FAA/NASA planning documents may be achieved with synthetic visual systems, i.e., (Kramer et al., 2008) with replacement of direct tower camera or sensor views with visualized electronic position data. This replacement of the direct view, however, will not be fully successful, and may even be tragically misleading, if the useful visual affordances provided by the real scene are not appropriately included or accounted for. Although Equivalent Visual Operations has primarily been considered from the pilot's viewpoint in terms of flight displays which use new sensor data for synthetic vision, it has a flip side for which synthetic vision or camera-based displays could be used to present useful visual information within a remote or virtual tower.

Significantly, this information need not be provided in the form of an image, but could be provided in a more map-like plan view format and conceivably could even come along non-visual sensory channels, e.g. auditory or haptic. In fact, it could be based on data directly down-linked to ground displays from an aircraft indicating its state, i.e. spoilers deployed (Hannon, et al, 2008).

The visual environment in an airport tower may be illustrated by considering the view from a specific tower such as that of San Francisco International Airport (SFO) (Fig. 3 top). Such tower views show significant perspective compression at the ~1 nmi range to runways and taxiways, making commercial aircraft subtend small visual angles, and posing viewing difficulties due to background visual clutter. Interestingly, during low visibility CAT III operations at SFO, airport operations may be conducted with the controllers never actually seeing the aircraft. Thus, since it is already possible for the controllers to continue many of their control tasks without visual contact, albeit with fewer aircraft, the idea of a remote tower may have some prima facie feasibility. But without visual contact, controllers must inform the pilot and those monitoring their communications that visual contact has been lost. Significantly, at the SFO tower where the parallel runways are ~750 ft apart, continued operation without
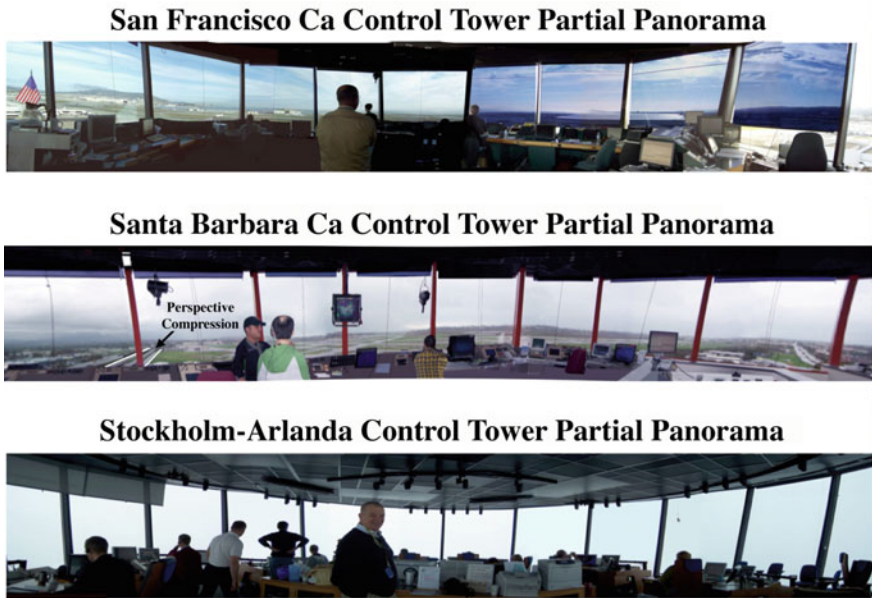
Fig. 3 The variation of visibility within airport tower's immediate environments is shown from unlimited visibility (San Francisco International, top) through partial occlusion due to low clouds (Santa Barbara Municipal, middle), to complete white-out (Stockholm-Arlanda, bottom)

visual contact is associated with a loss (~50%) of airport capacity.[1] In contrast at an airport such as Arlanda, Sweden (ARN) with the parallel runways ~1 km (~3280 ft) apart, total loss of visual contact can have virtually no impact on capacity when the ground radar is fully functional.[2] Thus, there exist some operational examples of tower operation with total loss of visual contact. During low visibility operations, it is not always necessary for the controller to maintain visual contact with the aircraft, but for the aircraft to have enough forward visibility to safely maneuver the aircraft during ground taxi operations.

SFO Operations

An analysis of the role of visual features in tower control can be developed from a more detailed discussion of operations for a particular airport, SFO. A sense of the overall strategy for some aspects of usual airport operation at SFO is best obtained from plan-view maps (see Fig. 4 for SFO map). Aircraft are taxied from their gates to the southwest ends of runways 1L and 1R and launched in staggered pairs to the northeast. Departing aircraft are interleaved between aircraft landing on Runways 28 Left and 28 Right which also are treated as staggered pairs. Current winds, weather, and special operational requirements, of course, can significantly alter this pattern.

---

[1] Personal communication, ATCO, San Francisco International Airport, 7/7/2006.

[2] Personal communication, Tower Supervisor, Arlanda International Airport, 4/23/2007.

Fig. 4 SFO airport diagram showing typical movement paths for United Airlines, departures (dark/red paths), arrivals (light green/paths)

For example, sometimes the longer 28 runways are needed for heavy, departing transpacific aircraft. Detailed descriptions of the alternative approach and departure procedures can be found in the Standard Instrument Departures (SID) and Standard Terminal Arrival Routes (STARS) associated with the airport, but the local controller's responsibility for arriving traffic generally begins with radio contact somewhat before the aircraft crosses the San Mateo Bridge and ends for departing aircraft 1 nmi. beyond the end of the departure runway. By FAA rules, the local controller is generally responsible for aircraft entering and leaving the runways whereas the ground controllers handle, in a coordinated way, most of the taxing to and from the gate. These two positions, in addition to that of the supervisor, are the ones that make the most use of the out-the-window information. The other two tower controller positions, Flight Data and Clearance Delivery, primarily use inside-the-tower information sources and voice communications.

2   Visual Information Used in the Airport Tower

The primary responsibility of the control tower is to ensure sufficient runway separation between landing and departing aircraft (FAA, 2006). A back propagating process may be used to understand the visual requirements supporting the tower controller's primary responsibility.

This process first identifies the visual affordances that the controllers' tasks involve. Affordances are the higher-level behavioral capacities that vision must support (Fig. 5). Controllers, for example, must be able to identify the aircraft type, company and flight status. They must control and recognize aircraft speed, direction and position. They must establish a movement plan involving a succession of spatial goals. They must communicate this plan to the aircraft, coordinate it with other controllers and pilots as necessary, establish whether aircraft comply appropriately, and recognize and resolve spatial and other conflicts that may arise. These

Fig. 5 Description of the dependency of the high level spatial information needed by controllers on progressively low and lower perceptual functions and visual mechanisms

Table 1 Analysis of tower control tasks that inherently involve visual information are printed in bold

	Job Task		Job Subtask
1.	**Separation**	1.	**Separation is ensured and maintained at all times.**
		2.	Safety alerts are provided.
2.	**Coordination**	1.	**Performs handoffs/point-outs.**
		2.	Required co-ordinations are performed.
3.	**Control Judgment**	1.	Good control judgment is applied.
		2.	Priority of duties is understood.
		3.	Positive control is provided.
		4.	Effective traffic flow is maintained.
4.	**Methods/ Procedures**	1.	**Aircraft identity is maintained.**
		2.	**Strip posting is complete/correct.**
		3.	Clearance delivery is complete/correct and timely.
		4.	Letters of Agreement (LOAs)/directives are adhered to.
		5.	Additional services are provided.
		6.	Rapidly recovers from equipment failures and emergencies.
		7.	**Scans entire control environment.**
		8.	Effective working speed is maintained.
5.	**Equipment**	1.	Equipment status information is maintained.
		2.	Equipment capabilities are utilized/understood.
6.	**Communication**	1.	Functions effectively as a radar/tower team member.
		2.	Communication is clear and concise.
		3.	Uses prescribed phraseology.
		4.	Makes only necessary transmissions.
		5.	Uses appropriate communications method.
		6.	Relief briefings are complete and accurate.

higher-level elements are supported visually by a number of visual functions: detection, recognition, and perception of the static and dynamic state of the aircraft. These functions are supported by still lower level visual mechanisms underlie luminance, color, control, position, and movement processing. These three levels of analysis provide a basis for describing the controllers' visual task.

The tower controller's overall task has, of course, been analyzed within and outside of the FAA. It may be broken down to six different job subtasks: separation, coordination, control judgment, methods/procedures, equipment, and communication. The five of these subtasks which involving vision have been identified by boldface type in Table 1 (Ruffner et al., 2003; FAA, 2006).

The assurance and maintenance of spatial separation is, of course, a visual task since regardless whether separation is determined by radar or direct view, it is definitely recognized visually. Handoffs and point-outs clearly are also intrinsically dependent upon vision, though the need for the controller to adopt the pilot's spatial frame of reference to direct attention toward objects and aircraft is also a significant cognitive task. Control judgment, being essentially a mental and cognitive issue, does not have an intrinsically visual component. But its connection with maintenance of effective and efficient traffic flow does emphasize the critical importance of time in traffic control. Three general methods and procedures directly involve vision. These include establishment and maintenance of aircraft identify, posting and correct annotation of flight strips, and continual scanning of the entire control environment.

Associated with these methods is the admonition to work quickly and rapidly recover from errors or off nominal conditions. Because each tower's environment is to some extent unique, the specifics of their procedures differ from tower to tower. All control techniques are, of course, consistent with the regulations cited and described in the FAA air traffic control, *Order 7110.65R*, but unique procedures and heuristics are passed on to future controllers by onsite training. The specific visual features tower controllers use can frequently be found in these locally developed heuristic rules. Some are presented in Table 4 below.

The overall tower control process has been formally analyzed and modeled including visual and nonvisual components (Alexander et al., 1989; Werther, 2006). For example, the MANTEA notation (Paul et al., 2000) has been applied to analyze controller activity in the tower. Some of the elements identified in the MANTEA analyses are, in fact, visual, but the visual components are only described in very general terms such as "visualize runway," "visualize meteo," etc. These descriptions really only identify the sensory modality used to gather the information and a general description of the content of the visual information, but they say nothing specific about the actual visual viewing conditions or about the specific visual stimuli. This feature is, in fact, common in other more recent and more sophisticated task analyses of visual features seen from the tower. Even the recent modeling done with Petri nets (Werther, 2006) does not identify specific visual stimuli but is more concerned with estimates of time required for the precision with which various visual sub-functions maybe executed and to the logical conditions and consequences associated with the functions.

The FAA has done some analysis of the specific visual performance expected from Tower Controllers. The work primarily focuses on the controller's surveillance function and has been based on visual performance models developed for the military by CERDEC at Ft. Belvoir (e.g., Vollmerhausen & Jacobs, 2004). These models primarily are intended to predict the probability of visual detection, recognition, and identification of known targets. "Detection" refers to users' ability to notice the presence of a particular object. "Recognition" refers to their ability to categorize the object into a general class such as a tank, light aircraft, or truck. "Identification" refers to their ability to determine the specific type of object, i.e., an Abrams tank, a Cessena 172, or a Ford refueling tanker. More modern similar visual performance models do not require same amount of calibration techniques to determine model parameters for specific visual targets and specific users. (Watson et al., 2009).

The CERDEC analysis, which predicts specific object perception from towers of various heights during a variety of atmospheric conditions and object distances, has been incorporated into a web tool to help tower designers ensure that specific architectural and site selection decisions for new towers will meet FAA requirements. Significantly, this tool also just focuses on the surveillance function and does not address the aspects of visual motion that tower controllers use for the information, separation and safety tasks (Fig. 6).

In order to understand the details of the visual features used in tower control it is first necessary to identify the range within which controllers use visual information. We can use the example of SFO. Informal voluntary discussions and structured
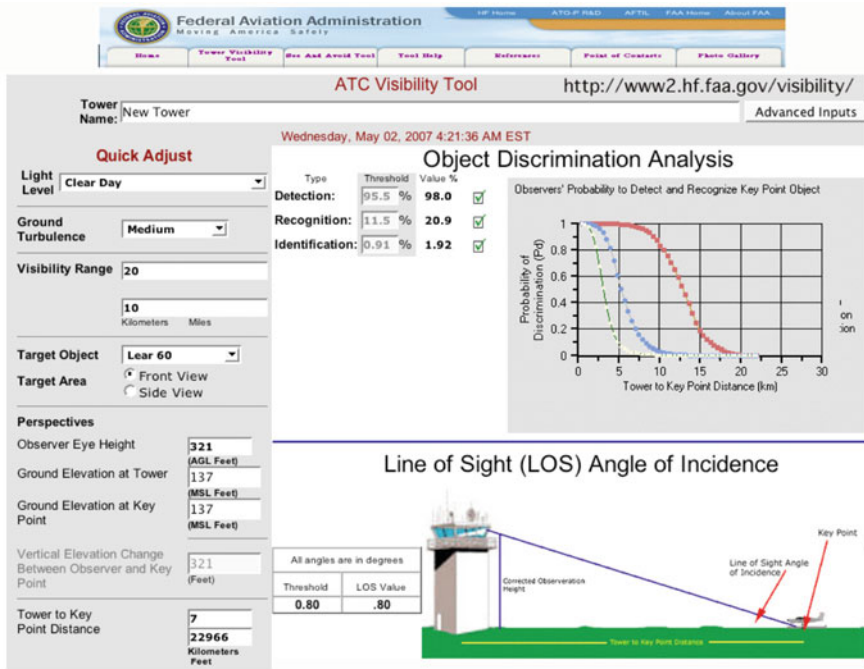
Fig. 6 The WEB interface to the FAA's tower design analysis tool that may be used by municipalities and others to test tower designs ultimately intended for FAA analysis and approval. Note website indicated in the upper right

interviews with ten active controllers and supervisors who work at this tower were analyzed for the physical locations identified as points where various types of visual references are used while controlling approaching or departing aircraft. These discussions, which were considered preliminary work, were conducted with the knowledge and approval of the SFO tower manager, his chain of command, and the local NATCA tower representative. All primary notes were taken without personally identifying markings and transcribed into secondary statistical summaries or grouped data so as to preserve the anonymity of the respondents. Primary notes were thereafter discarded.

These reported points where useful visual information could be seen primarily to include positions where visual contact with the aircraft is first or last considered to be helpful. These positions, marked in Fig. 7, include those for which aircraft come under or leave tower control, where they pass important ground references, or where visual contact provides other useful information. The points were determined independently from each of the controllers in response to the question, "When you are in the Local controller position, where are the aircraft when you usefully observe them visually, what visual aspects of the aircraft do you observe and why?" Controllers could designate more then one point of interest for departing and more than one for arriving traffic; only two controllers took this option. One point represents nine
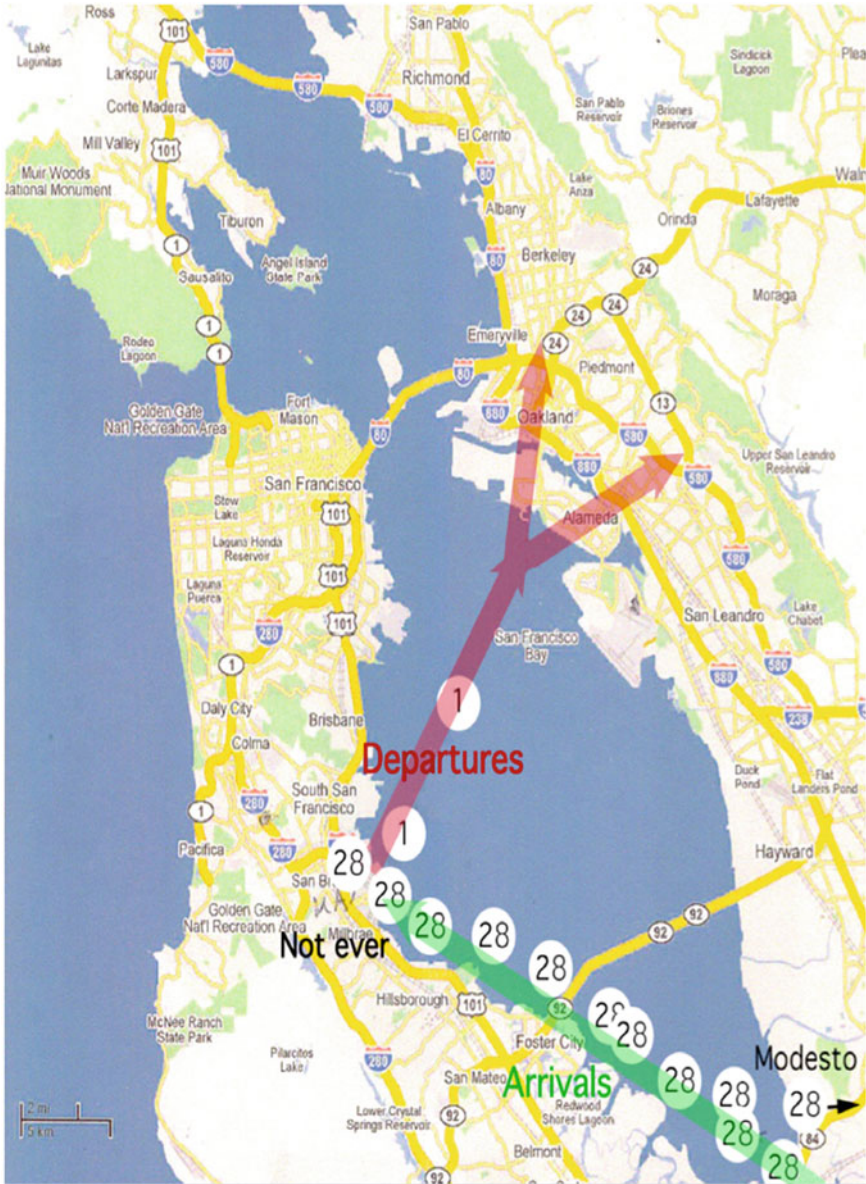
Fig. 7 The first and last positions where SFO controllers report useful visual information w/r to landing (Runway 28) and departing aircraft Runway 1). The arrows show idealized, most common approach paths (transparent green) to the west and departure paths (transparent red) to the north

Table 2 Visual features identified by interviews with ten SFO tower controllers. Boldface marks the predictive aspect of specific visual features

Feature	Times Mentioned	Commentary	Feature	Times Mentioned	Commentary
1. Relative visual motion used to interleave take-offs and landings	5	Controllers **verify predicted separation** using relative motion w/r stationary references to plan clearances	7. Visible wing dip predicts coming turn	3	Visible banking quickly confirms conformance to turn clearance
2. Visual check for obstacles or other A/C to verify a clearance	5	Obstacle checks include ground vehicles, aircraft, birds, people	8. "Mike and a mile" rule for interleaving take offs and landings	3	**Predictive rule**: Departing A/C must be rolling across taxiway Mike on RW1 when matched landing A/C on RW28 is at least 1 mi out for required separation.
3. "Taxing with authority" helps attention allocation	4	Fast and confident A/C motion allows controllers to distribute attention to pilots who maneuver hesitantly allowing **anticipation** of future problems	9. Engine smoke or heat confirms take off start	2	Currently less useful since modern engines don't smoke much and have cooler exhaust
4. Aircraft attitude/ altitude predicts a "Go Around"	4	Controllers anticipate "Go Around" by checking A/C passage through various approach gates defined by altitude and attitude	10. Onset of navigation lights predicts a tower call requesting service	2	Controllers can **anticipate** coming workload
5. Visually apparent acceleration, speed or turn rate anticipates taxiway selection	4	Controllers mentally integrate motion features to **anticipate** taxiway and ground route selection	11. Visual resolution of motion and position is better than by radar at airport	1	1-2 nmi. from the tower the "visual display" of the real world has more "pixels" than associated radar displays
6. Coordinate/cross-check visual and radar data	4	Specific visual land-marks are selected to cross check radar	12. Visual double check on A/C tail to verify company	1	
			13. Check landing gear	1	Probably an isolated comment because it's checked routinely. "Gear down" isn't a problem for major airlines.

controllers' overlapping responses identifying approximately the same location about 1 nmi. beyond the end of the departure Runway 1.

In general it is apparent from the distribution of points that controllers' visual attention is much more spatially distributed to the aircraft approaching the 28LR runways and rather abruptly drops off about 1 mile off the end of the usual departure runways 1LR. These observations refer to the most common aircraft flow at SFO but suggest the generalization that the local controllers' visual attention to approaching aircraft is distributed over a much large area than that corresponding to departing aircraft. A likely reason for this is that departing traffic is handed off to approach/departure control at 1 nmi beyond the end of the runway and generally not thereafter of concern to the tower.

A significant aspect of the controllers' remarks concerning when they first start paying visual attention, or when they last pay attention, to aircraft is that they rarely mentioned the aircraft's visual motion.[3] One reason is that for the viewing angles and distances to the aircraft approaching and departing SFO, this motion is very small in terms of degrees per second, often the azimuth rate is on the order of much less than 0.25°/s and rarely more than 0.5°/s. The visual accelerations are even much smaller and difficult to see because of atmospheric haze, thermal effects, and the visual range being beyond 5 miles. Visual rates of motion are more important for closer aircraft just seconds away from touch-down or from those on taxiways.

Probably the most obvious need for visual contact by controllers in the tower is to immediately note unusual events that are not detected by electronic sensors such as radar. Examples could be heavy bird activity or an aircraft leaking fuel onto a taxiway. But there are a wide variety of other visual features that controllers use on a more regular basis when aircraft are close enough for the visual motion to be more easily noticed. Discussions with controllers have provided a list of some that are used (see Tables 2 and 4).

---

[3] Visual motion is defined as the angular rate of change of the line of sight angle to an aircraft from the tower.

A tabulation (Table 2) of the visual features mentioned in the discussions with each of the SFO controllers shows the relative frequencies with which different features were mentioned. These discussions used a "cognitive walk-through" technique in which the controllers were asked to imagine representative approaching, departing, and taxiing aircraft under a variety of visual conditions and to report what they looked for visually to assist their control tasks. The consequent discussions were guided by the elements outlined in Ellis and Liston (2011, Appendix). The most frequently mentioned features were relative motion between landing or departing aircraft and obstacles that could be on the runway. The first of these features is probably prominent because SFO has intersecting runways commonly used for takeoffs and landings. An assessment of all of the features mentioned, however, shows what may be a more general element. Seven of the 13 features identified in the interviews note that the feature helps the controller anticipate future activity. This information provides insight into pilot intent, knowledge, and likelihood of aberrant behavior. These predictive cues help the controller with the short term trajectory planning needed for *anticipated separation* and help them allocate their attention to pilots either unfamiliar with the airport or maneuvering in unexpected ways.

3 Visual Features at SFO

In order to examine the generality of the visual features and produce a list as complete as possible, structured anonymous interviews were conducted with controllers from an additional seven airports. Because we were not able to obtain, timely agreement from the national NATCA office for the participation of line controllers, these additional discussions were limited to supervisory personnel. Anonymity was maintained since all written notes were taken without personally identifying markings and formal questionnaires were not used. To insure anonymity, original notes were transcribed into statistical or grouped secondary notes and the originals were thereafter discarded insuring that no personally identifiable information was recorded or could be reconstructed post hoc. In all cases tower visits to U.S. airports were conducted with the knowledge and approval of the specific tower's manager and FAA headquarters. U.S. airport towers in addition to that of San Francisco International Airport (SFO) that were visited were: Boston International (BOS) MA, Golden Triangle Regional (GTR) MS, Santa Barbara Municipal (SBA) Santa Barbara, CA, and Norman Y. Mineta San Jose International (SJC), San Jose, CA. Supervisory controllers from Denver International (DEN) Denver, CO, Laguardia Airport (LGA), New York City, NY, and Philadelphia International (PHL) Philidelphia, PA were included in the multi-airport analysis. They visited the first author at NASA Ames and provided information regarding the nature and location of visual features used by controllers while viewing airport diagrams and regional maps. The tower at Stockholm-Arlanda (ARN) in Sweden was the only foreign airport tower visited but was not included in any quantitative analysis. For a summary of the airports towers considered and the personnel interviewed see Table 3.

Table 3 Airport towers environments discussed and evaluated

Airport tower environments discussed	Number of controllers or supervisors	Notes
Stockholm (Arlanda) ARL	1	Discussions were held but visual features from the ARL tower were not analyzed
Boston International (BOS)	3	Supervisors only
Denver International (DEN)	1	Supervisors only without airport view
Golden Triangle Regional (GTR)	1	Supervisors only
La Guardia International (LGA)	1	One supervisor without airport view
Philadelphia International (PHL)	1	One supervisor without airport view
Santa Barbara City (SBA)	2	Supervisors only
San Jose International (SJC)	3	Supervisors only
San Francisco International (SFO)	11	One supervisor, 10 controllers
Total	24	

Figures 8 and 9 illustrate how the visual velocity of aircraft viewed from the tower could be determined for moving aircraft at or near the airport and those that were farther away in the airport vicinity but still visible. Figure 10 provides a breakdown of various classes of features as 14 general categories that were used to organize the features. Counts on the numbers in each category give some idea of their relative frequency of mention. At this stage of investigation no systematic attempt was made to determine the relative operational importance or frequency of use of the various features. Investigations are currently underway in collaboration with Jerry Crutchfield of the Civil Aerospace Medical Institute (CAMI) to determine the frequency of use and criticality of the visual features that have been identified[4] (also see van Schaik et al., 2010 and Chap. 3). In particular, the high frequency of mention of the points of first and last useful visual contact are undoubtedly an artifact of their mention in the structured interview as an example of the kind of visual information being sought. The point of the investigation was to collect as broad a range of visual features as possible for further analysis in subsequent studies.

When a visual feature was identified by a controller, its location was plotted on an appropriate map. Afterwards, the direction of flight and speed was determined from the appropriate airborne traffic pattern or ground path. Simple geometric analysis was then possible to determine the apparent visual rate of the aircraft as seen from the tower at the time the visual feature would have been noted. Because actual aircraft

---

[4] The project is called: Concurrent Validation of AT-SAT for Tower Controller Hiring (CoVATCH). AT-SAT stands for Air Traffic Selection and Training test battery.

Table 4 Visual and other perceptual features that aid tower air traffic control

Visual feature A/C = aircraft	Visual information provided	Information and display techniques: map-like displays	Information and display techniques: out-the-window image-like displays
Status			
1. A/C is prepositioned with an anticipatory rotation for a turn while holding short of a taxiway or runway	Pilot is correctly expecting to be cleared for a specific turn	Current and static A/C orientation should be shown on electronic map	Visual resolution of display should be sufficient for user to recognize A/C pose at crossing points
2. A/C type	Predicts likely ground acceleration, e.g. the difference between turbine vs. constant speed propeller. A/C type determines separ-ation techniques used	A/C type should be indicated by icon shape or data tag to relieve controller memory load	High resolution visual image required to support existing visual performance requirements for tower design
3. Dust up or thermal optical distortion from thrust	Applied power can confirm compliance with take-off or other clearances that require engine spool-up	Down-linked indications from A/C of engine spool up should be displayed A/C icon	Evidence of spool up should be visible on display or A/C icon associated with the power up should be dis-played based on down-linked information
4. Smoke, spray from wheel indicates ground contact, and touchdown point	Touch down indicates landing likely unless a touch-and-go is planned. Helps to identify likely taxiway to be used to exit runway	Down-linked information from wheel sensors indicating touchdown should be displayed on A/C icon to indicate touchdown point	Visual evidence of wheel contact should be visible or down-linked information from wheel sensors indicating touchdown should be displayed on A/C icon
5. Navigation lights being turned on	Call to tower is imminent, usually to the Clearance Delivery Controller at a big tower	Navigation lights when A/C at gate should be visible. Down-linked information regarding A/C before engine start should be displayed if visibility is insufficient before pilot calls tower	Navigation lights when A/C at gate should be visible, Down-linked information regarding A/C before engine start should be displayed if visibility is insufficient before pilot calls tower
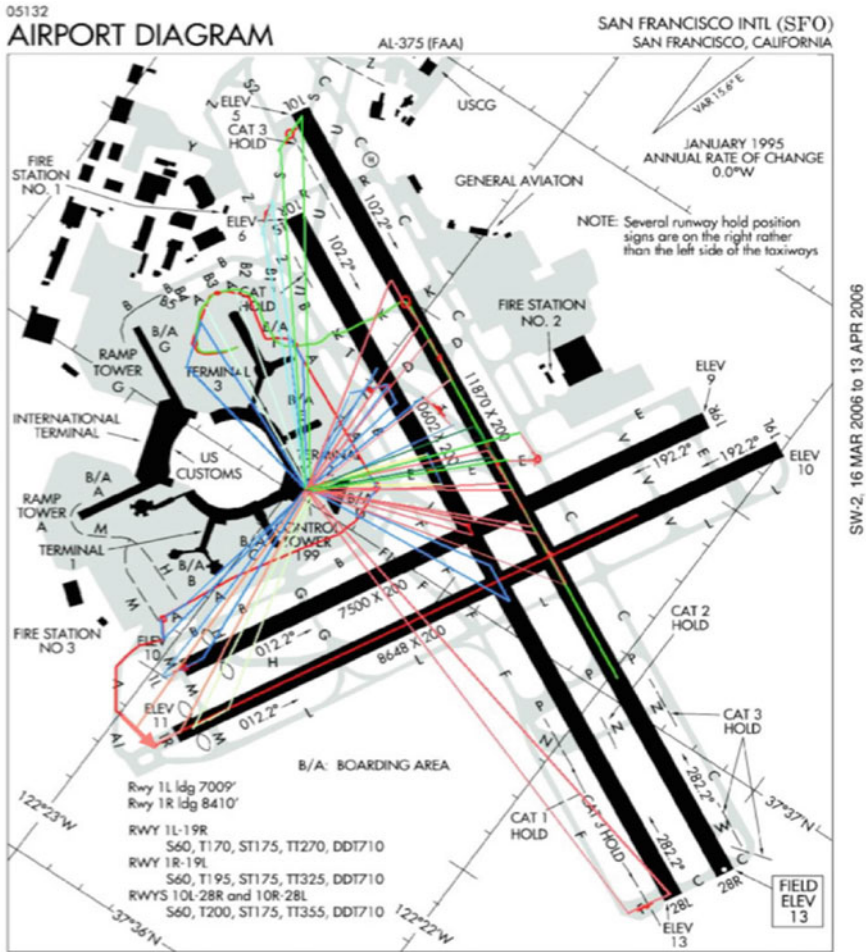
Fig. 8 Lines of sight from the San Francisco International Airport tower to positions on the airport where the visual motion was analyzed. Simple geometry allows calculation of rates of change of lines of sight from the tower to aircraft from knowledge of tower and aircraft position and aircraft velocity

speed was not actually measured, speed was estimated from typical rates mandated by approach procedures or estimated by controllers and pilots familiar with the airport and typical air and ground aircraft motion. Some reflection on the geometry shows, however, the aircraft speed to have a comparatively small influence on visual motion. Its impact is dwarfed by the effect of relative direction of flight. An aircraft flying directly towards the tower can have virtually 0 °/sec visual velocity! The relative direction of flight used for analyses was determined from the interviewees and the typical patterns of motion at and around the airport if the original notes did not include the needed information. Once the approximate visual velocity associated with each
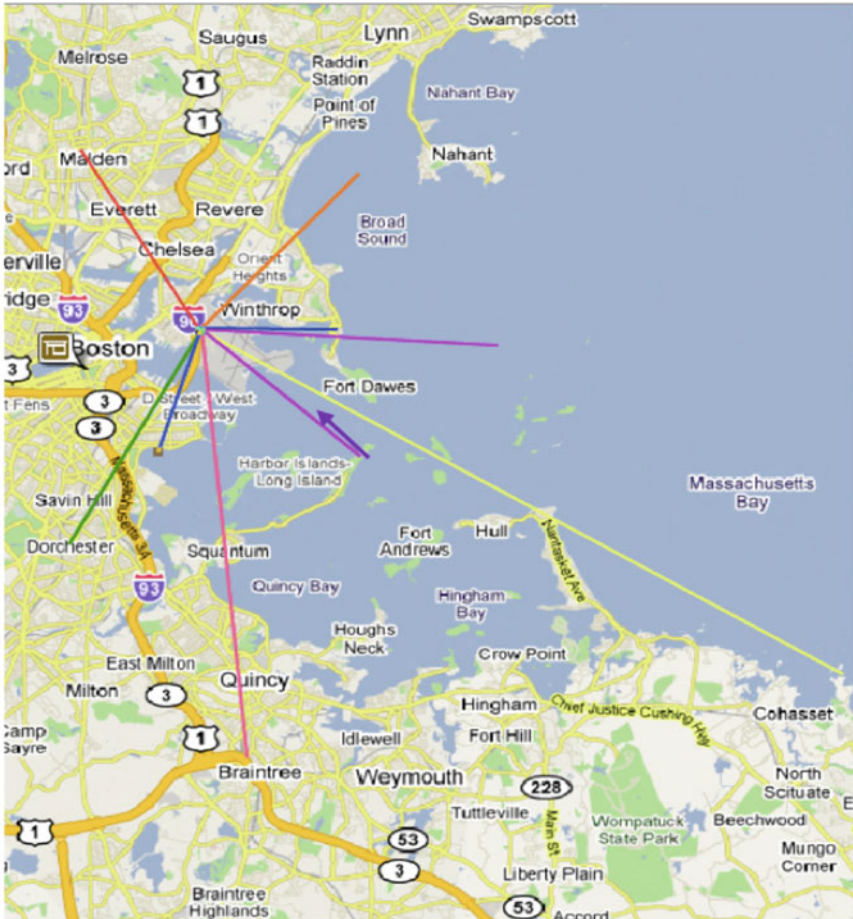
Fig. 9 Lines of sight from the Boston Logan International Airport (BOS) tower to positions in the airport region where the visual motion of moving aircraft were analyzed

visual feature was determined, a spectrum of visual velocities associated with each of the 14 feature categories could be determined. These are shown in Fig. 11 and summed to give an overall total. These spectrums of visual velocity for each of the categories of features reflect some of the physical aspects of each category. The first and last useful visual contact rates are slowest because these are in general the farthest from the tower. Visual rates during landing deceleration are high because the aircraft are generally closer to the tower yet still moving relatively fast compared to taxiing.

For the purposes of the present inventory the most important aspect of the distribution of motions is not its shape or arithmetic mean but its mode and range. As can be seen in Fig. 11, the vast majority of visual rates are less than 1 °/sec with the mode at a small fraction of a degree/sec. These visual rates are quite slow compared to those
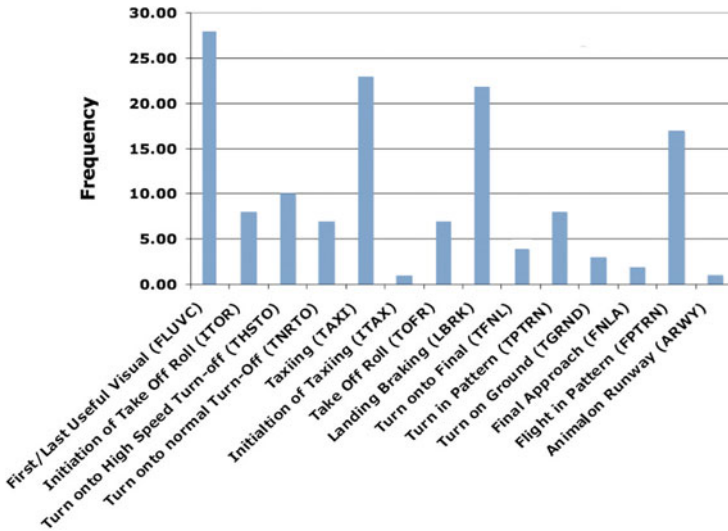
Fig. 10 Frequency of report of the use of various visual cues
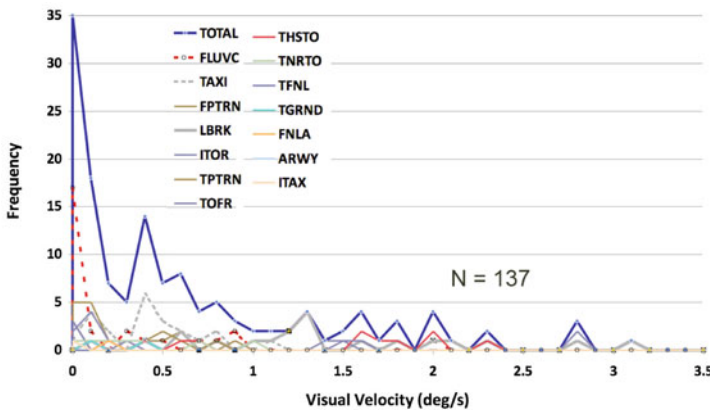
Fig. 11 Frequency distribution of rates of visual movement associated with a variety of different visual cues coming from moving aircraft. See Fig. 10 for the meaning of the letter codes of the variety visual cues identified

typically studied in visual psychophysics. If a concept of operations for a remote or virtual tower is to include visually presented targets that provide the information that controllers currently pick up from aircraft motion, the display techniques need to be able to represent this range of slow motion for visual cues that controllers currently use. It is important to note that the useful presentation of aircraft motion therefore benefits significantly from the use of very large format displays. To the extent that the display scales down visual motion due to screen size, the displayed visual rates,

which are already very slow, could well become imperceptible and require special signal processing to be operationally useful. An example of such processing could be the computational detection of the slow motion and its denotation by introduction of or changes in visible symbology. A second important caveat is that the visual rates are not seen in isolation but have a temporal context, in fact, the change in visual velocity itself can be an important cue which is identified for some visual features in Table 4 and discussed in more detail in the final section.

Table 4 provides a summary of all the visual features identified from discussions with controllers from all analyzed airports. It lists the identified visual feature, the information the feature provides the controller, and suggests some general information support characteristics that would be necessary to provide equivalent information on alternative displays that might be used in a virtual or remote tower: (1) A map-like display that could be driven by ground radar or other comparable positions information e.g. ADSB. (2) An image-like display that resembles the out-the-window view from a tower and could be driven by airport cameras or other sensors and computer graphics providing synthetic vision (Figs. 1 and 2).

A better understanding of exactly how some of these cues can be used can come from examining them quantitatively. An example of such analysis is presented below with respect to landing deceleration at SFO.

4   Deceleration During Landing at SFO

In order to analyze the deceleration of aircraft landing at SFO digital video images were recorded of the initial braking after touch down. Recordings of a wide variety of landing aircraft were made to examine a wide range of decelerations. The 45 observed and reported aircraft included 747–400's, a variety of models of 767, 757, 737, A319, A320, CRJ's, and small twin turboprops. The weather was clear with light winds from the west. The landing data from all the aircraft have been aggregated as there was no intention to make a more detailed analysis by type but rather to understand the range of visual rates and visual decelerations that would be visible from the airport tower as discussed below.

The following analysis begins to determine the magnitude of this visually sensed deceleration and how it could be used by controllers. Through this process we identify one of the dynamic visual features used in traffic control from the airport tower: the change in speed evident during a single glance a controller might make towards a decelerating landing aircraft.[5] In thinking about what specific aspects of the visual stimulus to which the controllers might be attending, it is helpful to remember that

---

[5] During normal vision, people make from 3 to 5 fixations per second (Rayner & Castelhano, 2007). However, when studying some aspect of an ATC image, fixations duration can increase but rarely grow longer than approximately 1.3 s (e.g. Remington et al., 2004). Consequently, a reasonable constraint for modeling the duration of a controller's glance would be to insure that they are 1.3 s or less.

perceptual discriminations of commonly experienced magnitudes of sensory quantities such as velocity are fairly well described by Weber's Law, which states that the just noticeable difference (JND) is a constant proportion of the quantity's magnitude. This so-called Weber fraction is roughly constant for a variety of psychophysical parameters, but under the best conditions is ~6% for changes in velocity viewed within a typical 0.5 s time period. For stimuli with random mixtures of spatial frequencies, i.e. mixtures of contours of different sizes, the JND grows to about 7.5%. Very significantly for the very slow visual velocities less than 1 °/s such as those commonly seen from the control tower for landing and departing aircraft, the JND can climb up to ~10% (McKee et al., 1986).

It is therefore important to understand that controllers may not be directly sensing the visual velocities per se even though they may claim to do so. They may, in fact, develop alternative viewing strategies allowing them to translate speed into displacement during relatively fixed time intervals, thus making the detection of unusual rates of change easier. Additionally, alternative visual cues to quantities such as deceleration could be used. For example, aircraft pitch while moving along the ground could be equally well a clue to the onset or offset of braking.

It is not so much the visual aspect of the visual information that is important as it is the fact that the information revealed by vision is relevant, real, direct, unmediated, immediate and continuous that makes it possible for the best possible anticipation of future action. This is why the visual input could be critical. Replacements for it need to capture the same predictive, informational features as suggested in Table 4.

In order to begin to analyze the visual features actually present in real landing in more detail we have initially focused on the deceleration profile of aircraft landing on the 28 Left and 28 Right runways at SFO. Controllers report that they use their sense of degree and timing of this specific deceleration to anticipate which taxiway would be needed for the aircraft to exit the active runway. Their decision is time critical during heavy runway use since landing aircraft are staggered in pairs and interleaved with departures on crossing runways 1R/1L.

We have made 15 frame/s video recordings at 1024 X 768 resolution of the braking phase of 45 aircraft landing on 28L and 28R and processed the recordings to measure changes in visual velocity. We have used a custom MatLab image processing technique that isolated the moving contours across a set of two frames and averaged them to localize the aircraft and provide their screen velocity in degrees per second. Using the viewing geometry described in Fig. 12, we have recovered the aircraft braking profile and computed the changes in its visual velocity as viewed from the control tower by re-projecting the movement, as it would have been seen from the tower. Thirty of these velocity profiles (low pass filtered with a 1 Hz cutoff) are shown in Fig. 13.

Because of the noise present in our current recording technique, we were unable to obtain velocity and acceleration values with acceptable noise levels. We were, however, able to obtain a directly recorded braking deceleration profile[6] for another

---

[6] The aircrafts deceleration was recorded just after touch-down using a arm rest stabilized iPhone in Airplane Mode running an application called Motion Data with sampling rates at 30 Hz.
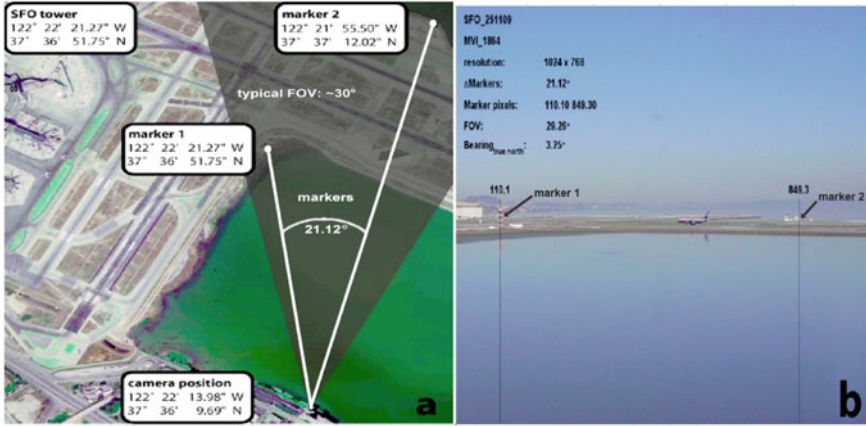
Fig. 12 Camera parameters and view at SFO. Markers at known ground positions determined from Google Earth ground images were used in combination with the known geometry of the runway to convert line of sight angles to aircraft from the camera position into position along the runway, thereafter into line of sight angles from the airport tower and thereafter into visual velocities as seen by controllers
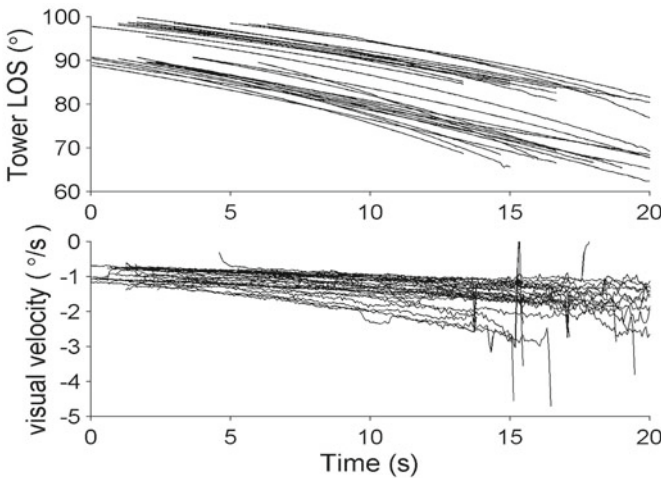
Fig. 13 Line of sight direction change and visual velocity

A319 aircraft landing on Runway 28L from the same company, comparably loaded and flying in the same wind and weather conditions as one of the aircraft we had recorded visually. Since we knew the touchdown points for these two A319 landings, we've combined the two trajectories to produce what we believe to be a fairly accurate landing profile as seen from the tower (Fig. 14).
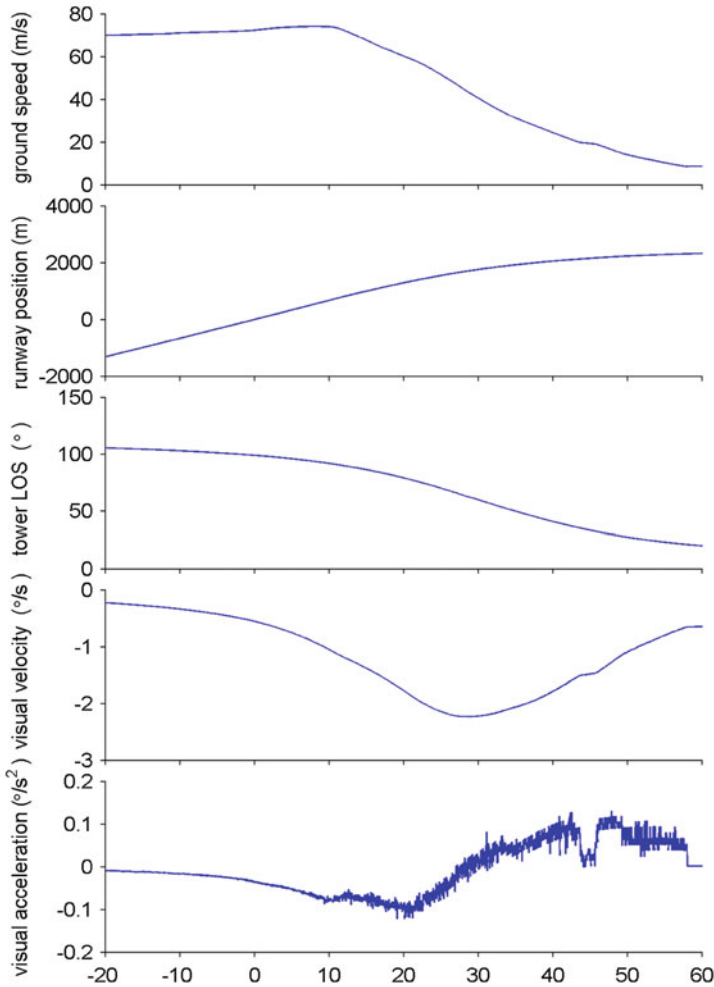
Fig. 14 Line of Sight (LOS) changes

The deceleration profile in Fig. 14 shows the aircraft approaching and passing the tower as it decelerates. In fact, during the approach the visual velocity actually increases during the deceleration because of the decreasing distance between the aircraft and the tower. It is clear from the deceleration profile that there are several phases of braking due to deployment of the thrust reversers, spoilers, and mechanical brakes and further data collection and processing needs to be done to more precisely identify these periods. However, the very smooth velocity plot in Fig. 14 (third panel from top) already shows that the amounts of velocity change in the braking within any short time window 2 s or less are well less than the ~6% usual Weber fraction for

a just noticeable difference of midrange psychophysical quantities such as perceived speed. This level is defined by convention to be that difference in a sensory quantity that can be detected correctly 75% of the time and is therefore not evidence of a very strong sensory stimulus. This observation leads to some skepticism that the controllers are detecting velocity change per se because controllers would likely wish to be more certain regarding their judgments than 75% correct. Accordingly, they may have developed a strategy to detect speed change by some other means, perhaps by comparing displacement for approximately equal time periods. Such a timing strategy might be evident in eye tracking records of controllers judging aircraft deceleration. Of particular interest will be future analyses and experiments to determine how well the controller's sense of aircraft deceleration can be maintained with airport imagery spatially degraded by pixilation and sensor noise, and temporally degraded by low sampling rate. The sampling rate issue has been addressed by research first published by Ellis et al. (2011) and more extensively analyzed in Chap. 6 of this volume.

5 Summary

1. Airport tower controllers use visual features observed during aircraft operations to provide information beyond simple detection, identification and recognition of aircraft.
2. Twenty-eight useful visual features have been identified from discussion with 24 controllers and supervisors. Some involve the static pose of the aircraft of interest but many of the most useful involve aircraft motion, especially aircraft acceleration and deceleration.
3. The visual features provide predictive or lead information regarding future aircraft position, pilot intention, and pilot airport familiarity that enable controllers to appropriately distribute their attention during operations and to anticipate possible conflicts.
4. The very slow rates of visual motion in terms of subtended visual angle suggest that the change in velocity reported by controllers is not directly sensed but must be observed by learned viewing strategies developed from tower experience.
5. Directional aircraft sounds audible in the tower are also used to assist operations.

References

Alexander, J. R., Alley, V. I., Ammerman, W.S., Fairhurst, W.S., Hostetler, C.M., Jones, G.W., & Rainey, C. L. (1989). *FAA air traffic control operations concepts, ACTC Tower controllers* (vol. VII). DOT/FAA/AP-87–01, FAA/DOT 800 Independence Ave., Washington, D.C.

Dorighi, N., & Sullivan, B. (2003). Future flight central: A revolutionary air traffic control tower simulation facility. In *Proceedings of the American institute of aeronautics and astronautics (AIAA)*.

Ellis, S. R., & Liston, D. (2010). Visual features involving motion seen from airport control towers. In *Proceedings of the 11th IFAC/IFIP/IFORS/IEA symposium on analysis, design, and evaluation of human-machine systems*, Valenciennes, France.

Ellis, S. R., & Liston, D. (2011). Static and motion-based visual features used by airport tower controllers: some implications for the design of remote or virtual towers. NASA/TM—2011–216427, NASA Ames Research Center, Moffett Field, California.

Ellis, S. R., Fürstenau, N., & Mittendorf, M. (2011) Visual discrimination of landing aircraft deceleration by tower controllers: Implications for update rate requirements for image-based virtual or remote towers. In *Proceedings of the human factors and ergonomics society* (pp. 71–75).

Fürstenau, N., Möhlenbrink, C., Rudolph, M., Schmidt, M., & Halle, W. (2008). Augmented vision videopanorama system for remote airport tower operation. In I. Grant (Ed.), *Proceedings of the 26th international congress of the aeronautical sciences.* Anchorage. ISBN 0-9533991-9-2

Gibson, J. J., Olum, P., & Rosenblatt, F. (1955). Parallax and perspective during aircraft landings. *American Journal of Psychology, 68*, 372–385.

Grunwald, A. J., & Kohn, S. (1994). Visual field information in low-altitude visual flight by line-of-sight slaved head mounted displays. *IEEE Systems Man and Cybernetics, 24*(1), 120–134.

Hannon, D., Lee, J., Geyer, T. M., Sheridan, T., Francis, M., Woods, S., & Malonson, M. (2008). Feasibility evaluation of a staffed virtual tower. *The Journal of Air Traffic Control*, 27–39

JPDO (Joint Planning and Development Office) (2007). *Concept of operations for the next generation air transportation system version 2.0* (pp. 2-27–2-36).

Kramer, L. J., Williams, S. P., Wilz, S. J., & Arthur, J. J., III. (2008). (2008) Simulation evaluation of equivalent vision technologies for aerospace operations. *SPIE Proceedings, 6557*, 69570K. https://doi.org/10.1117/12.772775

McKee, S. P., Silverman, G. H., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research, 28*(4), 609–619.

Paul, S., Zografos, K., & Hesselink, H., (2000). MANTEA final report, MANTEA/ISR DOC-D83–137-R1, TR 1036, Telematics Application Programme, (Transport/Air).

Rayner, K., & Castelhano, M. (2007). *Scholarpedia, 2*(10), 3649. https://doi.org/10.4249/scholarpedia.3649revision#54795

Remington, R. W., Lee, S. M., Ravinder, U., & Matessa, M. (2004). Observations on human performance in air traffic control operations: Preliminaries to a cognitive model. In *Proceedings of the 2004 behavioral representation in modeling and simulation (BRIMS'04).*

Ruffner, J., Deaver, D.M., and Henry, D. J. (2003) Requirements analysis for an air traffic control tower surface surveillance enhanced vision system. In *SPIE conference on enhanced and synthetic vision 2003. Orlando FL, 9–10 September 2003 SPIE proceedings* (vol. 5081, pp. 124–135). Society of Photo-Optical Instrumentation Engineers.

van Schaik, F. J., Roessingh, J. J. M., Lindqvist, G., & Fält, K (2010). Assessment of visual cues by tower controllers iwht implications for a remote tower control centre. In *Proceedings of the 11th IFAC/IFIP/IFORS/IEA symposium on analysis, design, and evaluation of human-machine systems.*

FAA Task order 7110.65R (2006). Air traffic control February 16, 2006, http://www.faa.gov/atpubs

Vollmerhausen, R. H., & Jacobs, E. (2004). The Targeting Task Performance (TTP) metric: A new model for predicting target acquisition performance. Technical Report AMSEL-NV-TR-230, U. S. Army CERDEC.

Watson, A. B., Ramirez, C. V., & Salud, E. (2009). Predicting visibility of aircraft. *PLoS ONE, 4*(5), e5594. Published online 2009 May 20. https://doi.org/10.1371/journal.pone.0005594

Werther, B., (2006) Colored Petri net based modeling of airport control process. Computation. Intelligence for Modeling, Control and Automation (CIMCA) Sydney, 0–7695–2731–0.

Detection and Recognition for Remote Tower Operations

F. J. van Schaik, J. J. M. Roessingh, G. Lindqvist, and K. Fält

Abstract Remote control of airports implies application of cameras to replace direct visual observation from airport control towers by projection of the airport and its traffic in a remote control centre. Surprisingly, hardly any literature can be found to list the required visual objects and phenomena for tower control, i.e. the visual cues that need to be seen for tower control. The composition and validation of the so-called visual cue list for tower control is the subject of this study. Tower controller task analysis was used to compose a 'long-list' of visual features. The long-list has been presented to a group of operational air traffic controllers to test the need and the circumstances to observe these visual cues. Our analysis shows that most of the visual cues are useful for operational tower control but are not strictly mandatory for applying the rules of the International Civil Aviation Organization. The requirement for visual image resolution of remote tower control is the second subject of the paper. Our analysis leads to definition of a 'short-list' of important safety-related visual objects and phenomena for tower control and the conclusion that state-of-the-art media are just able to provide the required image resolution for visual *detection* but not for *recognition.*

F. J. van Schaik · J. J. M. Roessingh (✉)
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

G. Lindqvist
Luftfartsverket LFV, Norrköping, Sweden

K. Fält
Saab Systems, Stockholm, Sweden

47

1 Introduction

In Europe, the first prototypes of remotely controlled airfields have emerged. Dedicated airfields are equipped with cameras, such that the air traffic controller (ATCO) can control the airfield from a distant virtual visual control room. The view on the airfield is displayed in real-time on a display in this room. From here, the airfield can be surveyed and the traffic movements can be controlled. This concept is particularly suitable for a group of relatively quiet airports at geographically dispersed locations, such that the control of multiple airfields can be centralized, thus making efficient use of air traffic controller resources.

The topic of this study focuses on two aspects:

1. The visual 'features' (cues, objects, phenomena) that air traffic controllers should be able to see for safety reasons in a remote tower;
2. The minimum resolution requirements for remote tower control.

A list of visual features to be seen from the control tower is of interest because it strongly influences the requirements on the surveillance cameras, the data-communication links and the display system. In this study, such a list of items, e.g. a flock of birds or debris on the runway, and the circumstances under which these items must be detected and recognized has been created. The basis for this list was established by considering the task-requirements of the air traffic tower controller.

Minimum required performance specifications are needed to determine the ability of camera surveillance and display systems to sufficiently display visual features. To see those features under widespread viewing conditions (day/night, sun/overcast, etc.) is key to the tower controllers' tasks and hence aviation safety. This means, that,

in order to detect for example birds at the runway, parameters such as the visibility range from the tower, the resolution of the image, and the contrast between object and its background must exceed certain threshold values. This paper will discuss the establishment of the resolution threshold values.

This small study was made possible in the context of the Advanced Remote Tower (ART, 2006) project. ART is a 6th Framework Program project funded by the European Commission and run under project lead by Saab AB in Sweden and the Swedish Air Traffic Control organisation LFV.

The properties of Tower Control may not be well known to readers. Therefore, the next section is included to explain the procedures and systems used in state of the art Tower Control. The focus of this contribution, i.e. the analysis of tower control visual features and resolution requirements for remote projection is found in last sections.

2 Tower Control

2.1 Basic Duties

The ICAO task definition for air traffic controllers is (ICAO, 2005a) to:

– Prevent collisions between aircraft, and on the manoeuvring area between aircraft and obstructions. The manoeuvring area is the section of the airport to be used for take off, landing and taxiing excluding aprons;
– Favour an expedite flow of traffic;

These tasks have to be performed by visual observation. The procedures change when visibility conditions change. Definition of visibility conditions can be found in ICAO (2005b). If the tower controller cannot exercise visual control over all traffic, e.g. because of fog, a special procedure called Procedural Control is applied. It means that an aircraft is cleared via radio telephony to a point at the airport, where the pilot has to report when reaching that point. Procedural Control and its safety depend largely on the quality of the VHF communication channel and the situational awareness in the cockpit. Procedural Control implies much lower throughput capacity for the airport (often only one aircraft can be moved at the time). ICAO does not specify how visual surveillance from Control Towers shall be implemented in detail. ICAO does not specify what objects or visual cues have to be seen.

2.2 Airport Radar and Surveillance Systems

Air Traffic Control in the towers of airports is thus based on visual surveillance tasks. However, for low-visibility conditions, Airport Surface Detection Equipment (ASDE) with radar screens and information from the Terminal Approach Radar

(TAR) are available at the larger airports. This kind of equipment serves the tower controllers, but controllers are allowed to take decisions based on the ASDE and TAR only in visibility condition 1 and 2 (ICAO, 2005b).

3 Analysis of Visual Features

3.1 Analysis of Tower Tasks and Visual Needs

We identified visual needs of the tower controller from our task analysis of tower operations, based on expert elicitation and task observation. The tower tasks were structured according to the time phases in ATC-handling of arriving and departing aircraft. Also general tasks (such as collecting weather information) and abnormal events (e.g. crash, bird strike, over-run of the runway) were taken into account. Our interest concerns the visual features at and around the airport which have to be surveyed as part of the task, such as specific features of aircraft (e.g. its apparent ability to land during final approach, flocks of birds, etc.). To make a more fine-grained assessment of the quality with which visual features can be viewed, different visual tasks can be distinguished:

– visual detection (you may or may not detect that an object is at a certain location);
– visual recognition (once you have detected the object, you may be able to recognise it, e.g. that an object is indeed an aircraft);
– visual identification (verify observed information, such as an aircraft at a particular position with other information, such as a flight-plan);
– visual judgment (concerns a more abstract relationship, e.g. a potential conflict between aircraft, or an unusual descent rate of aircraft).

These different visual tasks put different requirements on human visual characteristics (e.g. visual acuity see e.g. Stamford-Krause, 1997) and therefore lead to different system requirements when displaying these features in a remote tower. Moreover, visibility conditions, such as fog, may affect these visual tasks differently.

3.2 List of Visual Features

The list of visual features that was derived from the task analysis includes the following items:

| 1 | Large-size bird (e.g. goose) on the manoeuvring area or vicinity of the runway |
| 2 | One smaller bird (like a sea-gull) on the manoeuvring area or vicinity of the runway |

(continued)

1	Large-size bird (e.g. goose) on the manoeuvring area or vicinity of the runway
3	Flock of smaller-size birds (e.g. small type sea-gull) on the manoeuvring area or vicinity of the runway
4	Animal, like a deer or a dog, on the manoeuvring area/runway
5	Vehicle on the manoeuvring area
6	Aircraft entering the control zone of the tower
7	Stationary obstacles on the manoeuvring area
8	Aircraft in the circuit
9	Descent rate of aircraft
10	Aircraft undercarriage (main gear and nose wheel)
11	Aircraft position on final
12	Airspace for missed-approach/go-around
13	Foreign objects on the runway (e.g. plastic bag, pieces of metal, pieces of exploded tire)
14	Aircraft flare at landing (judgment)
15	Aircraft touch-down inside touch-down zone
16	Detect smoke or water spray from tires when touching down
17	Aircraft slowing down on runway (judgment)
18	Taxiing aircraft follows designated route
19	Water, snow or slush on runway
20	Aircraft acceleration during take-off run (judgment)
21	Aircraft lift-off (judgment)
22	Aircraft climb (judgment)
23	Cloud base
24	Clouds (type and coverage)
25	Visibility range (as judged from visibility of objects with known distance)
26	Aircraft lining up on the runway
27	Number or logo on skin of aircraft
28	Aircraft starts to move
29	Aircraft starts to turn
30	Aircraft landing lights
31	Precipitation (type) (judgment)

4   Method and Results

A questionnaire was presented to a group of seven controllers. This questionnaire referred to the visual features listed above. The controllers were asked to give their safety related experience about the maximum viewing range (detection range) for
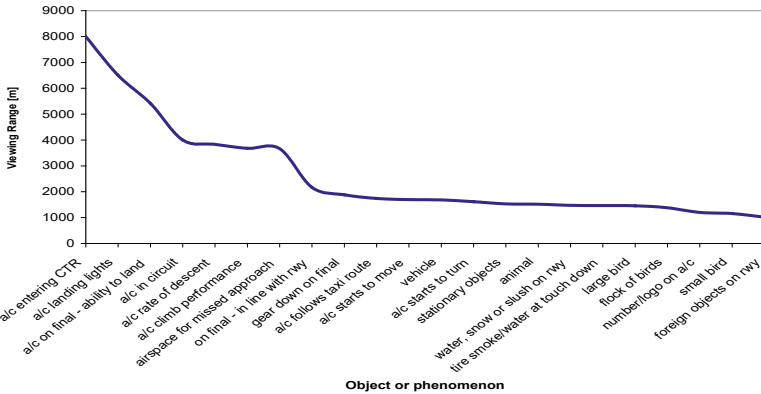
Fig. 1 Viewing range for detection at which tower controllers see objects or phenomena in good visibility and in daylight

each of the features. Subsequently, they were asked to state the importance for safety reason to detect or judge a feature on a scale from 0 (not important for safe control) to 6 (very important for safe control). If it was not required to see the feature, the controller should indicate "0".

The controllers had to indicate at which distance they would detect a feature in good visibility and daylight. Figure 1 summarizes the estimates of the seven controllers. With these distances, it should be considered that tower controllers often use binoculars to better see certain objects or phenomena, but this was not allowed for answering the questions. The middle section of the runway on the airport where these controllers are active is about 700 m away from the tower with the runway thresholds located at about 1100 and 1400 m from the tower. Note that the distances to the runway and to the runway thresholds lead to many responses close to these values. It is clearly important for safety that the runway is free of obstacles, wildlife and birds and that the monitoring of aircraft landing and take-off requires a visibility range of up to two kilometres from the tower. The "plateau" in Fig. 1 at approximately 3800 m distance is explained as a typical value for monitoring of the circuit.

It must be realised that these visual features are not equally important for the tasks of the tower controller. Therefore, it was attempted to impose an order on the list of features in accordance with the importance for the job. Figure 2 below depicts the importance to *detect* objects in order to do the job with emphasis on safety, on a scale from 0 to 6.

There was a high level of agreement among the controllers about the most important objects to detect. It should be considered that there is considerable variance in the importance-ratings of objects that are on average considered less important, sometimes depending on the way controllers interpret their job. Some controllers indicated that detection of certain objects (such as foreign objects on the runway) is not part of the tower controllers' tasks, but rather that of other airfield personnel, such as those responsible for runway inspection between flights.
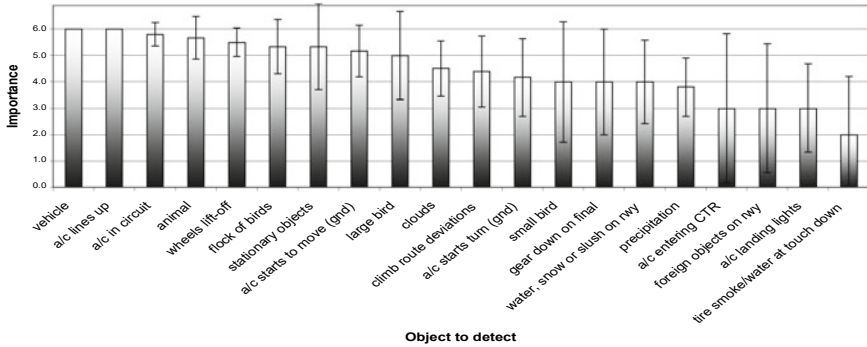
Fig. 2 Rated importance to detect objects (6 is high importance). Standard deviations (n = 7) in the rating are indicated
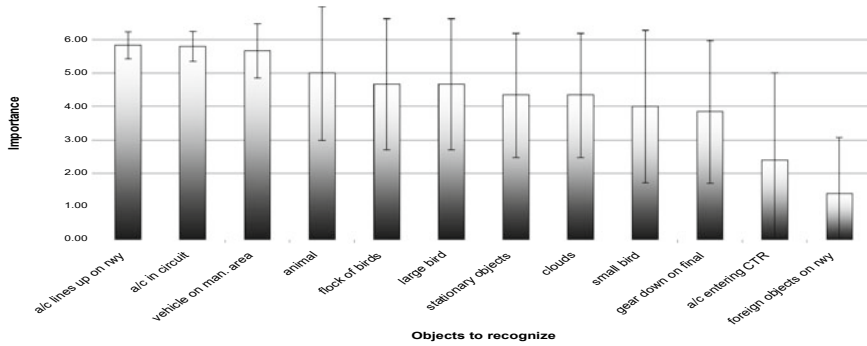
Fig. 3 Rated importance to recognise objects. Standard deviations (n = 6) in the rating are indicated

Figure 3 depicts the importance to *recognise* objects in order to do the job, on a scale from 0 to 6.

The five least important objects to detect are:

16. Tire smoke/water when touching down,
30. Aircraft landing lights,
13. Foreign objects on the runway,
6. Aircraft entering the Control Zone, and.
31. Precipitation.

The five most important objects to detect are:

5. Vehicle on the manoeuvring area (1700 m),
26. Aircraft lining up (1400 m),
8. Aircraft in circuit (4000 m),
4. Animal on the manoeuvring area and runway (1500 m),
21. Aircraft lifts up, wheels from runway (1000 m).

The five least important objects to recognise are:

13.   Foreign objects on the runway,
6.   Aircraft entering the Control Zone,
10.   Gear down on final,
2.   Small bird, and.
24.   Type of clouds.

The three most important features to recognise are:

26   Aircraft lines up on runway (1400 m),
8.   Aircraft in circuit (4000 m),
5.   Vehicle in the manoeuvring area (1700 m).

Again, there was a high level of agreement among the controllers about the three most important objects to recognise. It should be noted that recognition of objects requires a higher visual acuity (and imposes higher system requirements for displaying these objects) than detection.

Subsequently, controllers were asked to rate the importance to judge phenomena. The ratings are depicted in Fig. 4.

There is a moderate to good level of agreement among controllers about the importance of phenomena to be judged. The controllers are unanimous in the rating of importance that they must be able to visually judge the availability of (conflict free) airspace in case an aircraft has a missed approach and must make a go-around.

Finally, the controllers were asked to rate the importance of making an identification of an aircraft on the basis of logo or number visible on the skin of the aircraft. This importance was however rated as low.

Controllers were asked how they would rank the safety related importance of being able to see the features during night. Their answer was unanimous: no feature can be surveyed by visual observation in the dark unless it carries lights. Lights provide a high contrast and resolution against a dark background, making visual observation
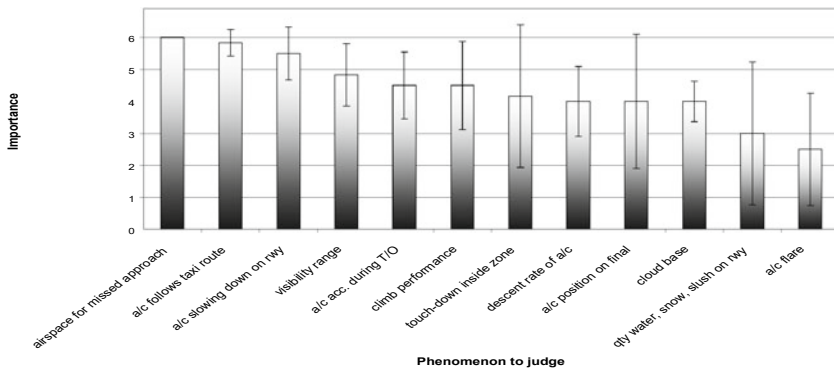

Fig. 4 Rated importance to judge phenomena (6 is high). Standard deviations (n = 7) in the rating are indicated

during night different from daylight conditions. Therefore night operations have not been further analysed in this paper.

The five least important phenomena to *judge* visually are:

26. Aircraft flare,
19. Water, snow or slush on the runway,
23. Cloud base,
11. Aircraft position on final, and.
9.   Descent rate of aircraft.

The five most important phenomena to *judge* visually are:

12. Conflict free airspace that must be available for approaching aircraft in case such an aircraft has a missed approach and must make a go-around (at approx. 3700 m maximum),
18. Whether aircraft follow or deviate from a designated taxi route (at approx. 2300 m),
17. Whether aircraft slow-down sufficiently after touch-down (at approx. 1500 m),
25. The visibility (range) from the tower, and.
20. Aircraft acceleration during take-off (at approx. 1000–1500 m).

5   Discussion and Effect on Image Resolution

This study was performed to investigate the features that benefit safety if observed under good visibility conditions during the day. A long-list of features was extracted from a tower controller task analysis. This list was presented to operational controllers in a questionnaire about the importance of these features for safe tower control. The responses to the questionnaire were ordered with respect to importance for the tower controllers' job, distinguishing between features that are important to be detected, to be recognized, or to be judged with emphasis on safety. More expensive visual systems would be able to detect the smallest objects at the largest distances and even recognize and assist in judgment but it would not be cost beneficial. Therefore we removed the features from the list that are ranked the least important either for detection, recognition or judgment. For daylight these are: (2) Smaller birds, (6) Aircraft entering the Control Zone, (9) Descent rate of aircraft, (10) Aircraft gear down, (11) Aircraft position on final, (13) Foreign objects, (14) Aircraft flare, (16) Smoke or water spray, (19) Water, snow or slush, (23) Cloud base, (24) Type of clouds, (30) Aircraft landing lights and (31) Precipitation. The reasons for controllers to find these features less important are obvious: these features are too small, too remote or not very important at all for safe control of aircraft. Off course, aircraft and vehicle lights are very important for night operations, but this investigation focused on daylight operations in good visibility. The low importance rating for item 27: Number or logo, was not expected. This result might stem from the typical traffic at a small airport with well-known users. This feature and result was therefore excluded from further analysis.

The results were translated to image resolution requirements. For the surveillance of objects at large distances, the capability of a display system to make small details visible (i.e. the resolution) is critical. The resolution depends on the camera, the addressable resolution of the graphics processor (expressed in pixels per degree visual angle in horizontal/vertical direction) and the resolution of the display (which in turn depends on such factors as pixel-size and video-bandwidth).

The size and distance of small objects determine the required limiting resolution of the image system and display. Ideally the limiting resolution of the image system should at least be equivalent to the ATCO's ability to perceive detail. Perception of detail is expressed as visual acuity (i.e. the inverse of the smallest perceptible object angular detail) in arcmin^{-1} (1 arcmin, or minute of arc, equals 1/60 deg.). ATCOs might have a minimum separable acuity between 10 and 40 arcsec-1, when tested in the laboratory (e.g. Boff & Lincoln, 1988). However, to set a minimum requirement for the 'noisy' tower environment we shall assume that the ATCO has a visual acuity of only 1 arcmin^{-1}, which is reached by 85% of the population. On this basis, we assume a limiting resolution of 60 lines per degree. This would ideally correspond to an 'addressable' resolution (that is, addressable pixels of the image generator) of 60 pixels per degree. However, to account for the loss of resolution in a system, we should divide the latter addressable resolution by 0.7 (the so called Kell-factor, Padmos & Milders, 1992). Hence, dividing the addressable resolution of 60 pixels per degree by 0.7 results in the *required* addressable resolution of 86 pixels per degree.

In this context, only the visual features for detection and recognition are contributing to the requirements. Table 1 is an inventory of remaining visual features that play a role in safe conduct of tower operations. These features are ordered according to importance in Fig. 2 for detection and Fig. 3 for recognition. For each feature, cross section area and typical size were estimated. The distance at which the features are observed comes from the data in Fig. 1.

For features for which a visual judgment is required, the resolution requirements are not specified and would need further investigation. These features are: (12) Airspace for missed approach/go-around, (15) Aircraft touch down capability, (17) Aircraft slowing down on runway, (18) Taxiing aircraft follows designated route, (20) Aircraft acceleration during take-off, (22) Aircraft climb, and (25) Visibility range. Feature (29) Aircraft starts to turn will probably be preceded by (28) Aircraft starts to move and is thus expected to bring similar requirements.

For recognition of objects, we can use the criterion that at least eight image lines overlay a recognisable object. Under optimal conditions in the tower, ATCOs may be able to recognize high-contrast features subtending a visual angle of 2 arc-minutes. This would mean that we need a limiting resolution of 240 lines per degree (343 pixels per degree). However, taking into account more realistic conditions, Padmos and Milders (1992) propose a more relaxed guideline for the required addressable resolution (in pixels per degree):

$0.14 \times$ object distance/object size (expressed in the same unit of length),

Thus, to *recognise* an object with a characteristic size of 1 m at a distance of 600 m requires 84 pixels per degree visual angle. The required visual subtended angles from

Table 1 Selected important visual features for visual tower control with their typical dimensions, observing distance from the tower and the consequences for remote projection in pixels per degree both for detection and recognition (daylight conditions, good visibility)

No.	Important visual feature	Cross section (m²)	Characteristic size (m)	Range (m)	Vis. angle (arcmin.)	Resolution for detection (pixels/deg.)	Resolution for recognition (pixels/deg.)
1.	Large-size bird	0.50	0.71	1500	1.7	51	289
3.	Flock of smaller-size birds	20.00	4.47	1400	11.1	8	43
4.	Animal	1.00	1.00	1500	2.3	37	210
5.	Vehicle on the manoeuvring area	4.00	2.00	1700	4.1	21	118
7.	Stationary obstacles	1.00	1.00	1500	2.2	38	215
8.	Aircraft in the circuit	6.00	2.45	4000	2.1	41	229
12.	Airspace for missed-approach/go-around	See text		3700	See text		
15.	Aircraft touch down inside t-down zone	See text		1500	See text		
17.	Aircraft slowing down on runway	See text		1500	See text		
18.	Taxiing aircraft follows designated	See text		1700	See text		
20.	Aircraft acceleration during take-off	See text		1500	See text		
21.	Aircraft wheels lift-off	1.00	1.00	1000	3.4	25	140
22.	Aircraft climb	See text		3700	See text		
25.	Visibility range	See text					

(continued)

Table 1 (continued)

No.	Important visual feature	Cross section (m²)	Characteristic size (m)	Range (m)	Vis. angle (arcmin.)	Resolution for detection (pixels/deg.)	Resolution for recognition (pixels/deg.)
26.	Aircraft lining up on the runway	6.00	2.45	1400	6.0	14	80
27.	Number or logo on skin of aircraft	See text		1200	See text		
28.	Aircraft starts to move	1.00	1.00	1700	2.3	37	210
29.	Aircraft starts to turn	See text		1700	See text		

the controller responses have been translated in the last two columns of Table 1 in to the calculated resolution for detection and recognition using our literature references given above. When traffic is labelled, such that detection and recognition of traffic is facilitated, the requirements listed in Table 1 may be lowered.

The visually most demanding task is to recognise a large bird at 1500 m, which is on the edge of what can be detected and recognised with unassisted eyesight. However, tower controllers will expectedly use binoculars if they detect distant objects or movements, for which an equivalent camera/display system, such a separate Pan-Tilt-Zoom (PTZ-) camera may be used in remote tower operations.

Most of the other features will be viewed with a minimum subtended visual angle of 2–6 min. The resolution required for detection and recognition of a flock of smaller-size birds would need additional investigation, since it will obviously depend on the actual size of each bird, and the number and distribution of birds. Typical camera and projection systems provide about 30–40 pixels per degree viewing angle. From the table it can be concluded that that resolution is sufficient to provide *detection* of the most important features for visual tower control. For detection of small features, for example, when a steady aircraft start to move (nr. 28), binoculars (zoom camera) and (automatic) tracking would be required. If features have to be recognised, five to six times the number of pixels per degree that are needed for detection will be needed. Video image enhancement techniques may be required to achieve resolution and contrast for recognition, therewith providing cost beneficial solutions to the requirements.

This survey was our first attempt to derive optical requirements for remote tower operations. It is planned to include more air traffic controllers (including military air traffic controllers) in the survey in order to fine-tune the analysis. Further analysis will also address minimum contrast requirements for remote tower control. For the ability to detect objects in a complex scene, contrast sensitivity of the human and hence image contrast in the projected image is at least as important as visual acuity/image resolution (e.g. Streid, 2007).

References

ART. (2006). Advanced Remote Tower project, 6th Frame Work Program project funded by the European Commission, DGTREN. Technical Annex to the Contract TREN/07/FP6AE/S07.73580/037179.

Boff, K. R., & Lincoln, J. E. (1988). *Engineering data compendium—Human perception and performance*. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio.

ICAO. (2004). Doc 9830—Advanced-Surface Movement Guidance and Control System Manual, First edition 2004. International Civil Aviation Organization.

ICAO. (2005a). Doc 9328: Manual of Runway Visual Range Observing and Reporting Practices, Third edition 2005. International Civil Aviation Organization.

ICAO. (2005b). Doc 4444 Air Traffic Management PANS-ATM, Edition November 2005. International Civil Aviation Organization.

ICAO. (2006). Annex 2 - Rules of the Air, Edition August 2006. International Civil Aviation Organization.

Padmos, P., & Milders, M. V. (1992). Quality criteria for simulator images, a literature review. *Human Factors, 34*(6), 727–748.

Stamford Krause, S. (1997). Collision avoidance must go beyond "see and avoid" to "search and detect". Flight Safety Digest, issue May 1997, Flight Safety Foundation.

Streid. (2007). Eye limited resolution displays: on the cusp? In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Remote Tower Research in the United States

Vilas Nene

Abstract The United States (U.S.) Federal Aviation Administration (FAA) has been conducting remote tower research since 2006. The focus of this effort has changed multiple times since the inception of the research. As a result, a number of different remote tower concepts were developed and validated to varying degrees. These included the Staffed NextGen Tower concept for all sized airports, the Select Services Concept for non-towered airports, and the Full Services Concept for non-towered airports. In 2013, the direction of the research changed again and the FAA began to work on a Colorado initiative that envisions the use of their Wide Area Multilateration (WAM) for improving services at the non-towered ski airports in Colorado. Currently the FAA is also initiating the evaluation of the camera-based concept at Leesburg, Virginia. All these efforts are described in the following sections.

1   Background

The early remote tower work in the United States (U.S.) began in response to the concept of the Next Generation Air Transportation System (NextGen) developed by the Joint Planning and Development Office (JPDO), which was established in 2003 by the White House to manage the partnerships between the private-sector, academia, and government agencies required to realize the NextGen vision. Some high-end estimates by JPDO indicated that the passenger enplanements would more than double and total aircraft operations would triple by the year 2025 in comparison to traffic in 2005 (FAA. Long-Range Aerospace Forecasts; Fiscal Years 2020–2025 and 2030, 2006). Historically any such increases in demand were addressed by constructing new runways and new Air Traffic Control Tower (ATCT) facilities across the National Airspace System (NAS). In addition to the projected increases in passenger enplanements and aircraft operations, the Federal Aviation Administration (FAA) was also facing the need for replacing a significant number of aging ATCT facilities that were rapidly approaching the end of their useful life. At the same time,

V. Nene (✉)

The MITRE Company, McLean, VA, USA

e-mail: vnene@mitre.org; vilasnene@gmail.com

the cost of new tower construction was escalating rapidly and the requirements for facility construction and refurbishment were exceeding the available budgets.

Therefore, in 2006 the FAA began the development of a concept of NextGen Tower (NT) facilities where airport traffic control services could be provided remotely without the need for constructing a tower on the airport property. Initially the FAA envisioned two types of NT facilities:

- A Staffed NextGen Tower (SNT), a ground-level facility from where controllers would provide full air traffic management (ATM) services to flights in and out of one or more airports, and
- An Automated NextGen Tower (ANT), a fully automated ground-level facility that would provide a limited number of basic ATM services without any human participation. The ANT automation systems would use available traffic, surveillance, and weather data to generate appropriate sequences, clearances, and advisories for transmission to the aircraft via synthetic voice and/or data link.

It was envisioned that an SNT would provide services for the presently towered airports, and an ANT would provide services mainly to small non-towered airports and possibly to towered airports during off-peak hours when tower services are terminated. Initially, however, the FAA focused its effort on the development of the SNT concept.

2 Present Use of the Out-the-Window (OTW) View

A tower controller typically uses his/her eyes and ears to maintain situational awareness (SA) of surface operations, arrivals and departures, and operations in the vicinity of the airport. In fact, the FAA requires (FAA, 2008a) the Local controllers to visually scan runways to the maximum extent possible and the Ground controllers to assist the Local controllers in visually scanning runways, especially when runways are in close proximity to other movement areas. The controllers also use visual scanning to observe changing weather between manual and automated weather updates, to check for animals, birds, and foreign objects and debris (FOD) on runways and taxiways, and to check for any emergencies on runways such as engine fire, smoking, blown tires, and other hazardous situations. In reduced visibility, however, the OTW view is of limited use; the controllers then supplement the OTW view by soliciting pilot position reports.

Based on the concept developed by the FAA in 2008 (FAA, 2008b), air traffic controllers in an SNT facility would not have an out the window (OTW) view from a tower cab. They would, therefore, need to obtain all the information they get from the OTW view from some other sources. In the past there have been several studies to examine controller use of OTW view in controlling traffic (Cardosi & Yost, 2001; Pinska & Bourgois, 2005). In support of the SNT related work, the FAA sponsored two investigations of controller activities in the tower to specifically identify the

different types of data they get from the OTW view and to understand the potential impact of removing this source of information.

2.1 SNT Walkthrough (FAA/George Mason University, Fairfax, VA, 2009)

A team of twelve controllers participated in a walkthrough study conducted by a team of researchers led by George Mason University (GMU), in Fairfax, VA. These controllers had experience in controlling traffic from ATCTs as well as from Terminal Radar Approach Control (TRACON) facilities. These controllers were first presented with the general concept of SNT operations. Then the controllers filled out questionnaires related to the use of specific information that is required to conduct Ground and Local control operations; they also indicated how that information was obtained during three visibility conditions—daytime good visibility, nighttime good visibility, and low visibility. The controllers were then presented with a list of potential off-nominal conditions and were asked to describe how their need for information would change under these conditions.

The controller responses indicated that the information they need can be grouped into three general categories:

1. Information about aircraft: type, size, capabilities, and related air carrier.
2. Information about aircraft location: location, orientation, distance from other aircraft, and their position in the sequence.
3. Information on constraints in effect: status of runways and taxiways, traffic management initiatives (TMIs) such as ground delays and stops, and flights requiring individual requests for approval (APREQ) from the overlying facility.

The above information enables controllers to predict and/or confirm an aircraft's route, threshold crossing, touch-down point, exit off the runway, and take-off point; in turn, these predictions help controllers in making tactical control decisions.

Most large airports in the U.S. and abroad, with high complexity and volumes of traffic, use integrated surface surveillance systems consisting of a surface movement radar (SMR), multilateration (MLAT) system, and Automatic Dependent Surveillance-broadcast (ADS-B). Such surveillance provides a two-dimensional (2-D) display of surface operations that controllers use to augment the OTW view. However, at all other towered airports the controllers exclusively use the OTW view to obtain the SA of surface operations. In the SNT operations, therefore, in the absence of the OTW view, controllers would need to find alternative ways of obtaining the necessary information that they would normally obtain via the OTW view. The SNT walkthrough asked the controllers to provide guidance on the characteristics of any future display that may be useful for this purpose. The controllers thought that an intuitive, immersive, adaptive three-dimensional (3-D) representation of air traffic would be more useful than the current 2-D displays.

2.2 *Verbal Protocol Analysis (Boehm-Davis, 2010)*

In a verbal protocol analysis (VPA) verbal data about cognitive processing is collected and analyzed in an effort to understand how humans perform certain tasks. In support of the SNT concept development, GMU researchers conducted such a VPA to evaluate towered operations. As a part of the VPA, a number of controllers used the ATCT simulator at the Ronald Reagan Washington National Airport (KDCA) in Washington, District of Columbia (D.C.) to control simulated traffic under varying scenarios. The scenarios covered many visibility conditions: daytime good and poor visibility; and Category (CAT) IIIC conditions. The scenarios also covered traffic volumes of 50–85% of the normal operations. Both the Local and Ground controllers "talked aloud" to provide a verbal protocol while they performed their tasks. A video recorder placed behind the controllers recorded the audio communications as well as various gestures made by the controllers.

The video tapes were analyzed and every utterance was examined to determine (i) where the specific information was coming from (OTW view, head-down display, or auditory); (ii) what information the controllers were looking at (scanning without a specific object in mind; specific object such as runway, gate, weather etc.); and (iii) why they were using the information (monitoring for potential conflict; aircraft compliance with control instructions, and SA).

As expected, the findings from the walkthrough indicate that tower controllers consider direct visual surveillance via the OTW view as an indispensable element of achieving full SA. Consequently, they use displays and other cues only when the OTW view is compromised under poor visibility. Although all controllers tend to use the information for all above purposes, the Local controllers tend to use it more for detecting potential conflicts while the Ground controllers tend to use it more for ensuring compliance.

The controllers like to know, under all visibility conditions, information about aircraft movements on taxiways and about all departing and arriving aircraft. Another common activity observed during the walkthrough was the act of scanning runways and taxiways. Under CAT IIIC simulated conditions, controllers were scanning surveillance displays and looking for speed and altitude information about aircraft. During the high traffic level condition there were fewer pauses between the control instructions.

As expected, both the Walkthrough and Verbal Protocol studies essentially confirmed the conclusions of the past studies (Cardosi & Yost, 2001; Pinska & Bourgois, 2005), which is that controllers seek various information elements in order to maintain SA of surface and surrounding traffic; to avoid aircraft-to-aircraft and aircraft-to-airspace conflicts and other potentially hazardous movements; and to monitor and confirm aircraft compliance with control instructions.

3 The Operational Concept for SNT (FAA, 2008b)

Tower controllers today depend on visual surveillance via the OTW view; on supplemental information provided by terminal radar, weather sensors, and displays; on flight data provided on the flight strips; and on air-ground verbal communications. At large airports, they also depend on surface surveillance information presented to them on a 2-D display.

The SNT operational concept (FAA, 2008b), as a first step, did not include camera surveillance but envisioned providing the information required by the controllers for controlling nominal tower operations with only cooperative surveillance. It was decided to visit at a later date the information needs of tower controllers during off-nominal events such as, for example, aircraft emergencies and aircraft non-conformance. It was felt that possibly some of the responsibility of off-nominal operations may be reassigned to airport personnel, and if necessary, the use of camera surveillance could be added to this concept as the concept matured.

3.1 Assumptions Related to the SNT Concept

The SNT concept was based on the following assumptions:

- There would be a fundamental shift in the roles and responsibilities of the Air Traffic Management (ATM) service provider, aircraft/pilot, and flight operations center personnel. Presently the service provider has a proportionately much larger influence on ATM decisions. Although the aircraft and the flight operations center share a lot of information between them, they share little information with the service provider. In the future, these three entities would share all the information in a net-centric environment and influence, somewhat equally, the ATM decisions.
- Some form of a cooperative surface surveillance system would be present at the airport. The status of all ground movement of aircraft and other vehicles would be presented on a two-dimensional (2-D) display for ATM personnel. Such a display would also present the necessary weather information.
- All aircraft operating in and out of *large* SNT-serviced airports would be equipped with a transponder. A significant number of aircraft may also carry multi-functional flight deck displays and data link equipment; these aircraft would be capable of providing aircraft-derived data (ADD) requested by the ATM system. However, the SNT concept would continue to accommodate aircraft unequipped with the multi-function display and those that could not provide ADD. Aircraft without a transponder or with a failed transponder would be accommodated at large airports only under emergency conditions.
- A significant majority of aircraft operating in and out of *small* SNT-serviced airports would be expected to be equipped with a transponder; however, the SNT concept would accommodate aircraft without a transponder. This would

be accomplished by having either some form of non-cooperative surveillance or
visual surveillance using digital cameras or other available technologies.

- Airports would be required to implement perimeter security to minimize
  runway/taxiway incursion by animals, pedestrians, and other unauthorized vehi-
  cles. Airports would also be required to minimize the presence of FOD on runways
  and taxiways.
- Necessary Decision Support Tools (DSTs) would be available to ATM personnel
  for minimizing the capacity/demand imbalances, balancing demand loads across
  runways, implementing traffic management initiatives, and generally improving
  the airport operational efficiency.
- Airports would be certified for SNT service, aircraft would be certified for oper-
  ating in and out of SNT-serviced airports, and ATM personnel would be certified
  for providing services at these airports.
- The envisioned integrated tower display for presenting weather and traffic would
  be certified for use by ATM personnel for providing separation services in the
  absence of the view from the cab window.

3.2 Substituting for the Window View

In a remote ground-level SNT facility there would be no tower cab and no OTW
view. Consequently, ATM personnel must be provided with an alternative means of
obtaining the information they depend on and get from the OTW view. Some form of
cooperative surface surveillance can provide locations and velocities of aircraft and
ground vehicles on the apron and movement areas including the runways. Conse-
quently, such surveillance can help the controller to determine aircraft conformance
with ATM instructions. However, such surveillance would be of minimal use with
aircraft emergencies and with unequipped aircraft. Information on a number of emer-
gencies related to aircraft could be obtained from the aircraft itself; this would require
aircraft to be equipped with a form of data link capability, although presumably such
information could be sent via voice as well. If a significant population of aircraft could
not provide ADD, and if one must accommodate aircraft without a transponder, one
would then need to provide some form of non-cooperative surveillance or camera
surveillance for the terminal area and for the airport surface. However, it was not
clear at the time if such cameras could provide sufficient fidelity and update rate for
use at airports with high volume and complexity of traffic. Consequently, various
options for substituting the window view would have to be examined in detail in
laboratory and operational tests before determining an acceptable NextGen Tower
configuration.

The concept of such a SNT facility would represent a paradigm shift and would
require a fundamental change of the present ATM culture. In addition, changes in
policy and procedures would require a partnership between the service provider,
aircraft operators, and airport operators. Some roles and responsibilities of the ATC
personnel may have to be shifted to airport and/or aircraft operator personnel located

at the airport. The level and the extent of such changes would depend on the surveillance infrastructure at the airport, ATM automation system capabilities, capabilities of the aircraft operating in and out of the airport, and the volume and complexity of the traffic at the airport being served.

There are a number of different options for substituting the OTW view. For example, relevant weather could be displayed on the SNT display and the time period between weather updates could be reduced. Also, digital cameras could be used to scan the weather between updates; such cameras are already in use to monitor weather in Alaska. Digital cameras could also be used to scan runways and taxiways or airport authorities could implement sensors for scanning these areas. Although digital cameras did not provide sufficient fidelity and update rates at the time for use at large airports, it was assumed they could be quite suitable for low and medium traffic airports. Also, performance of digital cameras was assumed to continue to improve in the future. Another option that was considered for the concept was for the aircraft to provide emergency-related aircraft status information to SNT automation. Such ADD may assist controllers in identifying emergencies on runways and taxiways; however, it was assumed that not all aircraft would be equipped to provide ADD. Also, some aircraft might not have onboard sensors to detect certain events such as wheel assembly fire and tire blowout; therefore, certain responsibility for handling emergencies would possibly have to be shared between service provider and airport/aircraft operators.

Clearly, all the above options for substituting OTW view would have to be validated by rigorous safety analysis and extended laboratory and operational trials; only then could these options be considered feasible.

3.3 2-D Surveillance Display for the Controllers

The proposed SNT concept envisioned the use of a large 2-D display that would present surveillance information (surface and surrounding airspace) and weather information integrated together in appropriate formats. Surveillance may consist of non-cooperative surveillance such as radar, inductive loops, and magnetometers; cooperative surveillance such as radar, MLAT, and ADS-B; and/or, if necessary, camera surveillance with or without image enhancement. The proposed concept did not specify or require any specific elements of surveillance. Controllers presently use 2-D displays and the altitude/speed information to create in their minds a 3-D picture of the airspace and the traffic within it. Such a mental 3-D picture is critical in maintaining the full SA required for controlling traffic. Researchers have long felt that the controllers may be able to avoid these mental computations if the traffic is presented to them on some form of a 3-D display. A number of 3-D and four-dimensional (4-D) presentations have been developed and examined by researchers around the world (Nene, 2007). A virtual tower cab mimicking an OTW view has also been developed (Nene, 2007). These studies have shown that these technologies must address several technical issues before these ideas can be tested in field trials.

Also, it is difficult to tell if these types of displays will be deemed acceptable within the present controller community. Consequently, this SNT Concept was based on the conventional 2-D display.

3.4   Decision Support Tools

Increasing airport capacity and efficient arrival/departure management were two important elements within the SNT concept. It was envisioned that ATM personnel would use different DST capabilities in achieving these two goals. Clearly, the list of these capabilities would evolve as the SNT technology evolved. Some candidate SNT DST capabilities were defined for, but not limited to, providing support for the following activities:

- Deciding the most efficient airport configuration for a given set of traffic and weather patterns
- Early planning of runway/taxiway assignments based on projected runway loading, surface congestion, user runway and gate preferences and other relevant factors
- Arrival and departure management for accommodating all traffic management constraints resulting from anticipated weather conditions and resource loading
- Providing information about airport weather conditions, runway visual range, surface conditions, braking action, current precipitation, runway availability, wake turbulence, and wind shear advisories to the aircraft via data link
- Providing pre-departure clearances, taxi clearances, and any revisions to clearances to the aircraft via data link
- Providing a coded taxi route to the aircraft via data link
- Monitoring aircraft conformance with ATM instructions and appropriately alerting aircraft and ATM personnel
- Updating flight trajectories based on rerouting, ground holding, and other TMIs
- Generally creating a common situational awareness between ATM personal, ramp personnel, airport operators, and flight operations personnel which should greatly improve the efficiency of all surface operations including de-icing operations.

Some of the functionality provided by automation within the SNT concept would be more effective if it could receive ADD from the aircraft. If, however, some aircraft were not equipped to provide ADD, it would reduce the performance of the tools and such aircraft may not receive services in full. The reduction in the performance of the tools would depend on the proportion of the unequipped aircraft, and in turn would reduce the overall throughput and efficiency of airport operations. The effectiveness of these DSTs would clearly depend on the quality and accuracy of the data used by the DSTs. If DST inputs exhibit large variability, the controllers may find these tools unreliable and unusable. Consequently, the tools themselves would need extensive validation before their implementation within NextGen tower automation systems.

3.5   Use of Aircraft Derived Data (ADD)

It has been mentioned earlier that aircraft can provide some of the data that controllers presently obtain from visual surveillance. It was expected that in the future the aircraft could provide an extensive amount of data that would improve the overall performance of the ATM system at the NAS level. Any available media of transmission could be used for such data transfer. A number of candidate data items for such data transfer had been identified at the time. These included, among others: aircraft identification (ID), route in the onboard flight management system (FMS), taxi route, aircraft braking performance, distance required for landing, and information on aircraft emergencies if applicable. It may not be necessary to obtain all of this information directly from the aircraft; some may be obtained from flight operations center handling the aircraft operations. Furthermore, it was assumed that this list would have evolved as the SNT technology matured over the coming years.

It must be noted here that the ADD technology would be useful well beyond its use in remote towers. However, presently there are no common formatting or communications standards for providing ADD to the ground ATM system and existing air-ground communications links may have to address bandwidth issues for such use. Consequently, if the use of ADD is envisioned going forward, the FAA may have to develop some commonly acceptable frameworks for its distribution and use by future ATM systems.

3.6   SNT Configurations

U.S. airports exhibit a large variability in volume and complexity of traffic. As an example, in 2009, the peak hourly traffic varied from a high of 216 operations at Hartsfield-Jackson Atlanta International Airport, Georgia (KATL) to a low of 15 operations at Branson Airport, Missouri (KBBG) (Nene, 2009). It would, therefore, be necessary to tailor the SNT configuration to fit the operational needs of the airport. The SNT concept proposed the following three configurations:

a. Display-only Configuration
    The display-only configuration would be suitable for implementation at small airports with low volume and complexity of traffic. An aircraft unequipped with a transponder would not be detected by cooperative surveillance. Consequently, either a mandate would be required for all aircraft to carry a transponder, or some form of non-cooperative surveillance or visual surveillance by digital camera(s) would need to be implemented at these airports for accommodating unequipped aircraft. It was envisioned that this configuration may also be useful in providing tower-like services at presently non-towered airports (NTAs), or in continuing tower services at towered airports when the tower facility is not in use.

b. Display + DST Configuration

This configuration would be similar to the display-only configuration, except that the automation system would provide some limited DST capability. The DST capability could include, for example, early runway/taxiway assignments, pre-departure clearance, and coded taxi route delivery. This configuration was assumed to be mostly applicable to medium sized airports.

c. Display + DST + ADD Configuration

This configuration would include a surface surveillance display, a variety of DST capabilities, and the availability of ADD. Clearly this configuration would be considered for implementation only at large airports with high volume and complexity of traffic. Although aircraft will not be required to carry ADS-B equipage, all aircraft operating in and out of these candidate airports would be expected to carry ADS-B-Out avionics; some aircraft may also carry ADS-B-In avionics. The airports would also have some form of non-cooperative surveillance for accommodating aircraft experiencing equipment failure.

4 Summary of the SNT Concept

The SNT Concept was developed for a range of airports. It was envisioned to support all ATC functions presently performed by tower personnel under nominal operations. Table 1 describes how various functions are currently performed by different tower personnel and how they would be performed in an SNT facility.

5 Assessment of the SNT Concept

The SNT concept was the first attempt to comprehensively understand potential remote control of airport operations without the benefit of an OTW view from the tower cab. It also attempted to provide all tower services without the use of a digital camera, exclusively depending on the use of co-operative surveillance on the airport surface. In addition, the concept required that either all ground vehicles operating on the movement area and all aircraft operating in and out of the airport would be required to carry a transponder, or non-cooperative surveillance would be implemented at the airport. Furthermore, the concept was envisioned to be applicable for all sizes of airports.

There were clearly a number of issues that needed to be addressed before the concept could proceed towards implementation, such as:

- Would the concept be able to handle off-nominal operations such as, for example, engine fire during take-off and unauthorized person or vehicle on the runway?
- Would the concept assure operational safety?
- Would the controllers accept the absence of both the OTW view and visual surveillance by a camera?

Table 1 Differences between present tower operations and SNT operations

Tower personnel	Functions that would be different	Present tower operation	SNT operations
Flight data controller	Weather sensing	• Automated weather updates with the use of variety of sensors are available at many airports • Controllers visually monitor changing weather between automatic updates	• Weather information would have to be updated at a faster rate than the present rate • Use of a digital camera may also be an option
Ground controller	• Push back into movement area • Taxi instruction	• SA is achieved mainly by visual surveillance via the OTW view • Limited SA is possible via ground surveillance systems if it covers the necessary gate areas and all taxiways and runways	• SA would be mainly via secondary surveillance that will cover the necessary gate areas and all taxiways and runways • Either all ground vehicles operating on the movement area would be required to carry a transponder or primary surveillance would be provided
Local controller	• Take-off clearance • Runway obstruction alerts • Separation assurance	• SA based primarily on visual surveillance via OTW view • At some airports a surface surveillance system is available for supplemental SA • The surveillance system detects a limited class of runway obstructions and provides appropriate alarms • Automated alerts are issued to cover a limited number of potential runway incursion scenarios	• Secondary surveillance would be available at the airport to cover all runways • Primary surveillance would be provided to accommodate unequipped aircraft • Automated alerting would be available at large airports

Table 1 (continued)

Tower personnel	Functions that would be different	Present tower operation	SNT operations
Traffic Management Coordinator (TMC) or Tower Supervisor (TS)	Traffic flow synchronization	• Surface congestion and gridlock conditions are noticeable via the OTW view and surveillance display if available • Flight plans, radar data, and applicable TMIs are available to TMCs and TSs	• Surface congestion and gridlock would be noticeable on the surveillance display • Flight plan, radar data, and applicable TMIs would be available to SNT personnel • DSTs would be available at some airports that would assist TMCs/TSs in implementing the TMIs

- If camera surveillance was found to be necessary, what images should be presented to the controllers and how?
- Would the passengers accept remote control of traffic in and out of large airports?

In an effort to examine some of the above issues, the FAA conducted a number of studies aimed at answering the following questions: how would controllers remotely respond to off-nominal events at the airport, what are the safety impacts of SNT operations, and what are the different ways to effectively present camera images to the controllers. These studies and related results are presented in the following sections.

5.1   Off-Nominal Events in SNT Operations (Nene, 2009)

ATCT controllers must be able respond to a number of off-nominal events such as, for example, aircraft non-conformance, failure of aircraft systems, and failure of ATC systems. Under present tower operations, controllers depend on the OTW view in responding to these off-nominal events. A number of off-nominal events typically occurring at airports were first identified (Nene, 2009) and possible ways in which an SNT controller would respond to these events in the absence of the OTW view were examined. The results of this study are summarized in Table 2. The study indicated that SNT controllers would be able to use available surface and terminal surveillance to detect and respond to a number of off-nominal events occurring at the airport, even without the presence of the OTW view. Under certain scenarios, it was assumed that visual surveillance with a pan-tilt-zoom (PTZ) camera would be useful although not required because of the availability of alternate means in dealing with the off-nominal

Table 2 Response to off-nominal events in the SNT environment (Nene, 2009)

Number	Event	Role of camera surveillance in the response of SNT controller		
		Not required	Would be useful but alternatives available	Required
1	Aircraft non-conformance		Monitor via surveillance display with or without automated taxi conformance monitoring and alerting; initiate appropriate action to remedy the situation	
2	Aircraft altitude falls below the minimum safe value		Controller would use radar position in relation to known terrain or obstruction to validate alerts issued by the automation system	
3	Potential collision between aircraft under tower control		• Automation system would issue conflict alert based on terminal area and surface surveillance information • Traffic and Collision Avoidance System (TCAS) would also alert the flight crew	
4	Aircraft in-flight system failure			Controller must use pan-tilt-zoom (PTZ) camera to observe and confirm aircraft problems with landing gear or flap control

Table 2 (continued)

Number	Event	Role of camera surveillance in the response of SNT controller		
		Not required	Would be useful but alternatives available	Required
5	Aircraft configuration problems such as an non-extended landing gear			Controller must use PTZ camera to observe and confirm aircraft problems with landing gear
6	Emergency during take-off			Controller must use PTZ camera to monitor and confirm emergency
7	Aircraft accident on the surface			Since surface surveillance cannot detect all accidents, controllers must use PTZ camera to detect and confirm accident on the surface
8	Aircraft accident within the vicinity of the airport		Controllers may use radar surveillance, reports by other pilots, and reports by airport personnel to know of the accident and initiate necessary response	

(continued)

events. However, there still would be a number of off-nominal events such as aircraft system failures, emergencies or accidents on the airport surface, and unauthorized presence on the airport surface of person or vehicle, especially with nefarious intent, that controllers would not be able to respond effectively without camera surveillance. As a result, this study concluded that if SNT has to provide all the functionality of the present towers, visual surveillance by digital camera(s) would be necessary.

5.2 SNT Safety Impact Assessment (Cheng, 2010; Colavito & Nene, 2010)

As a part of an overall effort to validate the SNT concept, a preliminary safety impact assessment of SNT operation was conducted (Colavito & Nene, 2010). This was not

Table 2 (continued)

Number	Event	Role of camera surveillance in the response of SNT controller		
		Not required	Would be useful but alternatives available	Required
9	Primary surface surveillance radar out of service			Although cooperative surface surveillance and terminal radar would continue their coverage, detecting non-cooperative targets, possibly with nefarious intent, would require the use of PTZ camera
10	Unauthorized person or vehicle on the airport surface			Detecting an unauthorized person would require the use of a PTZ camera
11	There is a need to move an aircraft on the ground to a designated area	Controllers can use surface surveillance display to relocate aircraft due to bomb threats, hazardous cargo, or any other reason		
12	There is a need to expedite the landing of an aircraft	Controllers can use surface surveillance and terminal surveillance displays to expedite any landing		

a full-fledged operational safety assessment of SNT operations as defined by the FAA's Safety Management System process that would be necessary as part of a formal acquisition program for an SNT facility.

In support of the safety impact assessment, a formal functional analysis of SNT operations was performed (Cheng, 2010). A number of specific functions were first identified as being necessary to be able to provide SNT services. These functions were then decomposed into sub-functions and were organized into a hierarchy of functions. At every level, functional flow diagrams and N^2 interface diagrams were then developed for the SNT functions; the N^2 diagrams were drawn to identify and represent all the functional interfaces within the system. The eleven first level functions and the related flow diagram are presented in Fig. 1. The first level N^2
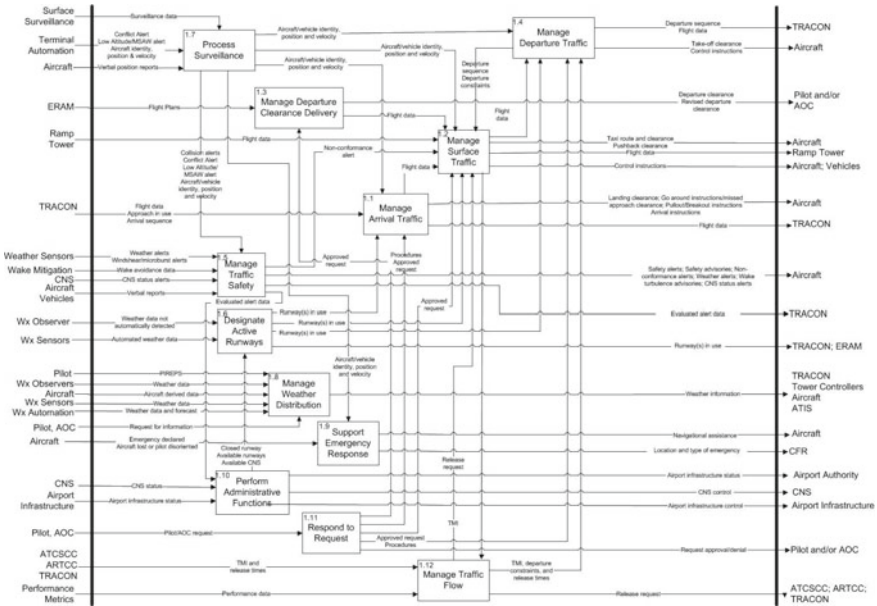
Fig. 1 First level functional block diagram of SNT operations (Cheng & Nene, 2009)

diagram is illustrated in Fig. 2. The functional analysis did not address the allocation of functions to the physical elements of the SNT system.

The functional analysis formed the basis of the safety impact analysis. Each function was evaluated in the present tower environment and in the proposed SNT environment. Potential safety impacts of the difference between the two environments was then identified. The SNT functions that were determined to have a negative safety impact from the loss of the OTW view are presented in Table 3.

A Massachusetts Institute of Technology (MIT) Lincoln Laboratory study (Grappel, 2009) also identified specific hazards associated with the use of surveillance radars, for example:

- Missing data—no surveillance data is provided for a real target
- Erroneous data—partly or fully incorrect or inaccurate data is provided for a real target
- False data—data is presented on the controller display that does not correspond to any real target on the airport.

These fundamental issues are also present for terminal radar coverage for the conventional tower operations. However, the surface surveillance radars add to the safety risk in SNT operations; camera surveillance may not be able to sufficiently mitigate these risks.

	Flight data / Approach in use / Arrival sequence	TMIs, Flight data	Flight plans		Weather alerts / CNS status / Verbal reports	Wind velocity / Ceiling height / Visibility	Surveillance data / Pilot reports / Pilot requests	Weather data / PIREPs / Weather forecasts	Emergency declared / Aircraft lost / Pilot disoriented	CNS status / Equipment status / Performance data	Pilot request / AOC request
Landing Clearance / Other Instructions / Flight Data	**1.1 Manage Arrivals**	Flight Data									
Flight Data / Control Instructions / TMI Coordination		**1.2 Manage Surface Traffic**		TMI Flight Data							
Departure Clearance		Flight Data	**1.3 Manage Departure Clearance Delivery**								
Flight Data / Departure Sequence / Control Instructions		Departure Constraints, Departure Sequence		**1.4 Manage Departure Traffic**							
Safety alerts / Weather alerts / CNS alerts					**1.5 Manage Traffic Safety**						
Runways(s) in Use	Runways(s) in use	Runways(s) in Use		Runways(s) in Use		**1.6 Designate Active Runways**					
Request approval	Aircraft ID / Aircraft position / Aircraft velocity	Aircraft ID / Aircraft position / Aircraft velocity		Aircraft ID / Aircraft position / Aircraft velocity	Safety alerts / Aircraft data		**1.7 Process Surveillance**			Aircraft ID / Aircraft position / Aircraft velocity	
Weather information								**1.8 Manage Weather Distribution**			
Data on Emergency / Navigational Assistance									**1.9 Support Emergency Response**		
CNS Status / Equipment Status						Closed Runway / Available runways / Available CNS				**1.10 Administrative Functions**	
Request approval / Request denial	Pilot Request	Pilot request / AOC Request	Pilot request / AOC Request	Pilot request / AOC Request				Aircraft ID / Aircraft position / Aircraft velocity			**1.11 Respond to Pilot Requests**

Fig. 2 First level N² diagram (Cheng & Nene, 2009)

5.3 Use of Digital Camera for Surface Surveillance (FAA, 2011a; Grappel, 2009)

The functional hazard analysis of SNT operations suggested that some form of digital camera surveillance may be required if an SNT facility were to provide all tower services to an airport.

Any SNT concept would require a surface surveillance display for the controller to obtain the necessary SA for controlling the surface operations. With added camera surveillance, the SNT concept was in need of updates to address the following issues:

- How should the camera images be presented to the controllers?
- How should the camera images be integrated with the surveillance display?

Consequently, possible alternate ways of presenting camera images to controllers were examined with the help of human-in-the-loop (HITL) tests in the laboratory and shadow-operation field tests.

Camera Integration CHI Evaluation (Grappel, 2009**)**

In order to begin to address the camera related issues noted in the section above, a study was conducted at the NextGen Integration and Evaluation Capability (NIEC) laboratory located at the FAA's William J. Hughes Technical Center (WJHTC), Atlantic City, New Jersey (Grappel, 2009). This part-task simulation had a limited scope: evaluation of the computer human interface (CHI) for the display systems, evaluation of a concept for integrating cameras and other surveillance and related

Table 3 SNT functions with negative safety impact (Cheng, 2010)

Function	Conventional tower operation	SNT operation	Negative safety impact
Monitoring surface condition (e.g., braking, FOD, snow)	Controllers fully rely on the visual surveillance via OTW view	Controllers would depend on third party surveillance such as by aircraft personnel and pilots	The loss of OTW view significantly would reduce SA related to surface condition. Controllers may not be able to make a safe judgment on airport surface conditions with the use of camera surveillance
Verifying system safety and non-conformance alerts	Controllers use OTW view to verify potential conflict or other hazardous conditions when the automation system issues such an alert. If verified, controller alerts the pilot	Controllers would depend on the 2-D surface surveillance display and any available camera images to verify the automation alerts	Controllers may not be able to verify all alerts with the use of the display and camera images. The controllers may have to issue unverified alerts that may eventually turn out to be false. The pilots may lose confidence in alerts and may tend to ignore them, potentially compromising safety
Scanning for emergencies	Controllers use OTW view to scan for emergencies on the surface and in the surrounding airspace	Controllers would depend on camera surveillance and third party reports	Controllers would not be able to scan for a number of emergencies such as, for example, aircraft system malfunction

displays, and evaluation of the use of SNT displays for responding to off-nominal events.

There are essentially three basic approaches to displaying camera images:

- A separate monitor to display a local image of a fixed or a pan-tilt-zoom (PTZ) camera
- a picture-in-picture (PIP) display of camera image(s) on the surface surveillance display
- a panoramic display of surface operations by using a number of cameras to cover the entire airport surface.

The part task simulation examined the first two approaches.

The tests were based on a simulation of the East Tower operations at the Dallas-Fort Worth International Airport (KDFW or DFW). The SNT supplemental configuration was used meaning the controllers used the SNT displays in addition to the

Table 4 List of off-nominal events (FAA, 2011a)

ID	Off-nominal description
1	Aircraft initiates missed approach/go around
2	Aircraft deviates from taxi route
3	Aircraft takes wrong heading after take-off (Flight Management System [FMS] programmed incorrectly)
4	Aircraft side-steps to alternate parallel runway during final approach without clearance
5	Aircraft rejects take-off
6	Aircraft fails to continue to climb after wheels up, continues on a runway heading at a low altitude
7	Aircraft initiates take-off roll after clearance to taxi and hold
8	Aircraft fails to hold short of active runway crossing
9	Aircraft crashes on airport and on taxiway(s) or runways(s)
10	Controller issues go-around. Vehicle entering movement area w/o clearance
11	Aircraft altitude falls below the minimum safe value
12	Aircraft taxis to the end of runway after rollout
13	Smoke coming from aircraft/brakes during landing or takeoff

OTW view. An electronic flight strip (EFS) capability was included in the simulation although DFW presently does not use this capability; an EFS capability was expected to be introduced at the large U.S. airports before any SNT implementation.

DFW airport is divided into the East and the West sides. The traffic on these two sides is independently controlled from two separate control towers. An older Center Tower is presently not in use; it is currently used as a back-up facility. A number of controllers were asked to control simulated East side traffic using nine 15-min scenarios with moderate traffic levels. Both day time and night time operations under Visual Flight Rules (VFR) were simulated. All scenarios contained one off-nominal event from the scripted events listed in Table 4. The basic surveillance display used by the controllers is illustrated in Fig. 3; the controllers also used an EFS display and a camera display. The two display configurations are illustrated in Figs. 4 and 5.

At the end of each scenario, the controllers rated different displays for their effectiveness in helping controllers maintain SA as well as efficient and safe operations. In general the controllers gave poor ratings for the use of all camera images—stand-alone, PIP, and PIP + stand-alone. They also did not believe that the camera surveillance in the current configurations helped them in detecting runway incursion events.

Field Demonstration at DFW

Shadow-operation tests were conducted at the DFW Center Tower during a few days in the spring of 2011 to provide a proof of concept for the supplemental SNT configuration. A number of controllers participated in the shadow operation tests

Fig. 3 Controller basic surveillance display (*Source* MIT-LL)
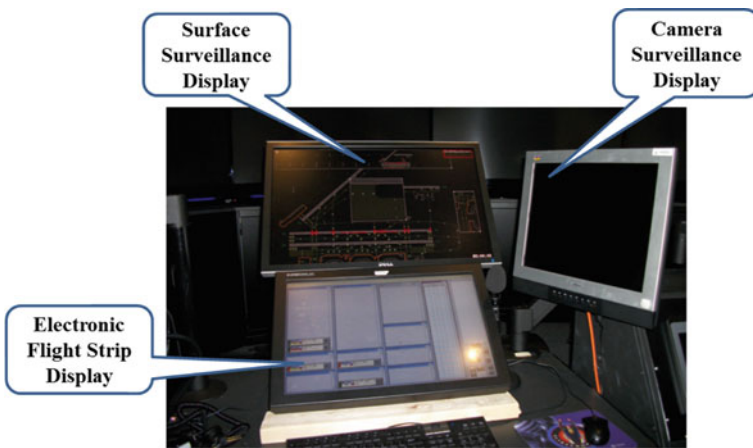

Fig. 4 Controller display configuration 1 (*Source* MIT-LL)

during normal DFW East side operations. The display suite used by the controllers is illustrated in Fig. 6. The display suite included a typical ASDE-X display, a terminal radar display, an EFS display, an SNT surface surveillance display, a camera display, and a communications panel.

The surveillance display was similar to the one shown earlier in Fig. 3. The controllers were also able to use a PIP window to display a selected camera view. The separate camera display was divided into three images as illustrated in Fig. 7.
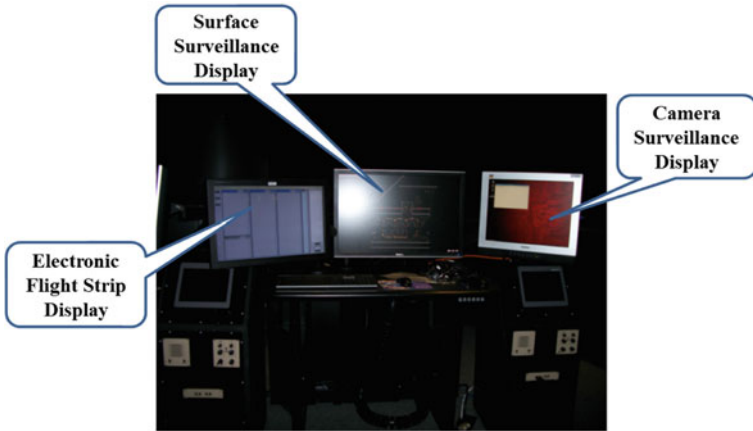
Fig. 5 Controller display configuration 2 (*Source* MIT-LL)

Fig. 6 DFW-2 controller workstation (*Source* MIT-LL)

Fig. 7 DFW-2 camera images (*Source* MIT-LL)

The top half presented a panoramic view of approximately 180° of the east side of DFW as seen from the Center Tower by stitching individual images from four fixed zoom cameras into a single picture (FAA, 2011b); the controllers were not able to change the presented image. The bottom half presented two different views. The left half presented a fixed view of the departure thresholds of the main parallel runways; the right half presented a view from the PTZ camera under the control of Local and Ground controllers. The PTZ image could also be shown on the surveillance display in a PIP window. It must be noted that the camera placement was not optimum and simply used available options for mounting the camera on the tower.

Each controller evaluated alternate ways of using the camera images. As with the part-task evaluation conducted at the WJHTC's NIEC laboratory, the controllers during this shadow-operations test did not see the potential for the use of camera for the SNT supplemental configuration. The study concluded that controller opinions were the result of, at least in part, technical problems such as, an unresponsive interface, poor performance of camera control, lack of sufficient display resolution, and the inability to track targets during windy conditions.

6 Some Observations Related to the SNT Concept

A number of observations can be made with reference to the SNT concept and related HITL tests:

- The concept was very ambitious to envision its application to all sized airports, including large airports
- Although it would be theoretically possible to control multiple airports from a single remote facility, it would not be feasible for the foreseeable future.
- Since the concept focused on large airports that were likely to have surface surveillance, the concept initially focused more on DSTs, the use of ADD, and surface surveillance, and less on the use of cameras to provide panoramic view on multiple monitors
- The display of camera images during tests was not optimal
- Due to the dependence on surface surveillance, the testing was done with operational scenarios related to a large airport (i.e., DFW)
- The tests resulted in lack of controller support and acceptance of camera images for the supplemental SNT configuration.

7 Change in Focus of the FAA's Remote Tower Research

At the end of the SNT tests and evaluation described above, it was becoming increasingly clear that full remote control of towered airports, especially of the large airports, may not be possible in the U.S. for the foreseeable future. Surface surveillance technology that can be certified for separation does not exist today and there are presently no plans to develop and implement such a technology across the NAS. Controllers that participated in the studies mentioned above also did not see significant use for cameras as a part of a supplemental SNT configuration; the use of cameras in the absence of the OTW view was not a part of any SNT-related HITL. Furthermore, the controller community in the U.S. did not support any concept for remote control of presently towered airports (http://natca.org/natca_insider.aspx?zone=NATCA-Insider&nID=4737). Therefore, given these challenges and perhaps others, the FAA changed the direction of its remote tower research and focused it on improving the services at the presently non-towered airports (NTAs). It was thought that the remote tower technology could address some of the operational shortfalls at the NTAs, and since these NTAs do not presently receive tower services, remotely improving services may be acceptable to the ATC community. As a first step in this effort, present NTA operations were analyzed for defining the need, if any, for improving services provided to these airports.

7.1 Current Operations at Non-towered Airports (NTAs)

NTAs predominantly serve unscheduled VFR traffic although some of them also serve scheduled IFR traffic. The VFR traffic in and out of these airports is uncontrolled and cooperatively organizes itself to maintain safety. The pilots are jointly responsible for collision avoidance, sequencing, and following local procedures. Separation services

are provided by controllers only to the IFR traffic; these are provided from the overlaying Air Route Traffic Control Center (ARTCC) or TRACON facility. IFR pilots are still responsible for safely merging with VFR flights operating on and around the airport.

The surrounding airspace is normally below 2,500 feet Above Ground Level (AGL) and within approximately a 5-mile radius from the airport center. Each pilot gathers information about nearby aircraft operations primarily by looking out the cockpit window. Pilots also listen to traffic advisories and airport information that may be provided over the radio by other participating pilots or a ground station, if present. Due to the complexity of operations, right-of-way rules, traffic patterns, and other procedures exist at NTAs primarily to prevent collisions in the air and on the ground.

VFR traffic is not required to participate with ATC at these airports, and as a result, adequate ATC radio coverage is often not available at many NTAs. Some airports, however, do provide some auxiliary channels to fill the communications gap. When available, TRACON or ARTCC personnel use a designated ATC frequency to communicate with IFR pilots at some airports for the purpose of clearance delivery. But at airports where surface radio coverage is not available, IFR pilots typically contact ATC personnel using various telephone media. Surveillance at NTAs varies widely and in many cases does not exist at lower altitudes or on the surface.

Airport advisories are typically broadcast for pilots via the common traffic advisory frequency (CTAF). The CTAF is normally the frequency for a Universal Integrated Community (UNICOM), MULTICOM, or Flight Service Station (FSS); it could also be the tower frequency, when used at times outside of the tower's hours of operation. The advisories include airport information, weather information, wind direction, or upon pilot request, the recommended runway or current runway in use, when known.

NTAs typically feature a mix of air traffic types, including general aviation (GA), helicopter, air carrier, air taxi, and military. Non-standard flight operations, such as ultra-light, parachute, balloon, and lighter-than-air, are also common. Aircraft performance can vary significantly among traffic types at these airports. Air carrier and air taxi operators typically operate large, heavy, and fast aircraft. In contrast, a large portion of GA traffic consists of light, single-engine piston aircraft that move at low speeds. Equipage levels also differ, as some aircraft are radio and transponder equipped while others do not even have an electrical system and therefore lack such equipment. Furthermore, pilot experience and skill level differ between professionals conducting passenger services, amateurs flying as a hobby, and students at a flight school.

Aircraft that operate in the controlled airspace beyond the NTA commonly receive ATC services from the ARTCC or TRACON controller associated with the non-towered airport. In controlled airspace, controllers communicate with the pilots on a designated radio frequency separate from the CTAF. Once in controlled airspace, VFR flights typically contact ATC to receive updates on traffic or weather information or request VFR flight following services. These services are simple compared to those requested by IFR flights, which require ATC clearance prior to departing or

approaching the NTA. Controllers provide departure and approach services using one-in one-out procedures in order to ensure separation between IFR flights. The difference between VFR and IFR operations at NTAs is summarized as follows:

a. **VFR Operations at an NTA**: The VFR traffic is not required to communicate with ATC and essentially organizes itself by following the right-of-way rules mentioned earlier. The pilots obtain their SA of the traffic around them via visual observations and CTAF communications.

b. **IFR Departures from an NTA**: The FAA requires IFR flights to file a flight plan in order to operate at NTAs. Prior to take off, IFR departures require a departure clearance from ATC. If the clearance is not immediate, ATC will then issue a departure release time and a clearance void time. The clearance is voided if the aircraft is not airborne by the specified void time, and the pilot must then request a new clearance from ATC. If there are radio coverage gaps at the airport, pilots typically call ATC via telephone or coordinate through the FSS via CTAF if available. Prior to departure the pilot typically tunes the radio to the CTAF and therefore is unable to communicate further with ATC until clear of the traffic pattern.

c. **IFR Arrivals to an NTA**: Pilots of IFR arrivals that are approaching an NTA but that are still in controlled airspace are in communication with ATC, and in turn, ATC monitors the position of the aircraft via available surveillance and provides instructions along the arrival procedure. Typically, prior to reaching the boundary of the controlled airspace, ATC clears the flight for approach and releases it to the CTAF. The ATC terminates radio communication with the pilot, and the pilot lands the aircraft following the right-of-way rules described earlier. The pilot then contacts ATC and cancels the IFR flight plan. Alternately, if the pilot executes a missed approach, he/she climbs to an altitude when communication with ATC is possible and reports the missed approach. ATC then directs the pilot to either repeat the approach procedure, hold, or divert to an alternate airport. When operating in Visual Meteorological Condition (VMC), the pilot has the option to terminate ATC services by canceling the IFR flight plan and operating the remainder of the flight by following the procedures described earlier; the pilots often choose this option if continuing with the IFR flight plan would result in unacceptable delays.

7.2   Present Shortfalls in NTA Operations (Colavito, 2013)

A number of the NTAs have sizable IFR operations. Some ski airports in Colorado, for example, have high levels of peak hourly IFR traffic during the ski season. The traffic in and out of these airports is also predominantly unscheduled and, as a result, there is large variability in the hourly demand for arrivals and departures. In view of these operational characteristics, the NTAs exhibit a number of operational shortfalls. The restrictive one-in one-out operations severely limit the IFR capacity resulting in significant holding and vectoring around the airport and ground

delays. The combination of unscheduled demand and low IFR capacity also introduces inefficiency in operations. Delays and inefficiency, in turn, result in excessive operating costs, increased fuel consumption, and increased emissions. Decentralized self-coordination of traffic, traffic complexity, and other factors also lead to flaws in SA of pilots, resulting in heightened safety concerns at these airports.

7.3 Concept for Remotely Providing Selected NTA Services

There are two possible approaches to remotely improving ATC services provided to the NTA airports. One is to provide all services presently provided at a towered airport. Another approach is to provide only a set of select services and establish an operational environment somewhere between an uncontrolled NTA and a fully controlled towered airport. A concept for provided a select set of services, referred to as the Select Services Concept, has been developed as part of the FAA's NextGen research efforts and is presented in this section.

7.3.1 An Overview of the Select Services Concept (Nene, 2013)

Under the Select Services Concept, ATC would organize both the IFR and VFR aircraft to and from the NTA airport, sequence the IFR aircraft closer together than when using the classical one-in one-out operations, and maintain safe separation on the airport runways. ATC would not provide the control of aircraft/vehicle movement on taxiways.

The proposed concept leverages three foundational air traffic control principles in use throughout the NAS today: an established area of ATC jurisdiction, use of surveillance information to monitor and separate traffic, and instantaneous two way radio communication between controller and pilots. Under the concept, controllers would use surveillance data to determine position information of aircraft in the airspace immediately surrounding the airport and of aircraft and vehicles on key airport surface areas. Controllers would use the surveillance data just as they do today to provide separation between airborne IFR aircraft and to provide traffic information to IFR and VFR pilots operating near the airport. In addition, controllers would use surface surveillance information to improve SA of operations on or near the runways. Although surface surveillance would depict aircraft on the surface, controllers would use two-way radio communications to obtain pilot position reports regarding key surface information such as *clear of the runway* upon runway exit or *holding short of the runway* when approaching a runway.

The operational environment under this concept is compared with the non-towered and towered environment in Table 5.

Table 5 Operating environment for the proposed select services concept (Nene, 2013)

Item	Current non-towered operations	Proposed select services concept	Current towered operations
Surface movement	ATC does not control surface movements on taxiways and parking/apron areas. Controllers do not issue taxi instructions. Pilots use taxiways and runways at their discretion	ATC would not issue taxi instructions; they would instruct pilots to report holding short of the assigned runway and report clear of active runways	ATC issues taxi instructions to pilots; they observe surface movements and that aircraft are holding short and are clear of the runway
Control of airspace surrounding the airport	ATC separates IFR aircraft from IFR aircraft. IFR and VFR aircraft self-organize in the airspace and in the VFR traffic pattern. Aircraft execute the basic turns in the pattern on their own	ATC would determine the landing order of all aircraft. ATC would control and integrate IFR and VFR aircraft in the airspace and establish a VFR pattern using control instructions based on the radar information. Aircraft on frequency but not displayed on radar would be managed as able based on other traffic	ATC determines the landing order of all aircraft. ATC controls and integrates IFR and VFR aircraft in the airspace and VFR pattern using direct viewing of the aircraft and uses radar information as an aid
Runway configuration	Pilots are free to determine their arrival and departure runway	ATC determines the active runway(s) and runway use	
Control of runway operations	IFR arrivals are instructed to proceed for landing; IFR departures released for departure with a void time VFR traffic self-organizes and is not controlled at all	Both VFR and IFR traffic receive landing and departure clearances. ATC ensures runway is clear of all known conflicts. Pilots self-separate on the uncontrolled taxiways and apron areas	
ATC participation	VFR aircraft are not required to participate in ATC operations or communicate with ATC	All aircraft are required to participate in ATC operations and communicate with ATC	
Transponder equipage	Aircraft are not required to carry a transponder. Unequipped aircraft receive service as they currently do in Class D airspace		

7.3.2 Assumptions and Constraints

The Select Services Concept is based on the following assumptions:

- ATC services will be provided by controllers located at a facility away from the airport; such a facility may be located at the overlying ARTCC or TRACON facility.
- The controllers at the remote facility will be radar controllers and will accept hand-offs from ARTCC or TRACON radar controllers providing approach services to the airport.
- If the airspace around the airport is presently designated as Class E or Class G, it will be designated as Class D. If the airspace classification is already higher than Class D, it shall remain unchanged.
- Consistent with the airspace classification, all aircraft operating in and out of the airport will be required to carry radio communication equipment and will be required to communicate with ATC and follow ATC instructions. However, aircraft without such communications equipment or with failed equipment will be accommodated, although they may receive a reduced level of services than those provided to radio-equipped aircraft.
- A unique frequency will be assigned to the airport for air-to-ground ATC communications; communication coverage will extend to all airport surface areas and the airspace immediately surrounding the airport.
- Secondary surveillance will cover all Runway Safety Areas (RSAs) and taxiways that are adjacent to or cross the runways.
- The status of all ground movement of transponder equipped aircraft and other transponder equipped vehicles in the movement area will be presented to the controllers on a 2-D display.
- The airborne location of transponder equipped aircraft will be presented to the controllers on a 2-D display certified for separation.
- Consistent with current operations in Class D airspace, aircraft will not be required to carry a transponder to be able to operate in and out of the airport. Aircraft unequipped with a transponder will be accommodated although they may receive a reduced level of services than those provided to transponder-equipped aircraft.
- The airport will be required to have an automated weather observation system.
- Controllers will determine the active runway configuration at the airport and issue clearance for landing on and take-off from the active runways; ATC may authorize pilots to use other-than-active runway(s).

7.3.3 Changes to NTA Operations

There are three significant ways in which the Select Services Concept would change the present NTA operations:
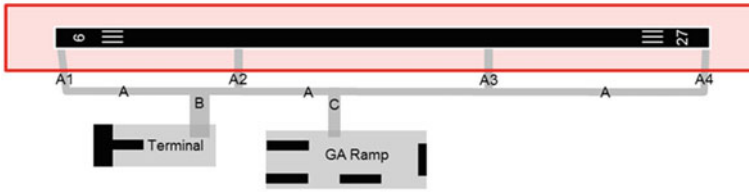
Fig. 8 Airport surface under ATC jurisdiction at a notional airport (Nene, 2013)

- All aircraft would be required to carry radio equipment onboard and participate in the ATC. This would be accomplished by making necessary changes in the airspace designation.
- Pilots would not be able to select the runway they use; the controllers would determine the runway configuration.
- The VFR traffic would no longer be able to land or takeoff at will. All VFR and IFR runway operations would be under control ATC control.

7.3.4 Airspace Jurisdiction

The airspace around the airport operating under the Select Services Concept would be classified as Class D as usually found around the small towered airports. Such a classification will require pilots to establish and maintain radio communication with ATC. Based on the local needs to effectively integrate IFR and VFR operations, the exact shape and volume of airspace might be larger than that found at towered airports.

7.3.5 Surface Jurisdiction

Under the Select Services Concept all runway surfaces and taxiways within the RSA would be under ATC jurisdiction. A typical RSA may extend up to 1000 feet beyond each runway end. All aircraft and vehicles would be required to receive a clearance from ATC to enter and move within the RSA. All other airport surface area outside the RSA would be designated as non-movement area and will be uncontrolled. These areas at a notional airport are illustrated in Fig. 8.

7.3.6 Surveillance Coverage

The Select Services Concept envisions that airborne surveillance would be continuous and seamless from the overlying enroute airspace, cover all charted paths to the Class D airspace, and cover the full extent of the Class D airspace.

The surface surveillance requirements would be defined for three distinct zones as illustrated in Fig. 9.
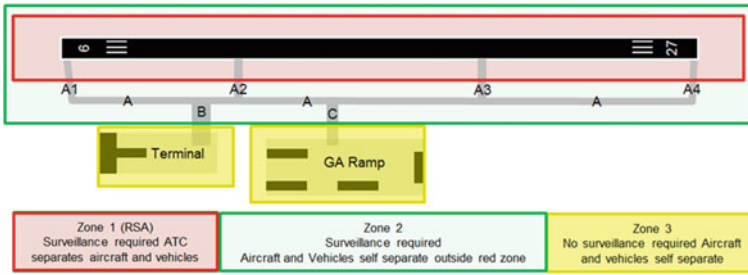
Fig. 9 Three surface surveillance zones for a notional airport (Nene, 2013)

Zone 1 is the RSA area defined earlier; surveillance would be required here and ATC would provide separation service. Zone 2 would typically cover all the taxiways in proximity to Zone 1. It would provide a buffer for ATC to detect when aircraft enters or leaves its area of jurisdiction. However, aircraft and vehicles would self-separate in this area. Surveillance would be required for Zone 2 as well. Zone 3 is for all other airport surface that would be uncontrolled and would not require surveillance.

As with the current surface surveillance systems at the high-end airports, the proposed surveillance would not be certified for use in separating aircraft. It would provide surface SA for the controller. The controller would use verbal pilot reports to determine if an aircraft is clear of a runway.

7.3.7   Surface Surveillance Display

The location of aircraft on the specified portion of the airport surface (runways and all taxiways that are directly adjacent to or that cross the runways) would be obtained from cooperative surveillance. The aircraft position in the terminal airspace around the airport would also be obtained from radar and other sensors that may be in use around the airport. The aircraft would be shown on a 2-D display overlaid on the geographical map of the airport. The display would appropriately differentiate between aircraft in the air and on the ground, as well as between arrivals and departures. All aircraft would also be tagged with the necessary data block showing the aircraft ID, altitude, speed, and other parameters. Appropriate weather maps and necessary weather information may also be made available on the surveillance display. The display would also be configurable to accommodate individual controller preferences. Applicable safety alerts would also be presented on the surveillance display.

Any aircraft unequipped with a transponder and any aircraft with failed onboard radio would be accommodated the same way as they are in today's tower environment.

7.3.8   Expected Benefits

Since all IFR and VFR would be under ATC control in the Select Services Concept, the IFR capacity is expected to increase significantly, and aircraft are expected to experience a reduction in holding and vectoring around the airport. The use of surface surveillance display would also significantly increase controller SA and operational safety at the airport.

7.3.9   Concept for Remotely Providing Full Services (Nene et al., 2013b)

If full tower services are provided remotely at the NTAs, rather than only select services, surface surveillance must be expanded to include non-cooperative surveillance as well as camera surveillance. This additional surveillance would provide additional SA that would be necessary for providing control instructions related to taxi movements and for accommodating unequipped aircraft. Camera images could be presented to the controller in a variety of ways such as by displaying a panoramic view of the airport surface on multiple monitors or on a separate single monitor, and/or through a PIP window on the surveillance display.

7.3.10   Status of Remote Tower Concepts for NTAs

In 2013, as the FAA was examining alternate paths for continuing remote tower research, the State of Colorado began an initiative for remotely improving services at its non-towered ski airports and requested the FAA to initiate a joint development program for such a concept. As a result, the FAA is currently not pursuing the development or validation of the Select Service or Full Service Concepts described above.

8   Present Effort on the Colorado Initiative

There are a number of non-towered ski airports in the mountainous areas of the State of Colorado that exhibit high traffic levels during the ski season. The lack of conventional radar coverage coupled with the one-in one-out operations resulted in limited capacity at these airports. During some periods of time, the single runway IFR arrival rate reached only 4–6 per hour (Payne, 2011). This low capacity resulted in limited access to the airports and in significant vectoring, holding, and delays.

As a response to the above shortfalls, the State of Colorado and the FAA jointly began the development and deployment of Wide Area Multilateration (WAM) and ADS-B in the mountainous areas of Colorado. The use of WAM now provides surveillance down to 500' AGL at eight airports (Payne, 2013) and has improved operations at these airports. Presently WAM coverage extends almost to the ground level

at a few airports, and it can be extended to the ground at all airports if additional transmitters/receivers are installed at these airports.

Currently, the State of Colorado is interested in a remote tower-like concept which would help achieve additional operational benefits at the non-tower airports now covered by WAM. The initiative seeks to use WAM surveillance information, knowledge of surface traffic, and appropriate ATC rules to allow a controller to manage traffic to and from a non-towered airport in an integrated and seamless manner across the airspace. The controllers handling the traffic at these non-towered airports could be located away from the airports, or in a nearby TRACON or ARTCC facility.

The FAA is presently pursuing the Colorado initiative to decide if a formal FAA acquisition process should be undertaken to implement the initiative. The acquisition management system (AMS) process requires the FAA to develop specific formal documentation such as an operational concept, functional and safety analyses, operational requirements, and cost and benefit analyses.

The Colorado initiative development began with the review of the earlier developed concepts for remote NTA operations (Nene, 2013; Nene et al., 2013a) and is presently formulating its own operational concept. It conducted independent HITL and other studies to examine surveillance requirements, safety impacts, and the necessary controller interfaces. Although the concept is still under development, it is expected that a field demonstration will be conducted at an airport in Colorado sometime in the year 2016.

9   Remote Tower Demonstration Project at the Leesburg Executive Airport (KJYO), Leesburg, Virginia

The Leesburg Executive Airport is one of the busiest NTA on the U.S. east coast. In view of the high cost of building a new tower, The Town of Leesburg is exploring the possibility of establishing a camera-based remote tower facility for improving the services at Leesburg Airport. The Virginia Department of Transportation's (VDOT) public–private research arm, called Virginia Small Aircraft Transportation System (VSATS), has agreed to help fund a test of such a remote tower facility. VSATS and the Saab Sensis Corporation have installed a remote tower (rTower) workstation inside a room at the airport terminal; the related camera array is located on the rooftop. A temporary trailer-mounted physical control tower would also be located on the ramp area so that safety of rTower operations could be compared to the safety achieved by the use of a traditional tower. The traffic in and out of Leesburg Airport would be controlled for approximately 12 h per day during a three-month test period sometime in the year 2015 while the FAA collects safety related data (Town of Leesburg, 2014). If the FAA finds the safety level of rTower operations acceptable, it is expected that the FAA would approve the use of the facility for normal operations. This demonstration project is still being organized and no additional information is presently available.

10 Future of Remote Towers in the U.S.

The FAA research to date on remote tower technology has clearly identified significant technical and operational issues that must be addressed before this technology can be implemented within the NAS. The first and the foremost is the need to unequivocally decide if there is a need for camera surveillance, and if so, determine appropriate ways of presenting camera images to the controllers. Some of the outstanding issues to be resolved include, among others, determining the best ways of remotely responding to off-nominal events in airport operations, addressing the need for requiring transponder equipage irrespective of the airspace classification at the airport, and determining if the remote tower concepts should accommodate only a subset of the services presently provided at the towered airports.

Both the Colorado initiative and the Leesburg effort are expected to continue to examine these issues for the next several years. It is, therefore, not clear at this time if one or both of these concepts will be developed and implemented across the NAS in the future.

References

Boehm-Davis, D. A., et al. (2010). *Report on verbal protocol analysis for Staffed NextGen Tower*. George Mason University, Fairfax, VA.

Cardosi, K., & Yost, A. (2001). *Controller and pilot errors in airport operations: A review of previous research and analyses of safety data*, DOT/FAA/AR-00/51. Volpe National Transportation Systems Center.

Cheng, J., et al. (2010). *Staffed NextGen Towers (SNT) safety impact assessment*, MTR100298. The MITRE Corporation, McLean, VA.

Cheng, J., & Nene, V. (2009). *A functional analysis of Staffed NextGen Tower (SNT) operations*, MTR090222. The MITRE Corporation, McLean, VA.

Colavito, A., et al. (2013). *Shortfall analysis for air traffic services at non-towered airports*, MTR130081. The MITRE Corporation, McLean, VA.

Colavito, A., & Nene, V. (2010). *A proposed classification of United States (U.S.) towered airports*, MTR100304. The MITRE Corporation, McLean, VA.

FAA/George Mason University, Fairfax, VA. (2009). *Draft report on Staffed NextGen Tower walkthrough.*

FAA. (2008a). Order 7110.65S, air traffic control, Section 3-1-12, 14, Washington, DC.

FAA. (2008b). *An operational concept for NextGen Towers, Version 5.1.*

FAA. (2011a). *Staffed NextGen Tower past-task camera integration CHI evaluation—Final report*.

FAA. (2011b). *Field demonstration #2 final report for Staffed NextGen Tower (SNT).*

FAA. Long-Range Aerospace Forecasts; Fiscal Years 2020–2025 and 2030, 2006. http://faa.gov/data_statistics/aviatino/long-range_forecasts/.

Grappel, R. (2009). *A safety hazard assessment for the Staffed NextGen Tower (SNT) system.* Report 42PM-SNT-0006, MIT Lincoln Laboratory.

http://natca.org/natca_insider.aspx?zone=NATCA-Insider&nID=4737

Nene, V. (2007). *Initial technology assessment for NextGen virtual towers*, MP070067. The MITRE Corporation, McLean, VA

Nene, V. (2009). *Analysis of off-nominal events in Staffed NextGen Tower (SNT) operations*, MTR090407. The MITRE Corporation, McLean, VA

Nene, V., et al. (2013a). *A preliminary operational concept for remotely providing select tower services to non-towered airports (NTAs)*, MTR130175. The MITRE Corporation, McLean, VA.

Nene, V., et al. (2013b). *A proposed operational concept for providing remote tower services (RTS) to non-towered airports*, MTR130058. The MITRE Corporation, McLean, VA.

Payne, W. (2011) The Colorado surveillance project—A NextGen success story. In *Advanced Technology Symposium*, Montreal, Canada.

Payne, W. (2013). *Colorado surveillance project phase III—Blended airspace.* W. E. Payne Associates, Denver, CO.

Pinska, E., & Bourgois, M. (2005) Behavioral analysis of tower controller activity, EUROCON-TROL Experiment Center Activity Report.

Town of Leesburg. (2014) *Town Council Worksession of 11 August 2014.* www.leesburgva.gov/WEBLINK8/0/doc/28199/Electronic.aspx

Remotely-Operated AFIS in Japan

Satoru Inoue and Mark Brown

Abstract The Japan Civil Aviation Bureau has been providing remote aerodrome flight information (advisory) services (remote AFIS) to small aerodromes such as island and local airports over 40 years. This remote AFIS is called "Remote Air-Ground communication" ("RAG" for short)". RAG services are provided to 36 small aerodromes in Japan. In RAG operation, an aeronautical flight information services operator (AFISO) at a Flight Service Centre (FSC) gives information to aircraft pilots through VHF radio communication. The RAG working position has equipment including a flight information display, a surveillance display (air route surveillance radar and/or wide-area multilateration, since small aerodromes do not have approach or surface movement radars), weather information, and a video display of images from a PTZ camera and/or 180-degree coverage fixed camera at the aerodrome. In some cases, an AFISO is able to handle two small aerodromes simultaneously when the traffic volume is low. The RAG working positions for multiple aerodromes are equipped with two video circuits, one for each aerodrome. The AFISO usually monitors traffic at each airport by a combination of radar and the video display. In this chapter, we give an overview of the RAG service system functions and typical operations processes. We also discuss the operational issues with the existing equipment and services that led to investigations of Remote Tower technology for the next system.

Keywords Remote Aerodrome Flight Information Services (AFIS)

1 The AFIS Situation in Japan

There are currently 92 airports in Japan, with three types of aerodrome Air Traffic Services (ATS) provided: Air Traffic Control Service (ATC), Aerodrome Flight Information Services (AFIS), and remotely-operated AFIS (Remote AFIS) (Airports in Japan, 2020; Remotely Operated Aerodrome Flight Information Service, 2012). The

S. Inoue (✉) · M. Brown
Electronic Navigation Research Institute, National Institute of Maritime, Port and Aviation Technology, 7-42-23, Jidaiji-higashi-machi, Chofu 118-0012, Tokyo, Japan
e-mail: s-inoue@enri.go.jp

major difference between ATC and AFIS/Remote AFIS is that ATC issues control instructions, with controllers assuming responsibility for safe separation of aircraft, while AFIS and Remote AFIS provide only advisory information and responsibility for safe separation rests with the pilot. ATC service is provided to 42 airports, including medium-sized and larger airports, and those that are likely to have poor weather conditions. The remaining airports are provided with AFIS, either from a tower at the airport, or from a remote facility, in which case the service is called a Remote Air-Ground Communication Service, or RAG. Figure 1 shows the current RAG airports in Japan. RAG is currently provided to 12 airports with a relatively low volume of scheduled traffic (about 10,000–15,000 flights per year), and to 33 airports with even less traffic (about 10,000 flights or less per year) including small island airports.

In Japan, aerodrome ATS for many small island airports, which have few regular flights, have already been remoted, and RAG services to small airports on the major islands of Honshu, Kyushu and Hokkaido have also expanded in recent years.

The purpose of AFIS is basically to provide necessary information to an aircraft at appropriate times. To achieve this, the AFIS operator visually monitors the condition of the aircraft either by looking out of the tower window in a tower facility, or by observing the aircraft on a closed-circuit television monitor at the RAG facility fed from an Industrial Television (ITV) camera at the airport. Visual monitoring is therefore an important factor in AFIS as well as ATC service provision.

The operating rules for conventional and remote AFIS are basically the same, but there are some differences in the content of the information provided, procedures,
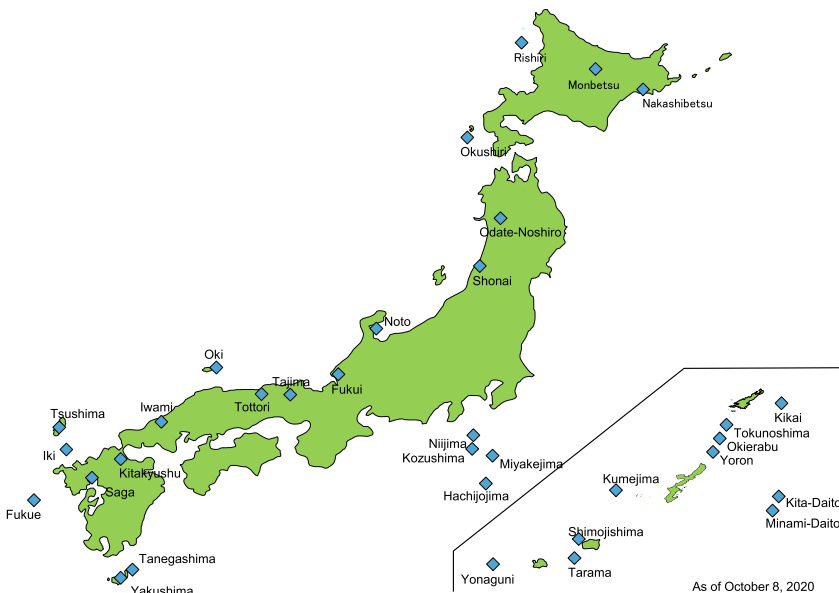
Fig. 1 Overview of remote AFIS airports

and communication phraseology due to the fact only limited visual observation of an RAG airport is possible. For example, the RAG operator does not provide information on weather, traffic, or airport conditions that rely on direct observation due to field of view and image quality limitations of the ITV camera at the airport, but instead information from sensor-based equipment such as weather data from a Terminal Display Unit (TDU) and traffic surveillance information from an Aircraft Position Display Unit (APDU) are provided as necessary. In addition, the RAG phraseology is adapted for the environment assuming that ITV video observation can be used only in a supplementary support role rather than directly for operations: that is, the basic premise of an RAG service is that no information derived from visual observation can be provided. For example, if the runway is clear of traffic or obstacles at the time of a departure and arrival, an AFIS operator at a conventional facility confirms the state of the runway by direct observation and states "Runway is clear". On the other hand, in the case of an RAG service, the runway is not observed directly so the phrase "Obstruction is not reported on runway" is used. Both phrases convey the same runway state, but the expressions are different. Furthermore, when an aircraft lands at an RAG airport the pilot is requested to "Report runway vacated" and then after vacating is requested to "Report down time" to communicate the landing time. A conventional AFIS does not require these reports because the status can be directly observed by the AFIS operator from the tower.

2 Remote AFIS in Japan

Operating of remote AFIS in Japan began more than 40 years ago. The Japan Civil Aviation Bureau (JCAB) started services in 1974 initially at two airports, Rishiri Airport (RJCR) and Okushiri Airport (RJEO), which are both on small remote islands with very little air traffic. This section describes the Remote AFIS service operated in Japan.

The Air Traffic Information Zone around an RAG airport is a circular area of airspace 5 NM radius centred on the airport, and is classified as Class E (AIM-j, 2020) according to the standard in ICAO ANNEX 11 (2018); that is, it is controlled airspace in which aircraft operating under Instrument Flight Rules (IFR) require ATC clearance and are separated from each other by ATC (in this case, by allowing only one IFR flight to fly within the zone at a time), whereas VFR flights do not require clearance and ATC assumes no responsibility for separating them from other traffic, but provides traffic information where possible. The RAG operator provides flights with flight information, weather information, and safety-related information on runway conditions and traffic conditions on the airport surface and within the Air Traffic Information Zone, and relays clearances between pilot and ATC.

There are eight AFIS operation centres called FSCs (Flight Service Centres) which are located at medium or large airports, and remote aerodrome AFIS services are provided from an RAG operation room in an FSC. The FCS provides services to multiple RAGs, and is connected to each by a communication line. There are currently

Fig. 2 Remote AFIS operation room in FSC

two types of RAG operations: operation of an airport by a single operator at time (single mode operation), and simultaneous provision of services to two airports by a single operator (multiple mode operation). Multiple mode operation is targeted at small airports with few scheduled flights, while single mode operation is mainly used for airports with more than a certain level of scheduled traffic.

Figure 2 shows an FSC RAG operation room with three operator positions (consoles), each of which is connected to two airports and is designed for multiple-mode operation. This FSC centre can serve six RAG airports simultaneously.

The RAG console has air-ground radio communication links with the airports and information displays, including weather information (TDU: Terminal Display Unit), flight plan information (FACE: Flight object Administration CEntre system), aircraft position information (APDU: Aircraft Position Display Unit), and a Multi-Purpose Information Display (MPID) that can show airport AIP and other information. In addition, a visual surveillance display (ITV monitor) displays live video that shows the situation at the airports. The ITV monitor and APDU information are used by the RAG operator as auxiliary (reference) information. Although as mentioned above the RAG AFIS procedures are based on the premise that visual surveillance is not available, using ITV video images as a tool to support operator awareness of the movement of aircraft and the timing of operation flow is efficient and safe. From this point of view, the ITV monitor has high utility and value. Figure 3 shows the RAG console for Remote AFIS multiple-mode operation, allowing an operator to handle two airports simultaneously. Displays for each airport are symmetrically arranged on the left and right on the console. A display for the FACE (Flight object Administration Centre) system that manages flight information is located in the centre of the console and presents information on both airports. Details of each display will be described later.

Air-ground radio voice channels for each airport are monitored through a speaker. The operator monitors the air-ground voice transmissions and the information displays of each airport on the left and right of the console. When there is traffic at an airport, the operator uses corresponding displays.

Figure 4 shows the communication connections between the RAG console and airport equipment. The ITV (1), weather (2), air-ground voice communication (3), and radar surveillance (5) systems are connected by Internet Protocol (IP) data virtual
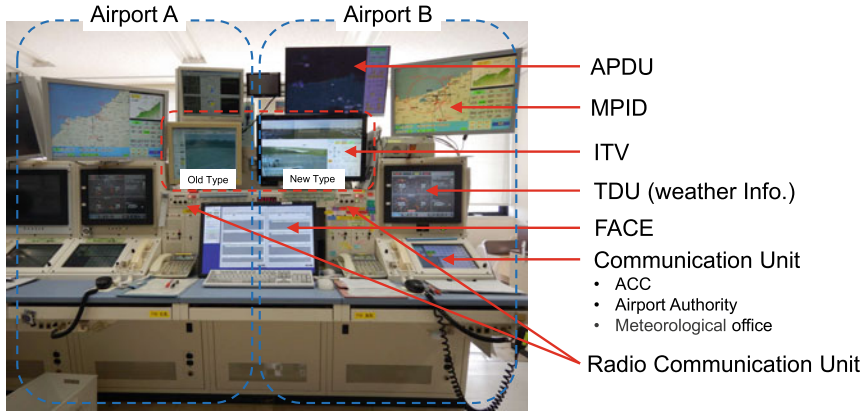
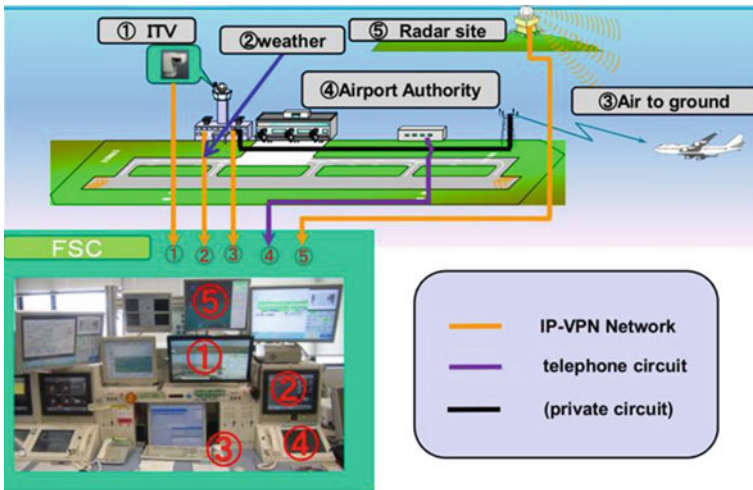**Fig. 3** RAG console for multiple mode operation

Fig. 4 Network connection structure of the RAG systems

private network (VPN) connections. Air-ground voice signals between the VPN interface and the air-ground radio transceiver use a dedicated circuit, and the airport authority can communicate with the RAG operator via a telephone "landline" circuit, (Communication circuits to other ATS facilities, airport fire and rescue services etc. are not shown.) To avoid degradation of the video due to packet loss video signals are transmitted by TCP/IP, which automatically retransmits packets if they are dropped by the network and guarantees in-order packet delivery. This creates latency, but since the ITV system is considered an auxiliary device, this is acceptable.

3   System Displays and Their Functions

The RAG console system consists of five main information systems (APDU, ITV, FACE, TDU, MPID), air-ground voice radio equipment, and landline voice circuits. An outline of each system is given below.

3.1   APDU (Aircraft Positioning Display Unit)

Shown in Fig. 5, the APDU is a display introduced in 2009 that presents aircraft positions with callsign, altitude and ground speed from Air Route Surveillance Radars (ARSR). It is purely a display with no operational information input capability. However, display controls shown on the right side of the position display in Fig. 5 allow the operator to vary the range and select information.

Typically, small-sized airports do not have a local surveillance radar, but when there are no obstacles such as mountains in the vicinity, the APDU provides for surveillance of the airspace above and around the airport's Air Traffic Information Zone down to a certain height. An ARSR has a relatively slow update rate (10 s) compared to a terminal radar, and ASRS are intended for high altitude surveillance, so the APDU cannot effectively cover the final descent and runway approach of inbound aircraft. However, it is effective in giving the operator awareness of the positions of aircraft approaching the airport.

Fig. 5   APDU display shows aircraft position based on en-route radar source

Operators use the APDU to provide traffic advisory information and to understand the traffic situation and anticipate a flight's information requirements. Although it cannot be used effectively for surveillance in the vicinity of the airport, for example during final approach and initial climb-out, the APDU can be used by the RAG operator to monitor an arrival until it enters the Air Traffic Information Zone. For example, if there is also a departure waiting at the airport, the APDU is useful for actively providing information on the situation of the arrival flight and to allow the departing pilot to time the departure clearance request.

3.2   ITV

The ITV system is a real-time video system that allows the operator to monitor visually the airport surface and traffic in the immediate vicinity of the airport, although its limited field of view and resolution, and communication latency means that it cannot serve as a replacement for visual observation from an airport tower and must be used in a support (auxiliary) role. Although AFIS is an advisory service and there is no obligation for RAG operators to maintain a continuous visual watch of the airport surface, by using the ITV system the operator easily grasp the positions of aircraft and vehicles and actively provide information necessary for safe and efficient operations.

There are two basic ITV system configurations in use depending on the size of the airport and when the RAG system was installed: a basic system and an advanced system. The basic system consists of two ITV cameras with PTZ (Pan-Tilt-Zoom) capability that can be used to monitor any point on the airport's manoeuvring area. As well as manual operation of the camera functions, the camera can be moved to preset position and zoom settings stored in a memory at the push of a button. Continuous PTZ movement patterns can also be stored in the memory, allowing automatic scanning along the runway or periodic monitoring of several positions in sequence.

The advanced ITV system was developed to give more performance and capability for airports with more traffic than the first system. It consists of three high-definition (Full HD, 1920 × 1080 pixel) fixed cameras that provide a 180° panoramic image that includes the runway and two Full HD PTZ cameras, one of which is a backup to improve system availability. In addition, there is a "box and follow" video-based moving object detection and tracking capability that can be used for both the PTZ camera and fixed camera views.

An ITV display from the newer system is shown in Fig. 6. The operator's interface is a 22-inch display screen with touchscreen input capability. The 180° panorama view is presented at the top of the display, while a PTZ view is shown in the lower left area and the lower right area contains indicators and touchscreen controls. Figure 6 shows an example of a box and follow (green box around an aircraft on the panorama view and a red box on the PTZ view. The box and follow function mainly assists the operator in detecting moving objects on the airport surface and allows more rapid

Fig. 6 Overview of the ITV display console

detection of moving objects. It can also help to prevent the operator overlooking objects by providing a cue as to their presence. The PTZ camera is also equipped with a video-based object-following function that automatically orientates the camera to keep a moving object near the centre of the view. The box-and-follow function can be operated by pushing a button.

Since both the box-and-follow and PTZ following are use image processing-based moving object detection, detection stops when the object stops. Also, continuous tracking tends to be difficult when visibility is poor. Despite these drawbacks, these functions greatly assist the operator in monitoring airport surface traffic from the remote FSC.

3.3 FACE (Flight object Administration CEntre system)

FACE is a system for managing departure and arrival schedules. The RAG operator uses the Traffic Information Log (TIL) shown in Fig. 7 for basic tasks. The TIL serves the same purpose as Flight Strips to show information related to each flight at the airport(s), including callsign or registration, estimated time of departure or arrival, and destination and departure airport, based on which the operator can anticipate and plan for each flight's management and information needs. FACE can also display time-varying operational information such as NOTAMs (Notices to Airmen).

3.4 TDU (Terminal Display Unit)

Meteorological information is one of the most important pieces of information provided to pilots when an aircraft departs or arrives. The TDU shows meteorological information and sensor data such as wind direction and speed, observation
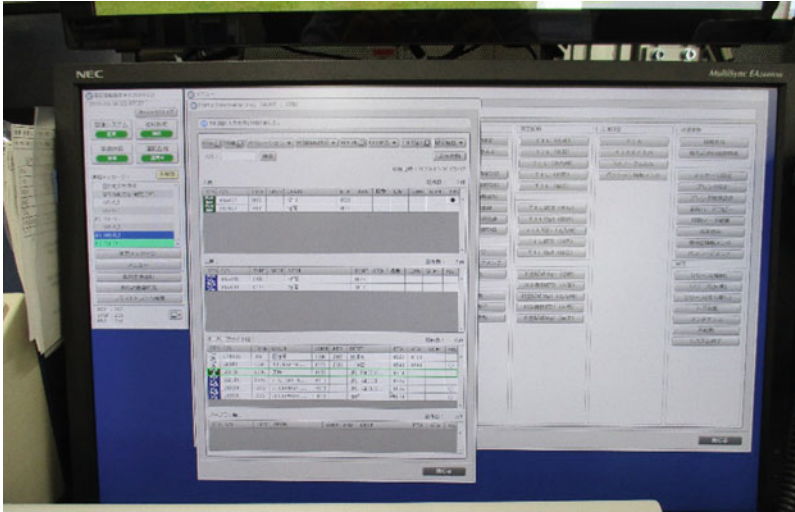
Fig. 7 FACE system display has a role of flight plan display and counter of actual traffic numbers

(METAR, SPECI) data, SCAN, and RVR (runway visual range). As shown in Fig. 8, wind direction and speed are shown for each runway with instantaneous readings as well as maximum and minimum values (ranges) over a period of time. The format can be customized as circles and bars. The airport name is shown at top left, and text information can be display in two windows on to right of the wind display. In Fig. 8,

Fig. 8 TDU Display shows meteorological information data

Fig. 9 Examples of the MPID display can show AIP such as airport map and charts and status check list of operation

METAR/SPECI information are displayed in the upper text window, and RVR and other text information is displayed in the lower.

3.5 MPID (Multi-Purpose Information Display)

The MPID can display static aerodrome information found in the Aeronautical Information Publication (AIP), including airport layout and instrument procedure charts (SID, STAR), and real-time runway and airport status information, as the examples in Fig. 9 show. Electronic AIP information retrieval is faster than using paper charts. The runway status function serves to remind the operator when the runway may be unavailable due to for example maintenance or inspection.

4 Remote AFIS Operational Challenges

4.1 Effect of ITV System

Aerodrome AFIS is an advisory rather than a control service, and can be provided even when visual observation of the airport is not possible. This may be perfectly adequate for airports with little scheduled traffic, and procedures ensure safety. However, as traffic increases, greater efficiency becomes necessary. Task and situation analysis has confirmed the effectiveness of using ITV video images to allow the RAG operator to positively understand the traffic situation and increase efficiency. Giving an example: An RAG operator observing the ITV monitor noticed a delay in preparations for the departure of a flight and was able to advise the departure controller of the delay instead of the pilot, which reduced the flight crew's workload and prevented unnecessary communication. In the case of a system that relies only on radio communication, actions must be triggered by receiving a communication and are reactive, and workload and communications increase in order to understand the context of an issue and take appropriate measures. [ENRI internal report (in Japanese), unpublished] On the other hand, ITV observation enabled the RAG

operator to take proactive steps without excess communication which mitigated an increase in workload.

4.2   Multiple Operation and Issues in Remote AFIS (Double Contact)

A common problem encountered when a single operator provides services to two airports simultaneously is a "double contact", meaning that the aircraft at each airport call the operator almost at the same time. This is possible since the two airports are operating simultaneously on different frequencies. The basic method of handling a double contact is to instruct one of the two aircraft (the one that made contact later) to "Stand by". However, this means that one party is kept waiting and service performance deteriorates even though the airport has low traffic.

A more efficient way of handling a "double contacts" is for an operator working the adjacent RAG console to assist if their workload permits. As shown in Fig. 2, the RAG consoles and operator seats are arranged in a row. Operators may be able to assist others when their workload permits, such as when they are not handling requests or if there are no traffic movements. If a double contact occurs and the RAG operator instructs the caller at one aerodrome to stand by while they are handling the other contact, the operator at an adjacent console may contact the caller that is standing by instead. However, in this case the assisting RAG operator must take over responsibility for providing service to the caller until it is completed. This poses risks such as the possibility of communication errors in information exchange and inconsistencies in context sharing. Frequent occurrences of "double contacts" increase risks of miscommunication and confusion. Therefore, multiple operations are not carried out at airports where the scheduled traffic flux exceeds a certain level.

5   Summary

This chapter introduced the remote operation RAG system that has been in operation for over 40 years serving small airports in Japan. The system has been repeatedly improved by upgrades such as the addition of remote video monitoring equipment at the airport (ITV) and surveillance displays utilizing airway radars (APDU). The existing RAG system was aimed at small islands and other small airports with low traffic volumes, so has been regarded as a relatively small system that can handle such traffic volume with a balance between operating cost and efficiency. However, the introduction of technologies such as "Remote Tower" is currently under consideration that will allow improved AFIS services and even ATC at small airports to be delivered remotely. The operation of a remote AFIS using a 360° panoramic display system will start in 2021.

References

Airports in Japan. (2020). Data from JCAB. Retrieved Oct. 01, 2020, from https://www.mlit.go.jp/koku/15_hf_000124.html.

AIM-j. (2020). Aeronautical Information Manual Japan, compiled by AIM-Japan Editorial Association, supervised by Japan Civil Aviation Bureau, Japan Meteorological Agency.

ICAO Annex 11. (2018). Air Traffic Services. 15th Edn.

Remotely Operated Aerodrome Flight Information Service. (2012). AN-Conf/12-WP/130 23/10/12, ICAO Working Paper, Twelfth Air Navigation Conference Montréa.

Development and Field Testing of Remote Tower Prototype

Remote Tower Experimental System with Augmented Vision Videopanorama

Norbert Fürstenau and Markus Schmidt

Abstract The goal of the research described in this chapter was the development and setup of an initial experimental version of the "Virtual Tower" with focus on replacement of the direct view out-of-windows. Specifically, the intermediate step of a Remote Tower Operation (RTO) work environment for remote surveillance and control of small airports is described which served for verifying the main functionalities. A structured work and task analysis detailed the requirements on the new human–machine interface (HMI) and emphasized the "far view" out of the tower windows as important information source. Consequently a digital high-resolution videopanorama system was implemented as central HMI-component to replace the airport tower out-the-window view. Field tests using this reconstructed panorama indicated the effective visual resolution for object detection to show reasonable agreement with theoretical predictions under ideal conditions. As addition to the panorama an integrated zoom function provided an enlarged narrow-angle "foveal" component by means of a remotely controlled pan-tilt-zoom camera with tracking functionality. The digital reconstruction of the far view allowed for integration of "video-see-through" augmented vision features by integration and superposition of e.g. weather and electronic surveillance data, and it allowed for video replay of stored surveillance information.

Keywords Remote airport traffic control · Human machine interface · Work analysis · Video panorama · Augmented vision

1 Introduction

This chapter provides results of the initial phase of Remote Tower research at the German Aerospace Center (DLR), starting with the "Visionary Projects" study

N. Fürstenau (✉) · M. Schmidt
German Aerospace Center, Institute of Flight Guidance, 38104 Braunschweig, Germany
e-mail: norbert.fuerstenau@dlr.de

M. Schmidt
e-mail: markus.schmidt@dlr.de

"Virtual Tower (ViTo, 2002–2004), with focus on the project RApTOR (Remote Airport Traffic Operation Research, 2004–2008). The chapter is based on a number of previous publications, (Fürstenau, 2004; Schmidt et al., 2006, 2007; Fürstenau et al., 2008a, b, 2011) and on the initial concept outlined in the Virtual Tower patent (Fürstenau et al., 2008a, b).

The growth of air traffic, the increasing use of small airports by low cost carriers, and the requirement for cost reduction of air traffic control have pushed the search for new solutions to increase efficiency of air traffic control. For traffic control on the airport movement areas and within the control zone the Virtual Tower idea was put forward by DLR since more than ten years, based on earlier suggestions by Kraiss & Kuhlen (1996) and Ellis (see the Foreword and references therein, and fundamental considerations in Ellis (1991). Specifically for small airports remote tower operation (RTO) with a new type of Remote Tower Center (RTC) provides the potential for multiple airport control from a single control room. Corresponding RTO/RTC-prototypes have been developed since about 2004 and are presently being tested under operational conditions.

RTO/RTC is considered as intermediate step towards the Virtual Tower for larger hubs, as a new kind of airport traffic control center without the need for an expensive tower building. In contrast to the small low-traffic airports large hubs usually rely on electronic surveillance (surface movement radar SMR) and so called (Advanced) Surface Movement Guidance and Control Systems (A-SMGCS) which support and partly replace the visual surveillance. Corresponding RTO and Virtual Tower projects were started in Europe (Germany: DLR/DFS, Sweden:Saab/LFV), Canada (Nav Canada and Searidge Technologies), and the US (Vogel, 2009; Hannon et al., 2008) (see also chapters "Visual Features Used by Airport Tower Controllers: Some Implications for the Design of Remote or Virtual Towers", "Detection and Recognition for Remote Tower Operations", "Remote Tower Research in the United States", and "The Advanced Remote Tower System and Its Validation").

A number of tower work and task analyses performed during recent years (Pinska, 2006; Tavanti, 2006; Werther & Uhlmann, 2005) partly accompanied by model based simulations of controller's decision processes (e.g. Werther & Schnieder, 2005; Werther et al., 2007) determined the importance of visual surveillance for creating the controllers situational awareness. In the tower work environment of large airports the permanent shifting of attention between far view and displays contributes to workload and generates head-down time problems (Pinska, 2006). Both may be reduced by augmented vision systems such as transparent head mounted or head-up displays which superimpose traffic or weather data on the out-of-windows view (Fürstenau et al., 2004; Schmidt et al., 2006; Peterson & Pinska, 2006). Consequently it was concluded that the digital reconstruction of the far view of the control tower by means of a high resolution videopanorama with a kind of video-see-through augmented vision element (Barfield & Caudell, 2001) will be an important component of the human-system interface in a future towerless work environment and it will support the acceptance of the controllers operating remote towers. The concept of a high resolution video panorama as potentially low cost human machine interface (HMI)

for replacement of the direct view out-of the tower windows is supported also by the fact that small airfields usually lack any advanced electronic surveillance.

A corresponding first experimental system was realized at the Braunschweig Research Airport (BWE) within the RApTOR project around 2004–2005, based on a 180° live video-reconstruction of the tower out-of-windows view (Schmidt et al., 2007). Video see-through augmented vision features were realized by integrating information from real-time moving object detection and from Mode-S multilateration via feeding transponder data into the reconstructed far view.

Initial verification of the basic RTO-design features was performed under realistic conditions using a DLR test aircraft for field testing (Fürstenau et al., 2008a, b; Schmidt et al., 2009). This kind of testing is a costly undertaking and easily exceeds a project budget. That is why a number of questions regarding the requirements, performance and acceptability of the new RTO-controller working position (CWP) was investigated in a specific Remote Tower simulation environment as extension of DLR's conventional Tower Simulator (see Sect. 5, part III of this book and Papenfuss et al., 2010). Naturally, many questions regarding the performance and acceptability of the video-based panorama reconstruction including zoom functions usually can rely on field tests only because no simulation is able to reproduce the reality in full detail. Nevertheless some usefull predictions and estimates can be derived also from appropriate theoretical considerations presented in Sect. 3.1 of the present chapter, and in chapters "Detection and Recognition for Remote Tower Operations", "Remote Tower Prototype System and Automation Perspectives", "Model Based Analysis of Two-Alternative Decision Errors...", "Model Based Analysis of Subjective Mental Workload..." and in Appendix A - D for the technical design, data analysis, workload modeling.

Section 2 reviews results of a structured work analysis, followed in Sect. 3 by a detailed description of the design and technical description of the augmented vision video panorama system realized within DLR's first RTO project RApTOR as basic experimental environment for field testing. Results of the initial field trials for verifying relevant performance parameters are reported in Sect. 4. In Sect. 5 a brief overview of the simulation environment is presented while details and simulation results are described in part III of the book. Section 6 provides a conclusion and outlook.

2 Work Analysis

In this section we will briefly review the initial work analysis which accompanied the basic RTO design and development as described in Sect. 3. A cognitive work and task analysis (CWA) was performed by means of structured interviews of domain experts (controllers) from medium sized airports (Werther & Uhlmann, 2005) which followed a method described by Vicente (1999). He separates the analysis into five levels, ie. analysis of: 1. work domain, 2. control task, 3. strategy, 4. social organization and cooperation, 5. operator competency. The latter was not considered in this

context because controllers due to the rigid selection process, highly specific training and formalized detailed work procedures may be considered a very homogenously qualified group of operators.

The formalised results provided the input data for a Formal Airport Control Model (FAirControl), developed for the computer simulation of the controller decision making processes at the tower work positions which supported the interviews (Werther & Uhlmann, 2005; Werther et al., 2007). In (Werther, 2006; Werther & Schnieder, 2005) it was shown how the results of a CWA can be transferred into an executable human machine model, based on Colored Petri Nets (CPN) for simulating the controllers work processes in relation to the airport processes. The formal model allowed for evaluation of different variants of work organization, and supported the design of the new work environment and the monitoring of psychological parameters, e.g. uncovering of reduced situational awareness.

The model was separated into submodels for the human agent (controller), interaction, and the traffic processes. The interaction model defines the controller-process interactions and includes sub networks for description of information resources, such as radio communication and visual perception of the traffic situation. The executable model with graphical representation of the controlled traffic process was useful in identifying the controllers' strategies in task organization and pursuance of goals. It supported the communication between domain experts and system developers by simulating different traffic situations to establish a basis for the structured interviews. Details can be found in previous publications, e.g. Werther et al. (2007).

One major focus of the repeated and model supported interviews of two tower supervisors concerned the visual information from the outside view. The following list summarizes the most important visual information ordered by area/distance (rating = 5 from a scale of 1 (= not important) to 5):

1. Approach-/ Departure Range (2–3 km, max. 5 km)

    • Recognition of A/C, direction of movement

2. All Airfield Areas (Taxi, Apron, Stand)

    • Recognition of all active objects
    • (A/C, vehicles, humans, animals)
    • Classification of A/C
    • Recognize Smoke at A/C (e.g. turbine fire)

3. Runway Range (800-1500 m, max. 2 km)

    • Observe Runway state, detect aircraft parts

4. Taxi Area (500–900 m, max. 2 km)

    • Recognition and position of passive objects
    • (A/C and parts, vehicles, obstacles)

5. Apron Area (200 m)

    • Recognize aircraft damage

6. Stand Area

    - Recognize Aircraft damage
    - Recognition and position of passive objects
    - (luggage, vehicles)

7. RWY/Taxiway Lights

    - Monitor Intensity
    - Monitor Function

Based on the CWA framework the information sources, work constraints, control tasks and decisions of two controller working positions (CWP) at a medium size airport tower (tower and ground executive controllers (TEC: responsible for landing and starting aircraft, GEC: responsible for ground traffic/taxiing)) were analysed. The contribution of information gathered through direct view out of the tower window (visual information) was of special interest. The information gained for three visibility conditions (normal vision, night vision, limited vision <2 km) were identified and rated according to their relevance. All decision tasks of the TEC- and GEC-CWP's were analysed. A total of 60 decisions ($N_{GEC} = 29; N_{TEC} = 31$) were modelled following this template (Papenfuss & Möhlenbrink, 2009).

Four main issues were found in this analysis: (1) small regional airports usually got no expensive electronic surveillance, leaving visual observation as the main information source on the live (traffic) situation on the airport surface. (2) safety relevant information like foreign objects on the runways or bird swarms can only be sensed—if not through direct visual surveillance—by expensive sensor infrastructure that is unlikely for a small airports. (3) the controllers' acceptance for a remote working place is expected to be higher if the working procedures and the look-and-feel remain as similar as possible to the known working procedures. (4) The reconstruction of the far view via digital video enables the augmentation of the video panorama via information superimposition, e.g. to reduce head down times.

The frequency (F) of use of different information sources over the whole spectrum of possible control task decisions derived from the CWA for the GEC and TEC controller working positions quantify their information requirements (Werther & Uhlmann, 2005; Schmidt et al., 2009). Oral communication via radio ($F = 97\%$) is the most often used information source, followed by weather information ($F = 35\%$), telephone ($F = 33\%$) and the direct view out of the window ($F = 21\%$). The latter number may be compared with values reported in the literature which vary between 20 and 50%, depending on airport class and CWP (for an overview and further reference e.g. Tavanti (2007).

In order select possible information for video augmentation a further analysis of the detailed control tasks was conducted. The use of information sources was furthermore analysed according to the working positions (TEC, GEC) and the kind of traffic that is controlled (VFR versus IFR). The availability of assistance systems, like departure coordination (DEPCOS) or extra monitors depends on available information (e.g. surface movement radar, usually only on large airports) and is very specific for every single airport (size, traffic amount). Compared to the GEC the

TEC uses more different information sources over all tasks ($N_{GEC} = 12; N_{TEC} = 16$) and more information sources in single tasks (mean values $\mu_{GEC} = 3.4$; (*std.dev.* = ±1.0); $\mu_{TEC} = 4.6$, (*std.dev.* = ±1.6)). TEC combines more information sources to achieve an integrated picture of the traffic. The analysis also revealed typical information source access profiles: whilst GEC mostly uses communicative items like radio ($F_{GECradio} = 97\%$) and telephone ($F_{GECtelephone} = 55\%$), the TEC, after the radio connection ($F_{TECradio} = 97\%$) to the pilot most often uses the radar ($F_{TECradar} = 55\%$), control strips and flight plan information ($F_{TECplan} = 35\%$), and weather information ($F_{TECweather} = 45\%$).

The quantitative analysis of information source usage showed that different information needs depending on working position and character of traffic can be identified. In particular the percentage of VFR traffic is significantly higher on regional airports as compared to international ones. The analysis showed that decisions for VFR traffic compared to IFR traffic required more often weather information ($F_{VFRweather} = 45\%, F_{IFRweather} = 25\%$) and control strips information ($F_{VFRstrips} = 31\%, F_{IFRstrips} = 11\%$).

One advantage of using a digital videopanorama as core of the RTO-HMI is the possibility of augmenting the far view with additional information. For the control of small regional airports with a lot of VFR traffic the augmentation of weather information can help to reduce head down times of the controller. Furthermore information normally saved on flight strips can be attached to the aircraft position on the video display. Through analysing the information needs of controllers in different situations a framework for the design of a work environment that reduces workload by integrating information and by adding automation can be achieved. Carefully added automation, such as an assistance system to reliably detect and signal moving objects for monitoring tasks, can support the controller and allow for the simultaneous control of several airports from a Remote Tower Center (RTC, see chapters "Multiple Remote Tower Simulation Environment", "Model Based Analysis of Subjective Mental Workload during Multiple Remote Tower Human-in-the-Loop Simulations", and "Planning Remote Multi-airport Control—Design and Evaluation of a Controller-Friendly Assistance System").

The work analysis outlined above with regard to visual surveillance may be compared with the comprehensive overview and discussion of cues presented in chapters "Visual Features Used by Airport Tower Controllers: Some Implications for the Design of Remote or Virtual Towers" and "Detection and Recognition for Remote Tower Operations" of this book.

3 Experimental Remote Tower System

In this section the development of the initial experimental RTO system at Braunschweig research airport is described. Motivated by the important role which visual information plays for the tower work processes according to the work analysis, in particular at smaller airports, a high resolution digital video panorama system with

augmented vision functionality was developed as outlined in DLR's basic Virtual Tower patent (Fürstenau et al., 2008a, b). It served as experimental environment for investigation of different aspects of the Remote Tower Operation and RTCenter concept and for development of a prototype demonstrator described in the following chapter "Remote Tower Prototype System and Automation Perspectives". This initial experimental system was used for verifying by field testing basic design features as realized within the DLR project RApTOR (2004–2007).

3.1 Optical Design and Expected Performance

The design of the experimental video panorama system described in the following Sect. 3.2 was based on the assumption that a digital video reconstruction with a visual resolution comparable to the real view out of the tower windows in combination with a PTZ camera for providing a binocular function should be sufficient to fulfill the requirements as derived from the work analysis reviewed in Sect. 2.

Performance predictions of the visual system were based on the assumption that signal delay (small latency effects), optical resolution, contrast and dynamic range are the most important parameters characterizing system quality, i.e. detectability of (moving) objects and discriminability of relevant operational situations. The effective resolution is largely determined by visual contrast which is discussed in more detail in Appendix A. With regard to aircraft detection and recognition basic work was published by Watson et al. (2009). Here we provide an (optimistic) estimate of the ideal resolution corresponding to the Nyquist limit of the modulation transfer function (MTF, see Appendix A of this volume). The MTF quantifies the contrast dependent resolution as dependent on the spacial frequency, i.e. the frequency of periodic black-white line pairs. The Nyquist limit is that spacial frequency value represented by the lowest discriminable light–dark spacial object-wavelength transmitted to the observer by the camera-monitor system. Of course this is an ideal value which is given by the pixel size and distance, i.e. the pixel resolution. In fact, not surprisingly the validation of the prototype panorama system (see the following chapters "Remote Tower Prototype System and Automation Perspectives", and "Which Metrics Provide the Insight Needed?") showed that under realistic environmental conditions this ideal resolution limit quite often will not be achieved.

The optical resolution and signal-to noise ratio depends on the technology and parameters of the image sensor (CCD or CMOS technology, sensor size and number of pixels), the focal width and quality of the selected camera objectives, which also determine the achievable contrast (through the MTF, see Appendix A of this volume). For the panorama system a compromise between achievable resolution and number of cameras, i.e. cost was made. Concerning size (=diagonal) of image sensor (horizontal/vertical width H/V) and focal width f the design criteria may be derived from the curves in Fig. 1 which are based on the fundamental (thin lens) approximation between field-of-view FOV = 2Θ and f (valid for large distances, where the image is close to the focal point at f:

Fig. 1 Horizontal field of view (FOV) as function of focal width f of the objective for panorama camera with 1"-image sensor, f = 12.5 mm (left endpoint of the curve) and PTZ-camera with f = 3.6–82.8 mm and ¼" image sensor
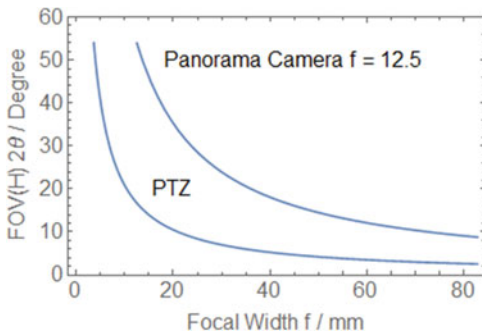
$$\theta = \arctan\left(\frac{H}{2f}\right) \tag{1}$$

with half angle Θ of FOV on the vertical axis.

For a 200° panorama four cameras require per camera FOV ≈ 50°. With 1" image sensors (sensor size H × V = 12.8 × 9.9 mm) this determines the required focal width of the wide-angle objective f = 13.7 mm. The graphic depicts the actually used commercially available objective with f ≈ 12.5 mm (FOV = 54°), the left end of the upper curve. The PTZ camera specifications were: 1/4" image sensor, H × V = 3.2 × 2.4 mm, with zoom objective f = 3.6–82.8 mm. The corresponding FOV range as depicted in the graphics varies between 53°–2.2°.

From (1) the angular pixel resolution of the panorama may be estimated within the paraxial approximation (for details see also Appendix A1; $\Delta H = p < < f$) as $2\Theta = \alpha \approx 2$ arcmin $= 0.033°$, with ΔH = pixel size p = 7.5 μm (+ 0.5 μm gap). This may be compared with the diffraction limited value of the human eye (≈1 arcmin, e.g. (Bass, 1995), see also Appendix A). Towards the edge of the image sensor the pixel FOV decreases which reduces the received light power per pixel accordingly. This in turn would reduce the contrast towards the edges of the image and add to a number of other image degrading effects that are corrected more or less with high quality lens systems (for some additional details see Appendix A).

In other words, by using the fundamental relationship G/B = (g/f–1) ≈ g/f derived from Newtons (thin) lens equation as a paraxial estimate, with g = object distance, G = object size, B = image size ≈ p, the minimum vertical object size at g = 1 km distance corresponding to 1 camera pixel p is G/p = 0.6 m/1 Pixel, or again ca. 2 arcmin angular resolution. With the given vertical camera position we get 1 m/1 Pixel along the line of sight. This optimistic estimate is idealized and does not include limitations due to contrast of cameras and display and possible image distortions and diffraction effects with small lens aperture stop. It represents the Nyquist limit of the modulation transfer function (MTF) which states that as minimum condition for resolution at least two pixels of the image sensor have to cover the distance (spacial wavelength λ_{min}) of an alternating dark–light pattern (i.e. 2 measurements = pixel distance = λ_{min}). For some more details see Appendix A of this volume.

Moreover the observable resolution at the videopanorama HMI is reduced due to imperfect optics of the camera, the limited dynamic (illumination dependent) image compression, and influenced by the resolution of the display system (ideally the same as the camera) and Gamma-adjustment of camera and display (see Appendix A). The estimated resolution value of about 2' (about two times the diffraction limited value of the human eye) may be approached with decreasing camera aperture D (increasing depth of focus, reducing lens-imaging errors), which is of course possible only under good light conditions and sufficient object–background contrast. Furthermore, under bright illumination (bright sunlight) the automatic aperture control of the lens decreases aperture (increases aperture f-number f#f/D, see Appendix A) which may start decreasing resolution if f# > ca. 6. This decrease is due to diffraction effects originating from the wave nature of light leading to blurring of the idealized point focus. The aperture is focused into a light disc diameter corresponding to the so called Airy-radius $q_1 = f\#/\mu m$ (for wavelength $\lambda = 0.6\ \mu m$ (green), see Appendix A) of the point-spread function that exceeds pixel size $p \approx 8\mu m$ with decreasing aperture diameter (increasing f#).

For realization of the panorama only 1424×1066 Pixels of each camera (50° horizontal viewing angle) are actually used in order to match the 180° panorama angle. For improving the pixel-resolution to match the human eye, the focal width would have to be doubled, resulting in a fourfold number of cameras (for covering roughly the same horizontal and vertical FOV). This of course would also mean a multiplication of the system cost (including data processing and and HMI) by a comparable factor.

From the above discussion it becomes evident that RTO-controllers require a zoom camera (with reduced FOV, however increased pixel resolution) not only as equivalent for the binocular but also to compensate for the limited videopanorama resolution. The theoretical (idealized, paraxial) angular pixel resolution of the PTZ-camera used for the experiments below is given by $\alpha_Z \approx p_H/Z\ f_0$, yielding $\alpha_Z = 1$ arcmin ($Z = 4$, viewing angle $2\Theta = 15°$) and $\alpha_Z = 0.2$ arcmin ($Z = 23$, viewing angle $2\Theta = 2.5°$), with $p_H =$ horizontal pixel size $= 4.4\ \mu m$, $f_0\ (Z = 1) = 3.6$ mm, f_{max} ($Z_{max} = 23) = 82.8$ mm. With larger Z of course als a conventional binocular function is obtained (although limited by the individual MTF and Gamma-adjustment of the individual camera-display systems, see Appendix A).

3.2   Digital Reconstruction of the Out-of-Windows View

A block diagram of the initial augmented vision video panorama system is shown in Fig. 2. The basic sensor component in the initial design consisted of four high resolution (1600×1200 pixels) high dynamic range (14 bit/pixel) CCD cameras ($P_{1, 2, 3, 4}$) covering the Braunschweig airport runway area within 180° viewing angle, complemented by a remotely controlled pan-tilt zoom camera (P_5: PTZ).

The cameras (upper photo in Fig. 2) are positioned 18 m above the airport surface, horizontally aligned on top of a building at the southern boundary of the airport, ca.

Fig. 2 Schematic block diagram of augmented vision video panorama system as set up in 2005 (initially without IR camera), reprinted from (Fürstenau et al., 2011) with permission. Arrows indicate flow of information with GBit fiber-optic data link between sender and receiver. Wide light-blue arrow indicates visual information for the controller. 180°—FOV RTO videopanorama shown in the bottom photo: panorama version with backprojection displays (see Fig. 5). Compressed angular arrangement (ca. 125°) for cameras no. 1–4, PTZ display and touch-input interaction display integrated in the controller console
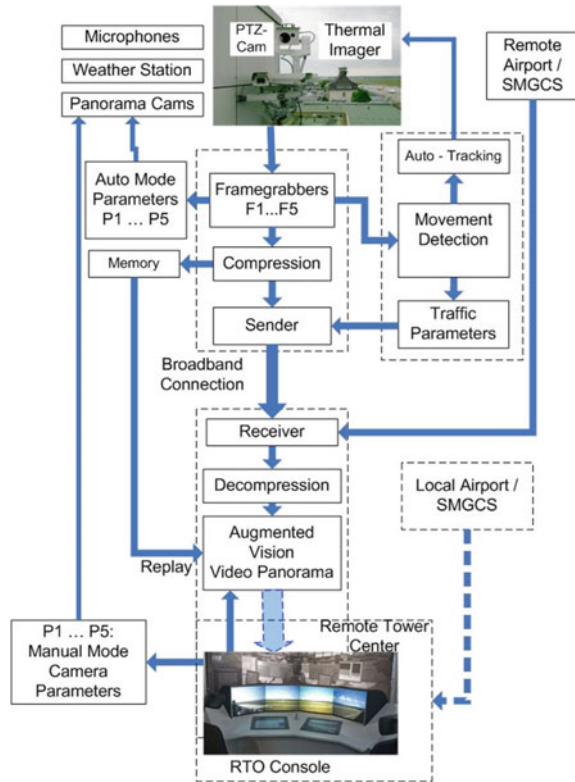
100 m east of Braunschweig tower, 340 m south of the main runway 08/26 (1670 m, until extension to 2500 m after 2008). The vertical aperture angle of about $\pm 20°$ (with respect to the horizontal line of sight) allowed for a closest surveillance distance of about 60 m and about 365 m observation height at 1 km distance or ca. 125 m above the runway. The latter value was critizised lateron by domain experts during a more detailed requirements analysis as being too low. This resulted in a re-design for the validation experiments described in the following chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation". Upon request of several domain experts the visual system was extended by (stereo) microphones at the camera site and a digital connection to loudspeakers at the controllers console.

For each camera (framerate = 25 frames/s) the signals are split into two outputs, according to (Fürstenau et al., 2008a, b). One feeds the data compression and encryption (AES256) for transmission to the RTO HMI, while the other drives the simultaneous real time image processing for movement detection. The five recording PC's with the compression software near the camera position allow for storing panorama and zoom data (roughly 500 GByte of data per day) and provide the possibility of complete panorama replay. Figure 3 shows an aerial view of the Braunschweig
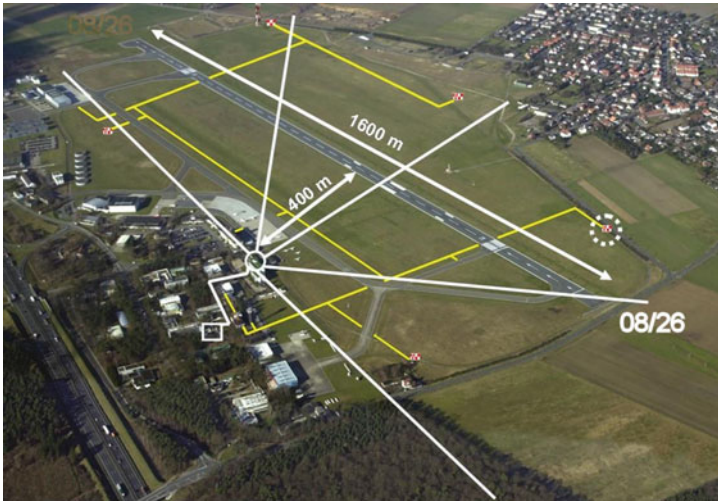
Fig. 3 Braunschweig research airport BWE (2005, runway extended to 2.5 km after 2008) with 1.67 km runway 08/26 extending E-W, fiber optic data link (thin yellow lines) connecting sensor containers (enlarged, one with broken circle) used for measuring static resolution. Circle with radiating lines indicates panorama camera position and field-of-view respectively. Reprinted from (Fürstenau et al., 2011), with permission

research airport from S-E direction indicating camera position and camera viewing sectors.

A GBit ethernet switch feeds the images from the five sensors into a single mode fiber optic data link which transfers the typically 100 MBit/s data (night + day average) of the panorama system and PTZ over a distance of 450 m to the visualisation system. A second GBit ethernet switch splits the incoming data into five output channels for decompression, with one PC per camera in the initial setups. The PCs also synchronize the displays of the four segments. Each camera is remotely controlled with respect to aperture and γ-correction (see Appendix A). The PTZ camera is controlled with respect to azimuth, vertical angle and zoom (Z = 1–23-fold, focal width 3.6 mm–82.8 mm, corresponding to 54°–2.5° visual angle).

The Augmented Vision Videopanorama (AVP-) HMI for a single operator/single airport surveillance in an early version is depicted in Fig. 4 (Fürstenau et al., 2007). It was based on four 21"-LCD-monitors (UXGA, 1600 × 1200 Pixels) for displaying the reconstructed panorama and a separate one for display of the remotely controlled PTZ-camera.

The monitor frames may be considered a disadvantage of this realization of the video panorama, although the resulting discontinuities were not seen as a major negative aspect by the domain experts during the system validation in human-in-the-loop simulations and field tests. An alternative videopanorama setup is shown in Fig. 5 (Schmidt et al., 2009). It allowed for seamless stitching of the single camera images without disturbing display frames. This backprojection system used a custom

Fig. 4 Early version of 180°-videopanorama (1600 × 1200 pixel) with pan-tilt zoom camera display on top, demonstrating augmented vision function: superimposed transponder label tracking the landing aircraft using automatic movement detection (see Sect. 3.4, Fig. 7). Two pen-touch input displays integrated in the console display additional information required by the controller: the one on the right was used for electronic flight strips (flight data) and control of the camera parameters, PTZ, weather information, and tracking. Initially published in (Fürstenau et al., 2007)

Fig. 5 Experimental RTO-Videopanorama (180°) backprojection system. It was installed in the RTO-simulator environment (see Sect. 5) and could display simulated as well as live traffic (the latter shown in the photo). PTZ display is integrated in the console (right side) beside the interaction and camera control display on the left (Sect. 3.3). Reprinted from (Schmidt et al., 2009), IEEE with permission.

made compact construction with video projectors generating images on 31" screens, with 1400 × 1050 pixels, 3500 ANSI lm, and contrast of 3000:1.

This version (initially published in Schmidt et al., 2009) was used as an initial RTO-extension of the DLR tower simulator (see Sect. 5 of this chapter and chapter "Multiple Remote Tower Simulation Environment" of this book). Besides displaying the live video panorama the interface could accept video and simulator output signals. This allowed for using the video panorama display system without changes for performing human-in-the-loop experiments with domain experts.

3.3 Videopanorama Interaction and Control Display

Interaction of the operator with the panorama system (panorama cameras, PTZ, weather station, microphones) was performed via a pen touch-input display which can be seen integrated in the consoles of Figs. 4 and 5. Figure 6 shows details of the display.

In order to obtain a compact RTO operator HMI which, e.g. should fit into a typical tower environment of a medium size airport, the pen touch-input display is designed to incorporate video panorama control features as well as traffic information, e.g.

Fig. 6 Initial design of pen touch-input interaction display (reprint from: Schmidt et al., 2009), IEEE with permission). Minipanorama on top as option for PTZ positioning via touch (actal position visualized by yellow square). Center: electronic flight strips; right, from top: PTZ control buttons, virtual joystick as another option for PTZ positioning, weather data with wind direction (from weather sensors); left: optional buttons, e.g. for preset PTZ position

electronic flight strips. A mini-panorama at the top is updated with 5 Hz and serves for commanding the PTZ-camera orientation via pointing of the touch-pen. The display also contains buttons for optical PTZ-parameters and activation of automatic object tracking via movement detection, a virtual joystic as an additional option for commanding PTZ orientation, and weather data.

For PTZ positioning the target can be defined manually or by automatic movement detection. A yellow square is positioned at the respective location of the panorama, defining the target area to be zoomed-in. With the tracking mode turned on the square moves coherently with the corresponding object, aftoer tuching it on the display. An algorithm for real time movement detection is running on a separate parallel processor of the image compression PCs of each camera (see next Sect. 3.4).

3.4   Augmented Tower Vision and Movement Detection

One of the design goals was the minimization of the number of additional interaction systems and displays and for improving low visibility conditions by integration of additional sensor data (e.g. for IR see chapter "Designing and Evaluating a Fusion of Visible and Infrared Spectrum Video Streams for Remote Tower Operations" and for Multilateration see chapter "Integration of (Surveillance) Multilateration Sensor Data into a Remote Tower System") and relevant traffic information into the videopanorama by using augmented vision techniques.

Augmented vision as defined for 3D-virtual reality systems discriminates between so called "optical see-through" and "video see-through" systems (Barfield & Caudell, 2001). With optical see-through information displayed e.g. in a transparent head-mounted or head-up display is superimposed on real world objects whereas with video see-through the real world is displayed as a digital video image with relevant data digitally superimposed. The advantage of the latter option is zero latency time between environment and superimposed information whereas fast image processing with object- and head-tracking and minimization of latency effects by predictive filtering is required for appropriate superposition in optical see-through systems.

Within the video panorama real-time aircraft position information is integrated as obtained from the (radar based) multilateration system at the Braunschweig airport via the aircraft (a/c) transponder. An example is shown in the inset of Fig. 7 with an enlarged section of display no. 4 (looking east) depicting a yellow transponder code with multilateration position attached to the landing aircraft. It indicates a/c position on the approach glide path. Under reduced visibility this Augmented Tower Vision (ATV) feature allows for localizing the A/C near the correct position because the transponder code, A/C label and numerical information are integrated near the nominal object image location in real time.

Another example of augmented vision data is the integration of GPS-ADS-B position information transmitted via transponder. An example is shown in Fig. 7

Fig. 7 Part of the initial version of the panorama display with PTZ-display on top. Inset: Screenshot of camera # 4 (=East) display showing augmentation during landing (reprinted from Schmidt et al., 2009, IEEE with permission). Superimposed glide path (violett, added for replay), GPS-trajectories (red), live multilateration position (yellow, from transponder) and automatic movement detection (red square with object number)

where D-GPS data measured during flight testing (see Sect. 4) are superimposed (off-line) on the video in the form of flight trajectories (red) that, after geo-referencing are transformed from geographical into display coordinates.

Contours of the movement areas and the 3°-glide path are superimposed on the videopanorama for guiding the operators attention during low visibility or nighttime to those areas where moving vehicles are expected. Movement areas are also the preferred targets of a high resolution (640×512) infrared camera system with PTZ function operating in the mid-IR range (2–5 μm) which was integrated into the experimental system in a later phase for investigating improved night vision and visibility under CAT I conditions (see chapters "Remote Tower Prototype System and Automation Perspectives"" and "Designing and Evaluating a Fusion of Visible and Infrared Spectrum Video Streams for Remote Tower Operations" of this book).

For image processing two different strategies were followed, with the initial goal of automatic object tracking with the PTZ camera via movement detection: (a) hardware implementation of algorithms on FPGA's, (b) software processing with a second processor of the video-compression PC (at the sensor system location).

An initial version of automatic moving object tracking with the PTZ–camera was realized by method (b) with a simple video–frame difference method for object-movement detection (for more sophisticated approach see chapter "Remote Tower Prototype System and Automation Perspectives", Sect. 4). In practice an update rate of 5 Hz was used although theoretically 20 Hz was estimated to be achievable. The

Fig. 8 Object detection (aircraft in cloudy sky) using automatic background subtraction (contributed by DLR Unit Optical Information systems, W. Halle)
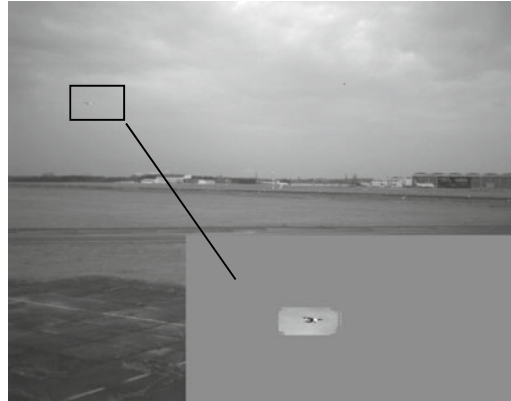
basic performance of the software based movement detection algorithms could be verified with automatic PTZ-tracking activated on the interaction display. It demonstrated the practical usefulness of this feature, however with limited reliability due to relatively simple algorithms based on image subtraction and texture analysis of detected clusters. An example is depicted in Fig. 7 (inset) with the numbered frame surrounding the approaching aircraft.

An advanced approach (using strategy (a), see chapter "Remote Tower Prototype System and Automation Perspectives", Sect. 4) for automatic object detection and discrimination, e.g. between cars and aircraft on the movement areas, or between aircraft and birds in the sky was described already in the Virtual Tower patent (Fürstenau et al., 2008a, b). The challenge in the present context for achieving a reliable performance lies in the fact that different algorithms have to be implemented for flying objects in the air with a dynamic background (moving clouds) and for moving objects on the ground with a more stable background. An example based on automatic background subtraction is shown in Fig. 8. Detection relies on the combination of different criteria: 1. speed (a/c faster than clouds), 2. a/c texture different from clouds.

3.5 Triangulation

In order to determine the position of aircraft on small airports without electronic surveillance a stereophotogrammetry method (triangulation) was investigated for analysing the high resolution videos. An additional camera was positioned at a distant position and it was used for observing the same field of view of one of the fixed panorama cameras (realizing a stereo system). In order to achieve the required spatial resolution it is necessary to choose the appropriate base length between both cameras. The analysis and initial tests showed that for a sufficient position resolution the second camera of this stereo system has to be placed at the opposite side of the airport.

With a baseline of 450 m (ca. position of nearest multilateration system container in Fig. 2) theoretically the position of an aircraft at a distance of three kilometers can be obtained with an accuracy of 12 m (1.5 m at 1 km distance, based on pixel resolution). In the present case only preliminary field test were performed with a reduced camera baseline of ca. 20 m so that the additional stereo camera could be placed on the roof of the same building as the videopanorama sensor system that provided the first camera of the pair of stereo pair.

A video processing framework was realized for this purpose in order to unify and simplify the design and development of the video processing loops. Based on this framework a heterogeneous object tracker for various image sections and object types detected the different objects. The tracker worked on each stereo camera separately. For triangulation of the object position corresponding object pairs in both images have to be identified automatically. In order to retrieve the desired accuracy, the cameras were time synchronized and calibrated with regard to interior and exterior orientation. This requirement was fulfilled already for the panorama camera system with augmented vision function which was based on the transformation from geographical into display coordinates. However, this experimental triangulation system allowed only for initial proof of concept. The tests showed that for a larger measurement campaign allowing for reasonable quantification of performance in the approach and departure direction, the viewing angle between runway and line-of-sight (and in particular the angle difference between cameras) would have to be increased significantlythe by extended baseline which was not within the scope of the initial RTO project.

4   Field Testing for Verification of System Performance

In this section we present results obtained by field tests, in particular flight testing with the DLR DO-228 test aircraft for verifying the theoretically expected system performance. The main question to be answered refers to the comparability of the video panorama with the real view out of the tower windows. For this purpose different experiments and measurements were performed for determining signal latency and visual resolution under realistic environmental conditions The results allowed for an initial experimental estimate of the effective (subjectively experienced) visual resolution of the reconstructed far view.

4.1   Latency

In order to react quickly to critical situations domain experts require a low delay (<0.5 s) between real world events and the video reconstruction in the RTO-HMI. The video-system delay basically consists of computing time contributions from the color image construction at the camera site using an implementation of the Bayer-algorithm

and the image compression-decompression (CoDec). A special laser arrangement with beam-shutter was used to determine this time interval. The 450 m single-mode fiber optic data link allowed for feeding the laser beam via a beam splitter into one of the fibers of the installed cable at the panorama site, and use the output at the camera site for illuminating one of the cameras. The returned camera signal could be compared with the non-delayed beam from the beam splitter. An overall latency time between image acquisition and panorama visualization of 230 ms–270 ms was measured. Of course, compared with this value the delay due to camera-monitor separation is neglectible for distances up to some 100 km, given the speed of light = 300 km/ms.

In realistic long distance connections between remote airports, however, due to additional (opto-)electronic equipment with potential additional sources of delay the actual latencies should be verified for each individual situation (see chapter "Remote Tower Prototype System and Automation Perspectives" Sect. 5).

4.2 Optical Resolution: Static Measurements

With the known size and distances of static objects on the airfield it is possible to evaluate the practically achieved effective video panorama resolution as compared to the theoretical (optimistic) estimate of ca. 2 arcmin. For verification we used the red-white (1 m squares) multilateration sensor-containers at the end points of the fiber-optic A-SMGCS data network as reference objects (see Fig. 2, height and width G = 2 m). The nearest containers as captured by the NE and E-looking camera $P_{3,4}$ are located at distances g_E = 400.8 m (Ref.-Obj. 1) and g_{NE} = 588 m (Ref. Obj. 2, broken white circle) respectively.

With the above mentioned lens equation we obtain 7.8 and 5.3 pixels respectively of the camera chip covered by the container images in the vertical direction. These values correspond to a measured resolution of α_v^{exp} = 1.7 arcmin for Ref.Obj. 1 and α_v^{exp} = 1.4 arcmin for Ref.Obj. 2, which appears reasonably close to the theoretical estimate.

4.3 Performance Verification: Flight Tests

In what follows we review previously published results (Fürstenau et al., 2008a, b; Schmidt et al., 2009) which were obtained by measuring the detectability of repeated simple flight situations for determining the subjectively experienced visual resolution of the reconstructed far view. For generating statistically relevant performance data of the videopanorama system including the zoom function a flight-test plan was set up for the D-CODE to observe with controllers and non-experts repeatable scenarios for object and maneuver detection under real view and video panorama conditions. Here we report on two experiments under VFR conditions performed on two days

in May 2007, one with clear sky, one under reduced visibility (<10 km), with one pre-test in December 2006.

4.3.1 Experimental Design

Flight tests of two hours duration each, with the DLR DO-228 (D-CODE) test aircraft were performed with successive approaches, touch-and-go (or low approach) and takeoffs. Five subjects (2 controllers of the Braunschweig Tower (S_1, S_2) and 3 non-experts (S_3, S_4, S_5, members of the human factors department)) observed the flyby from a position near the panorama camera system and monitored times of 11 characteristic events e_1–e_{11}: out of sight, low/steep departure angle, take-off, touchdown, approach main/grass runway, landing gear down/up, steep approach, first sighting. The measurements were performed with notebook (touch input) computers used by each participant, using a specially developed data input software (GUI). Pilots received the flight plan for up to 16 approaches. For the trials a W-LAN with time synchronized camera and data acquisition was used. One of the GPS trajectories recorded for each flight with the onboard Omnistar satellite navigation system is shown in Fig. 9, including event observation positions $x(e_i)$ of the corresponding observation times $t(e_i)$. For the present task of determining the perceived video resolution only the six well defined events with the lowest event-time variances were used (see Table 1).
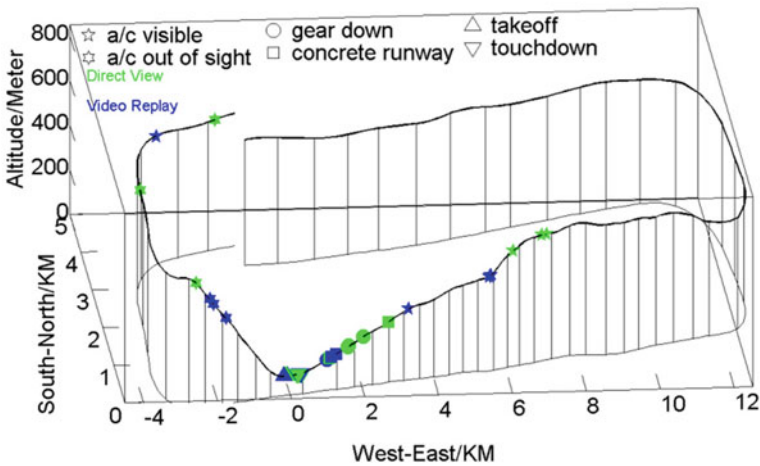
Fig. 9 GPS trajectory no. 4 out of 11 test flights during the pre-experiment on 13/12/2006 (clockwise direction). green/blue symbols represent event observations under real view (green)/ video panorama replay conditions (blue). Approach direction 260° at RWY 08/26 with touchdown near ARP at 0 km (52°19′09″ N, 10° 33′22″ E). Vertical lines = 10 s intervals on flight trajectory. Final speed ca. 100 kn

Table 1 Trial #2(May 21/07, clear view, >10 km) and #3(May22/07, cloudy, <10 km). Mean, standard deviation and std. error of event observation time difference $\Delta t = t(\text{real view})$–$t(\text{replay})$. Reprinted from (Schmidt et al., 2009), IEEE with permission

Trial #2 (clear) Event e_i	N	Mean $\Delta t/s$	S.D/s	S.E/s
e_{11}: A/C visible	54	85.1	77.9	10.7
e_8: Gear visible	42	13.0	12.9	2.0
e_6: main RWY	28	34.3	49.5	9.5
e_7: grass RWY	22	29.4	45.5	9.9
e_5: touchdown	22	+1.8	1.0	0.2
e_4: takeoff	17	+ 2.3	2.5	0.6
Trial #3(cloudy) Event e_i	N	Mean $\Delta t/s$	S.D/s	S.E/s
e_{11}: A/C visible	54	26.5	18.3	2.5
e_8: Gear visible	44	13.2	7.6	1.2
e_6: main RWY	28	15.7	16.0	3.1
e_7: grass RWY	20	25.8	24.5	5.6
e_5: touchdown	25	+2.0	1.0	0.2
e_4: takeoff	23	+2.0	1.4	0.3

The distance between the airport reference point ARP and departure and approach turning points was ca. 4 km and 14 km respectively. Each flyby was characterized by 6 parameters, with parameter values statistically mixed:

1. approaching main (concrete) or grass runway; 2. approach angle normal or high; 3. landing gear out: early, normal, late; 4. low level crossing of airport or touch and go; 5. touch down point early or late; 6. departure angle normal, low angle, steep angle.

While pilots had a detailed plan to follow for the sequence of approaches with different parameter values, the participants only knew about the different possibilities within the approaches. They had to activate the corresponding field of their input display of the tablet PC and set a time mark at the time of their observation of one out of the 11 possible events during each of the D-CODE approaches and flybys respectively. Also all approaches of additional (non-D-CODE) a/c were monitored. Experts and non-experts were briefed separately before the first experiment, with both groups filling separate questionaires. For each trial raw data from all subjects and for all approaches under real view conditions were collected into a single data file.

During test #1 significant time drifts between the individual computers were observed which were corrected for by comparing with the P_1-camera time as reference before and after the 2-h experiment for generating correction factors. For trials 2, 3 a W-LAN with time synchronized camera and data acquisition touch-input laptops was used.

On December 13 2006 the first out of three 2-h trials was performed (as a pre-trial for testing and improving the procedures) with lower cloud boundary at 600 m. Two more experiments were performed in 2007 on May 21 with clear sky and May 22 with reduced visibility (<10 km). The latter results are listed in Table 1.

4.3.2 Experimental Results and Discussion

Videopanorama

For each trial raw data from all subjects and for all approaches under real view conditions were collected into a single data file. Evaluation of the different approach, touch-and-go, and departure conditions (in trial #1 14 approaches with 11 D-CODE and 3 other aircraft) yielded the inter-subject event time-measurement scattering with mean and standard deviation (stdev) of the sample and standard errors (sterr) of mean for the $n = 5$ subjects.

In pre-trial #1 typical unbiased estimates of sample stdev for event e_{11} (first sighting during approach) were between 2 and 25 s (sterr = 1–15 s). Comparing approach detection time with low stdev with the GPS track yielded first sighting (e_{11}) of A/C (headlight) at distance 9 km. The minimum sterr. of e.g. 1 s for e_{11} and 0.2_s for e_5 (touchdown) presumably represented the optimum observation conditions for all subjects (i.e. all $n = 5$ attending first sighting direction during expected apearance time).

Quantitative data on the difference in event-detection times between real view and video panorama were obtained by repeating the experiments with the video panorama replay after a week or more in order for the subjects to no longer remember the different flight conditions. It was expected that due to lower resolution of the videopanorama as compared to the real view, distant events of approaching/departing a/c (like first/last sighting of a/c) should receive an earlier/later mark under real view as compared to video observation. Correspondingly within-subject evaluations of the direct viewing and video panorama replay observations yields time differences t(real view, e_i)–t(video, e_i) <0 and >0 for approaching (app) and departing (dpt) a/c respectively.

Results of the initial pre-trial #1 were reported in (Schmidt et al., 2007; Fürstenau et al., 2007), showing experimental visual resolution between 1.3 and 2 arcmin in reasonable agreement with the theoretical prediction and with the static verification measurements (Sect. 4.2). In Table 1 the results for six of the 11 possible observation types are listed for the trials #2, 3 (May 21, sunny day & 22/07, cloudy day), averaged over all participants and all flights with pairs of observation (time marks) of real view–video, with number. of observation pairs N, mean time difference Δt(real view–video), standard deviation S.D. and std. error of mean.

All displayed events exhibit reproducible and significant pos.(dpt.) and neg.(app.) delays between video panorama and real view conditions. For example the significant negative delay measured as overall mean for e_8 (landing gear visible, -13.0 ± 2.0 s and -13.2 ± 1.2 s respectively) shows this event to be observable with video only

0.7 km closer to the airport (a/c speed ca. 100 kn = 185 km/h), as compared to the real view conditions (e.g. e_{11}(real view): a/c (lights) recognized at ca. 8 $_{km}$). If we assume that detection time difference is determined by the difference of optical resolution between real view (resolution of the human eye ca. $\alpha_E \approx 1$ arcmin = 1/60°) and videopanorama system, the measured time difference Δt(real view-video) = t_E–t_V from Table 1 can be used for calculating the effective resolution α_V of the optical system. The extremely large observation time difference and s.d. of e_{11} in trial #2 was due to real view event registration under clear view conditions (mostly expert subjects S1, S2) long before the a/c turned towards approach at the ILS turning point. For the video observation only after passing this point and entering the glide path the A/C became visible.

For suitable events with known object size G the single Δt-values allow for calculation of α_V via:

$$\alpha_V = \alpha_E(1 + \alpha_E v_E \Delta t/G)^{-1} \qquad (2)$$

where the resolution angle α is given by $\alpha_{E,V} = G/x_{E,V}$ measured in rad, with event observation distance $x_{E,V}$ under real view (E) and video replay (V) conditions. G is the object size, e.g. aircraft cross section for e_{11} or landing gear wheel size for e_8. For e_8 we obtain in this way $\alpha_V = 1.4\ \alpha_E$ (with G(main wheel) = 0.65 m, $v_E = 100$ kn). For e_{11} (using G(cabin) = 1.8 m) trial #3 yields 1.3 α_E. Both values are in agreement within the experimental uncertainty, although smaller (i.e. even better) than the theoretical (optimistic) estimate.

In order to obtain a statistically relevant and model based mean value, a linear regression procedure is employed for those events where the visual resolution (more or less modified by image contrast) may be assumed to play the dominant role for event timing. Because e_1 was unreliable due to observability problems (the aircraft quite often vanished from the P1-camera observation angle before e_1 was observable), only e_4, e_5, e_8, e_{11} were used for this evaluation.

For applying a regression procedure the independent variable "event e_i" has to be replaced by a quantifiable variable. A linear model is obtained when considering the observation distance x as obtained from the GPS reference trajectory instead of the observation time, yielding the Δx(E–V) $_=$ $v_E(t)\ \Delta t$ versus x_E (= distance from event position x_E to airport reference point ARP). The plot of the four data points ([x_E, $\Delta x = v_E\Delta t$], for e_4, e_5, e_8, e_{11}) in Fig. 10 for trial #3 (cloudy day) is obtained by averaging real view–video observation delays for all flights and all subjects and correlating the measured time values with the corresponding time-synchronized GPS position data $x_E(t)$.

The corresponding linear model with video and real view resolution α_V and α_E respectively was derived as (see Appendix A)

$$\Delta x(eye - video) = (1 - \alpha_E/\alpha_V)x_E \qquad (3)$$

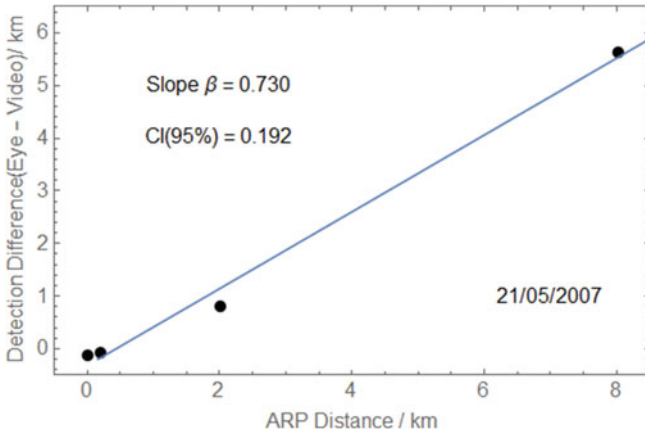$$\alpha_V = \alpha_E(1 - \beta_1)^{-1} \qquad (4)$$
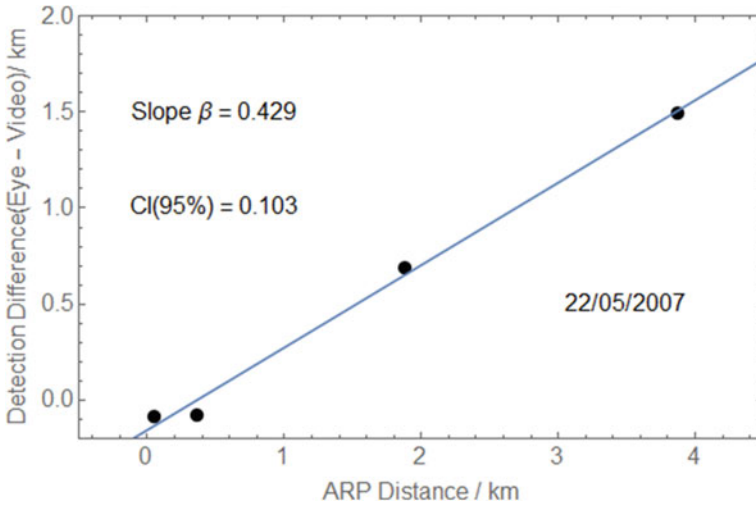
Fig. 10 Experiment #3 (reduced visibility): Mean event-observation position differences Δx (real view–video replay) for e4, e5, e8, e11, between real out-the windows view and video replay conditions versus mean GPS-position estimate of x_E for trial #3 including linear regression

With the slope $\beta_1 = \Delta x / x_E = 0.429$ (± 0.02 std.err.), 95% confidence interval of parameter estimate ci(95%) = 0.1, $R^2 = 0.99$, significance level F = 321 at p = 0.003) a corresponding α_V estimate of 1.75 arcmin (± 0.08) is obtained, again exhibiting surprisingly good agreement between the detection threshold of perceived events with the theoretical predictions and also the experimental visual resolution data from field observations of known static objects.

Although, in contrast to the above described results from the #3 test under reduced visibility, the #2-test results were obtained under good visibility conditions, they exhibit a decreased detectability as shown in the following Fig. 11. In this case the slope $\beta_1 = \Delta x / x_E = 0.730$ (± 0.04 std.err.) with 95% confidence interval of parameter estimate ci(95%) = 0.2, $R^2 = 0.99$, significance level F = 268 at p = 0.004), a correspondingly higher α_V estimate of 3.7 arcmin (± 0.6) is obtained, although with a reduced confidence.

The reason for this apparently low detectability and corresponding resolution value despite sunshine and clear weather was already discussed above with respect to the large real view–video detection time and scattering of event e_{11} (initial detection during approach). Under real view observers quite often detected the A/C already before turning into final approach on the glide path, resulting in a wrong (too short) distance. A too early detection (too shoort distance of e_{11}) in fact results in an overestimation of slope. Because this systematic error did not occur during all observations also the large scattering can be explained in this way.

As suggested already by the reduced confidence (large uncertainty range) of the #2-experiment results, we may conclude that the #3-experiment provides the more reliable resolution value, that moreover agrees with the time-based analysis, relying

Fig. 11 Experiment #2 (good visibility): Mean event-observation position differences Δx (real view–video replay) for e4, e5, e8, e11, between real out-the windows view and video replay conditions versus mean GPS-position estimate of x_E for trial #3 including linear regression

on aircraft speed instead of satellite position data, with the static measurements in Sect. 4.2, and with the theoretical estimate in Sect. 3.1.

Zoom Function

In order to decrease the duration of the replay experiments for evaluating observations with the PTZ camera (e_{11}, e_8) only the approach sections of the videos until touchdown (event e_5) were used. Because due to this procedure time synchronization with real-view experiments was lost, PTZ experiments were related to panorama replay with touchdown time as common reference. For data evaluation Eq. (4) with substitution of a_V through α_{PTZ} and α_E through a_V was used, yielding

$$\alpha_{PTZ} = \alpha_V(1 + \alpha_V v \Delta t/G)^{-1} \tag{5}$$

The experimental results for the effective zoom camera (PTZ-) resolution are presented in Table 2.

Table 2 PT-Zoom experiment for determining effective resolution. Δt = measured event observation time difference t(Panorama)–t(PTZ). Z = 3.6: day 1, clear; Z = 4: day 2: cloudy. 2Θ = field of view

Trial #2 & 3	α_{PTZ}/arcmin for (Δt/s)	
Zoom Factor Z (2Θ)	e_{11}: 1st Sighting	e_8: Gear down
3.6 (16.2°)	1.07 (52)	1.35 (10)
4.0 (14.5°)	1.30 (32)	1.23 (14)
Mean	1.2 (42)	1.3 (12)

The experimental data are reasonably close to the theoretical value $\alpha_{PTZ} \approx 1$'
$= \alpha_E$ as obtained under the hypothesis of resolution limited object detection times
(see Sect. 3.1). These data were obtained with 20 participants observing those three
rounds around the airport of each of the two days, which included a touchdown (e_5)
to be used as common PTZ–videopanorama reference with Δt (Panorama–PTZ) $\approx$
0 s.

5 Simulation Environment

Detailed descriptions of the (advanced) RTO simulation environment and human-in-
the-loop experiments using this system are presented in separate chapters of this book
(see Part III of this volume). Here only a brief overview of the initial RTO-simulator
is given.

For investigating possible RTO/RTC work system alternatives with different traffic
scenarios and determining RTO system specifications under reproducible experi-
mental conditions DLR's Apron and Tower simulator (ATS, depicted in Fig. 12) was
extended by a remote tower operator (RTO) console as shown in Fig. 5.

Besides the possibility of displaying the live stream of the panorama camera
system it was the main purpose of the ATS-RTO console to provide a simulated
real-time panorama as derived from an image generator (IG) with simulated airport
traffic generated by the ATS simulation engine. The simulation included PTZ camera,
displayed on the touch input display integrated into the console. Position data, flight
plan data, weather information, and airfield lighting were provided by the simula-
tion. This allows simulation of the advanced augmented vision capabilities of the
RTO-controller work position (CWP) via integration into the simulated far view for
trials within the validation setup with professional controllers as test subjects. For
simulation trials an eye tracking measurement system could be added to the system
with optical head tracker for obtaining quantitative fixation and dwell-time data of
the areas of interest attended by the operators.

Fig. 12 RTC simulator environment used for Remote Tower experiments (reprinted from Fürstenau et al., 2011, IFAC, with permission). Photo depicts the previous 200° vision system of the DLR tower simulator (ATS), extended by the 180°-RTO backprojection console (left). The latter could alternatively display RTO-simulations or the live Panorama

Fig. 13 RTC Simulator
setup for human-in-the-loop
simulation with
simultaneous surveillance of
two remote airports using
two augmented vision
videopanorama systems

There are several reasons for validating the RTO/RTC concept, besides field tests
also by means of integration into a real-time simulation environment. First of all it
ensures control and reproducibility over experimental conditions and constraints.
Variation of traffic mix and load, environmental conditions and the creation of
possibly conflicting situations allows the evaluation of human factors and safety
related issues. Furthermore, the ability to vary between different CWP configura-
tions and operational procedures in the simulator enables a more comprehensive
analysis of related organisational and operational constraints for the implementation
of the new remote tower center (RTC) concept. The simulation setup supports also
the analysis of HMI and RTC work system design. Special real-time simulation capa-
bilities were prepared for validating RTO working positions located within a RTC
for two small airports as depicted in Fig. 13.

The real-time simulator experiments were an integral part of the (iterated) concept
development- and validation process. Due to its characteristics, the experiments
carried out within the simulator focused on certain specific issues. The experimental
design covered the analysis of operational procedures, the dedicated work environ-
ment and the evaluation of its influences on controller workload and situational
awareness. Additionally the developed work share within the combined RTC envi-
ronment was observed and analysed as well as attention- and perception related
factors. Within the experiments variation of different visual conditions, including
reduced visibility conditions (fog) as well as a variation of available light situations
(day- and night-time conditions) were examined.

Another important issue was the specification of technical system parameters like
video frame rate (see in this chapter). These parameters were varied systematically in
order to investigate the limitations of the reconstructed far view, search for alternative
solutions, and derive specifications for operational use.

Results of RTO/RTC simulation experiments including structured interviews with
professional controllers are published in Möhlenbrink et al. (2010), Papenfuss et al.
(2010) and in Chapter 12 of this book.

6 Conclusion and Outlook

Basic elements of DLR's initial experimental Remote Tower Operation (RTO) system at the Braunschweig Research Airport are described and the theoretically expected performance of the RTO video panorama including zoom function as core of the new controller working position is estimated. The RTO human–machine interface was developed under the guideline of human centered automation and basic results of structured cognitive work analysis are briefly reviewed. Initial field test results are reported which were evaluated by assuming the visual (pixel) resolution to play the dominant role for event detection. Quantitative evaluation of field trials by comparing real view and video panorama detectability of different events confirmed the theoretically predicted video panorama resolution of ca. 2 arcmin. Resolution of the pan-tilt zoom camera (PTZ) was near the predicted value of ca. 1 arcmin with zoom factor Z = 4 and exceeded it with increasing Z. Advanced features like PTZ-object tracking based on real time image processing for movement detection, and augmented vision by superimposed flight data such as multilateration position transmitted by Mode-S transponder were demonstrated. Separately investigated extensions for improving low visibility conditions include a high resolution thermal imager that is a component of the improved prototype RTO-videopanorama system (see Chapters 7, 19) developed for the shadow mode validation experiments (see Chapter 9).

Besides the experimental system for field testing a first version of RTO simulator environment was realized for high-fidelity human-in-the-loop (HitL) RTO simulation experiments (see " part III of this volume"). It included an eye tracking measurement system for obtaining quantitative data on the areas of interest attended by the operators during simulation trials. High-fidelity HitL experiments complement field trials due to the improved possibility for experiments under reproducible and more laboratory-like controlled conditions.

References

Barfield, W., & Caudell, T. (Eds.) (2001). *Fundamentals of Wearable Computers and Augmented Reality*. Mahwah/NJ, London: Lawrence Erlbaum.

Bass, M. (1995). *Handbook of Optics* (Vol. 1, pp. 24.13 ff). New York: McGraw-Hill.

Brinton, & Atkins. (2006). Brinton & Atkins Remote Airport Traffic Services Concept, Proceedings, I-CNS Conference, Baltimore 5/2006. Baltimore

Ellis, S. R. (1991). *Pictorial Communication in Virtual and Real Environments*. Taylor & Francis.

Ellis, S. R., Fürstenau, N., & Mittendorf, M. (2011). Determination of frame rate requirements of video-panorama-based virtual towers using visual discrimination of landing aircraft deceleration during simulated aircraft landing. *Fortschritt-Berichte VDI, 22*(33), 519–524.

Fürstenau, N. (2013). *Remote Airport Traffic Control Center (RAiCe).* Braunschweig: DLR IB-112-2013/20.

Fürstenau, N. (1996). From Sensors to Situation Awareness. *DLR Mitteilung 96–02.* Köln: DLR.

Fürstenau, N. (2004). *Virtual Tower. Visionary Projects Competition 2001–2004* (pp. 16–19). Köln: DLR.

Fürstenau, N., Rudolph, M., Schmidt, M., & Lorenz, B. (2004). On the use of transparent rear projection screens to reduce head-down time in the air-traffic control tower. In *Proceedings of the Human Performance, Situation Awareness and Automation Technology (HAPSA II)* (pp. 195–200). Lawrence Erlbaum, Mahwah/NJ.

Fürstenau, N., Rudolph, M., Schmidt, M., Werther, B., Hetzheim, H., Halle, W., & Tuchscheerer, W. (2008). Flugverkehr-Leiteinrichtung (Virtueller Tower). *Europäisches Patent EP1791364.*

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., & Halle, W. (2008). Augmented vision videopano-rama system for remote airport tower operation. In *Proceedings of the ICAS 2008, 26th International Congress of the Aeronautical Sciences.* Anchorage: I. Grant.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Kaltenhäuser, S., & Schlüßler, E. (2011). Experimental system for remote airport traffic control center research. In F. Vanderhaegen (Ed.), *IFAC Proceeding of the Analysis, Design, and Evaluation of Human-Machine Systems. 11/1,* pp. 111–116. Valenciennes / France: Elsevier.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, M., & Werther, B. (2007). Development of an augmented vision videopanorama human-machine interface for remote airport tower operation. In *Proceedings of the 6th Eurocontrol Innovative Research Workshop* (pp. 125–132). Bretigny, France: Eurocontrol.

Hannon, D., Lee, J., Geyer, M., Mackey, S., Sheridan, T., Francis, M., … Malonson, M. (2008). Feasibility Evaluation of a Staffed Virtual Tower. *The Journal of Air Traffic Control,* 27–39.

Kraiss, K., & Kuhlen, T. (1996). Virtual reality—principles and applications. In N. Fürstenau (Ed.), *From Sensors to Situation Awareness. DLR-Mitteilung 112-96-02* (pp. 187–208).

Möhlenbrink, C., Friedrich, M., Papenfuss, A., Rudolph, M., Schmidt, M., Morlang, F., & Fürstenau, N. (2010). High-fidelity human-in-the-loop simulations as one step towards remote control of regional airports: a preliminary study. In *Proceedings of the Internatioanl Conference Research Air Transportation.* Budapest: Eurocontrol.

Papenfuss, A., & Möhlenbrink, C. (2009). *RaiCe—Cognitive Work Analysis.* Braunschweig: DLR IB-112–2009/22.

Papenfuss, A., Möhlenbrink, C., Friedrich, M., Schmidt, M., Morlang, F., & Fürstenau, N. (2010). High-fidelity human-in-the-loop simulations as one step towards remote control of regional airports: a preliminary study. In *Proceedings of the of 11th IFAC Symp. Analysis, Design, and Evaluation of Human-Machine Systems.* Valenciennes.

Peterson, S., & Pinska, E. (2006). Human performance with simulated collimation in transparent projection screens. In *Proceedings of the 2nd International Conference Research in Air Transportation* (pp. 231–237). Belgrade: Eurocontrol.

Pinska, E. (2006). An investigation of the head-up time at tower and ground control positions. In *Proceedings of the 5th Eurocontrol Innovative Research Workshop* (pp. 81–86). Bretigny: Eurocontrol.

Schmidt, M., Rudolph, M., Papenfuss, A., Friedrich, M., Möhlenbrink, C., Kaltenhäuser, S., & Fürstenau, N. (2009). Remote airport traffic control center with augmented vision videopanorama. In *Proceedings of the 28th IEEE-DASC* (pp. 4.E.2–1–4.E.2–15). Orlando: IEEE.

Schmidt, M., Rudolph, M., Werther, B., & Fürstenau, N. (2006). Remote airport tower operation with augmented vision videopanorama HMI. In *Proceedings of the 2nd International Conference on Research in Air Transportation (ICRAT II)* (pp. 221–230). Belgrade: Eurocontrol.

Schmidt, M., Rudolph, M., Werther, B., Möhlenbrink, C., & Fürstenau, N. (2007). Development of an augmented vision video panorama human-machine interface for remote airport tower operation. In M. J. Smith, & G. Salvendy (Eds.). *Lecture Notes Computer Science: Human Interface II* (Vol. 4558, pp. 1119–1128). Berlin, Heidelberg: Springer Verlag.

Tavanti, M. (2006). Control Tower Operations: A Literature Review of Task Analysis Studies. *Eurocontrol Experimental Center EEC Note 05.*

Tavanti, M. (2007). Augmented reality for tower: using scenarios for describing tower activities. In *Proceedings of the 28th Digital Avionics Systems Conference (DASC'07)* (pp. 5.A.4–1–5.A.4–12). Dallas/Texas: IEEE.

Vicente, K. (1999). *Cognitive Work Analysis.* Lawrence Erlbaum Associates.

Vogel, B. (2009). ANSP's explore control from afar. *Janes Airport Review*, pp. 16–17.

Watson, A., Ramirez, C., & Salud, E. (2009). Predicting visibility of aircraft. *PLoS ONE. 4(5)*, e5594. https://doi.org/10.1371/journal.pone.0005594.

Werther, B. (2006). Colored Petri net based modeling of airport control processes. In *Proceedings of the Internatioanl Conference on Computer Intelligence for Modeling, Control & Automation (CIMCA).* Sydney: IEEE. ISBN 0-7695-2731-0.

Werther, B. (2006). *Kognitive Modellierung mit farbigen Petrinetzen zur Analyse menschlichen Verhaltens.* Ph.D. Dissertation, German Aerospace Center (DLR).

Werther, B., & Schnieder, E. (2005). Formal Cognitive Ressource Model: Modeling of Human Behavior in Complex Work Environments. In *Proceedings of the International Conference Computational Intelligence for Modelling, Control & Automation (CIMCA)*, (pp. 606–611). Wien.

Werther, B., & Uhlmann, H. (2005). Ansatz zur modellbasierten Entwicklung eines Lotsenarbeitsplatzes. *Fortschrittberichte VDI, Zustandserkennung und Systemgestaltung* (Vol. 22, pp. 291–294).

Werther, B., Möhlenbrink, C., & Rudolph, M. (2007). Formal Petri-Net based Airport Control Model for simulation and analysis of airport control processes. In *Proceedings of the Conference Human Computer Interact. (HCII 2007). Publ. in: Lecture Notes in Computer Science* (vol. 4558). Beijing: Springer.

Remote Tower Prototype System and Automation Perspectives

Markus Schmidt, Michael Rudolph, and Norbert Fürstenau

Abstract In this chapter we describe the development of the video panorama based Remote Tower prototype system as the main goal of the second DLR-RTO project (RAiCe, Remote Airport traffic Control Center). One focus was on the implementation of an advanced RTO-environment at a second airport (besides a comparable system at the Research airport Braunschweig). It was used for the worldwide first RTO-validation experiments with controlled flight scenarios for directly comparing RTO versus tower conditions using a DLR test aircraft (see separate chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" and "Model Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position"). The advanced RTO system served for analysing the performance of the near-prototype level of hard and software solutions and for preparing and executing passive shadow mode field test with participation of domain experts for providing more realistic operational conditions. We will describe the design and setup of this RTO-system which was realized in cooperation with the German air-navigation service provider DFS. A detailed work analysis with DFS domain experts during workshops and RTO simulations provided a breakdown of the specific requirement specifications. The analysis showed that it would be impossible to consider all of these requirements in an RTO design within a reasonable cost frame. This concerned the selection of type, numbers and focal width of cameras, their visual resolution, contrast, dynamic range and field of view, zoom functions and the corresponding number and type of displays or projection systems for the reconstructed panoramic view. E.g., the vertical FOV turned out as a crucial factor, the visual surveillance up to an altitude of 1000 ft. above the runway in the panoramic view as one of the basic design condition. In the present chapter we will describe hard- and software aspects of the system design, its setup, initial tests and verification, as precondition for the RTO-validation experiments. Furthermore we include an outlook on the automation potential using image processing. The requirement for automation of functions such as pan-tilt zoom camera based object tracking

M. Schmidt · M. Rudolph · N. Fürstenau (✉)
German Aerospace Center (DLR), Institute of Flight Guidance, Lilienthalplatz 7, 38108 Braunschweig, Germany
e-mail: norbert.fuerstenau@dlr.de

via movement detection was derived from the results of validation experiments
described in chapters "Model Based Analysis of Two-Alternative Decision Errors in
a Videopanorama-Based Remote Tower Work Position", "Multiple Remote Tower
Simulation Environment", and "Assessing Operational Validity of Remote Tower
Control in High-Fidelity Simulation".

1 Introduction

One of the basic goals of the DLR-project RAiCe (Remote Airport Traffic Control
Center, 2008–2012) as follow up of the initial experimental system develop-
ment within the project RApTOr (Remote Airport Traffic Operation Research,
2005–2007, chapter "Remote Tower Experimental System with Augmented Vision
Videopanorama") was the setup and test of an improved RTO-system at a second
airport (Fürstenau, 2013). It served for demonstrating and analyzing the next (near
prototype) level of hard and software solutions, for designing and testing long
distance real-time video data transmission, and for preparing and executing RTO-
passive shadow mode test at this airport. These goals were achieved within a coopera-
tion with the German air navigation service provider (ANSP) DFS (project RAiCon,
Remote Airport Cooperation, 2011–2012).

In what follows we will briefly review in Sect. 2 the results of in-depth work
and task analysis addressing the specific design requirements for Remote Tower
Operation at low traffic airports selected for the prototype verification and validation
experiments. In Sect. 3 the concrete design and setup of the advanced prototype
systems at airports Braunschweig and Erfurt is described, including RTO-specific
software development, the controller working position (CWP), and the long distance
wide area network (WAN) connection between the remote airport and the DLR tower
lab in Braunschweig. Section 4 addresses RTO-specific aspects and perspectives
of advanced image processing and thermal imaging. In Sect. 5 verification with
functional tests for quantification of relevant RTO-system features is addressed. We
finish this chapter with a conclusion and outlook in Sect. 6.

2 Work and Task Analysis for Requirements Specifications and Prototype Design

The major characteristics of the type of airport under consideration are given by a
few points only:

1. Low ratio of IFR versus VFR traffic. Only the former requires controlled airspace. On low-traffic airports quite often only a handful of regularly operating flights (IFR, usually commercial international flights) has to be controlled.
2. VFR traffic (usually general aviation) is highly weather dependent and often generates unexpected events due to heterogeneous pilot skills and experience.
3. Low-traffic and regional airports usually have only a low level electronic sensor infrastructure (e.g. no surface movement radar SME).
4. The tower provides 1–2 working positions, with often a single operator during low traffic hours.

The specific low-traffic airport requirement specifications for the prototype RTO system-design built upon the initial results obtained within the initial DLR remote tower projects (ViTo, RApTOR, 2002–2007, see previous chapters). Increased cooperation with the German ANSP DFS and several RTO-simulation campaigns with voluntarily participating (and paid) controllers increased significantly the number of experts (tower controllers and supervisors) contributing their domain knowledge, and provided a more detailed requirement breakdown through discussions, formalized interviews and design workshops. The comparison of the updated requirement specifications with achievable technical parameters showed that in particular for the reconstruction of the visual information (the "far view") it was hardly possible to consider all of these requirements for a realizable design under reasonable cost constraints. This concerned the numbers of cameras and the corresponding number of displays for the panoramic view, the visual resolution, dynamic view, contrast and the field of view in vertical and horizontal direction.

Of course there exist ICAO and DFS-requirements for tower construction concerning direct visual observability conditions which may be used as guidelines for the reconstructed far view. These are more or less specific, e.g.:

- the tower should have an appropriate distance from the landing threshold
- the vertical FOV should be at least 1000 ft. above the RWY, i.e. at least ca. 40° rel. to the horizon at a distance of 400 m between tower and runway (ICAO, 2013) (DFS: BA-FVD343)
- good 360° panoramic view (implicitly assuming visual resolution of the human eye), including airport circles, approach sectors, runway, movement areas
- at least one CWP with places for 2 persons for training purpose, including consoles equipped with different VHF/UHF stations, control strips, telephone communication

The following graph (Fig. 1) depicts the percentage of use of different information sources for all identified controllers tasks, separated for the ground (PG) and tower (PL) controller obtained from controller interviews (Papenfuss & Möhlenbrink, 2009).

The graphic is based on structured interviews with two supervisors and relates to medium sized airports. It provides an indication that PL and PG make quite different use of the different available information sources. Out-of-windows view appears to be used mainly by the PG whereas exclusively the PL uses the (approach) radar.

Use of Information Sources for Ground (PG) and Tower (PL) Controller
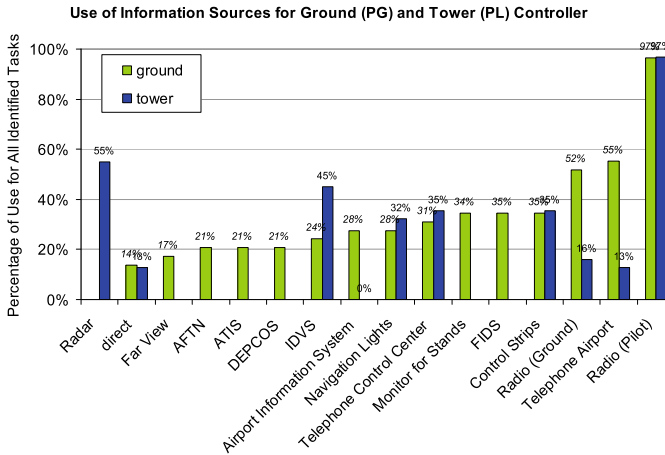
Fig. 1 Use of information sources of tower and ground controller for decision and support tasks. Average of 29 PG-tasks and 31 PL-tasks (clearances, communications etc.)

Radio communication with pilots, control strips (flight plan information), telephone communication with approach control and airport, and weather information display IDVS are other important communication/information channels. Clearly the distribution of work will differ at small and low traffic airports from medium and large ones (the latter ones generally have separate apron controllers which usually are airport operator employees) so that the PL–PG work distribution may be different at different airports and of course vanishes for the single operator situation (see also chapters "Remote Tower Experimental System with Augmented Vision Videopanorama" and "Assessing Operational Validity of Remote Tower Control in High-Fidelity Simulation").

During prototyping of the advanced RTO-system the DLR team together with DFS domain experts discussed the concept and the design of the improved RTO controller working position (RTO-CWP). Within two specific design workshops the participants derived the framework for an operational design based on the requirement specifications and other conditions. From a complete requirements list (separated for day and night operation) we will extract here only some aspects of interest for the far view reconstruction, the core HMI-component of the RTO-CWP.

- 360°-panoramic view is considered necessary for observability of all movements of aircraft in the airport vicinity (including airport circling within the control zone) and of cars and persons on the movement areas. The direction opposite of the RWY shall be visualized at least on demand within 3 s.
- Recognition of traffic situations/movements (before reaching of clearance boundaries) and emergencies without delay for timely control actions
- Discriminability of weather conditions (wind, cloud parameters) and recognition of weather changes

- Recognition of aircraft and vehicle positions, operating states and movement state (direction, acceleration, braking)
- Recognition of fixed and mobile obstacles
- Object detectability: size 0.3 m/1 km distance (=1 arcmin visual resolution)
- Zoom function (=binocular function): continuous, Z_{max} within 2 s, pre-set and hot-spot viewing positions (pan-tilt zoom, PTZ)
- manual and automatic tracking of 0.3 m objects in 1 km distance
- all requirements are valid for day and night operation

It seems worth while to remind the experimental results of the field tests reported in the previous chapter "Remote Tower Experimental System with Augmented Vision Videopanorama": the initial experimental video panorama system at Braunschweig research airport showed an optimistic visual resolution of about 2 arcmin (0.6 m/km, about half as good as the human eye, not considering contrast effects; see also Appendix chapter A) which was obtained with the best high-resolution CCD cameras available at that time (ca. 2006: (Fürstenau et al., 2007, 2008b; Schmidt et al., 2007)). Meanwhile technology changed to HD-format as standard, even quad-HD available, and cost for high-resolution cameras decreased from >10 k€ to <5 k€.

3 The RTO-System Setup and Human-System Interface

3.1 Video-Panorama Camera System

The crucial aspects for the selection of the most important system design requirements with regard to the reconstruction of the tower out-of-windows view (the "far view") derived from the task analysis, are the visual resolution and the vertical FOV, according to the requirement for object detectability in the video panorama up to an altitude of 1000 ft. (300 m) above the runway (ICAO, 2013).

Table 1 shows the theoretically available options used for comparing the different configurations, as dependent on number of cameras and focal width. The pixel resolution of the cameras for all options is determined by HD-format, i.e. 1920 × 1080 pixels.

Because the FOV by most experts was considered more important than the maximum achievable resolution (within the given cost frame for the system components and complexity) the RTO-system setup was defined following option 1, with the following characteristics:

- 5 industrial HD-cameras, 2/3″-CCD-technology, with f = 8 mm lenses
- single camera housings with heating and air blaster cleaned front window
- PanTiltZoom-camera (PTZ) with VGA-resolution, continuously horizontally rotatable with tilt angle from −30° to 90° rel. to the horizon.

Figure 2 shows the panorama camera system with pan-tilt-zoom (PTZ) camera on top after adjustment (viewing direction north, 360° horizontal FOV, with electronic

Table 1 Different design options for RTO-video panorama camera system (1 arcmin = 1/60° = 0.3 mrad)

	Option 1	Option 2	Option 3
Numbers of HD-cameras/displays	5	7	10
Vertical/horizontal FOV	68°/190°	46°/182°	34°/190°
Resolution/arc minute per pixel	Ca. 2	Ca. 1,44	Ca. 1
Main focus	High FOV	Medium FOV and resolution	High resolution
Conclusions	Medium effort Medium resolution	Affordable but realistic	High effort Huge required space for the R-CWP

Fig. 2 Panorama camera and PTZ setup at the Airport Braunschweig/Wolfsburg

box beyond. In addition to the power supply and opto-electronic components for data transfer it contains an air-blast system for remotely controlled cleaning of the camera housing windows.

The cameras originally included infrared filters which served for optimising the RGB color definition. These were removed in order to increase the overall sensitivity at the cost of color reproduction fidelity. The effective dynamical range of the cameras (including video image (color) pre-processing was 8 bit (=255 intensity steps) for each of the three RGB color channels. The video framerate typically was set to the maximum possible value of 30 Hz (see chapter "Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration During Simulated Aircraft Landing").

3.2 RTO-Controller Working Position

Based on the experience gained within the first DLR RTO project (see chapter "Remote Tower Experimental System with Augmented Vision Videopanorama") and a close cooperation with DFS domain experts, including two specific design workshops (see Sect. 2) an improved design of the new RTO-controller working position (RTO-CWP) was realized. A corresponding version although somewhat extended for experimental purposes with 360° panorama was set up in the DLR Tower-Lab at the Inst. of Flight Guidance. It replaced the initial version shown in chapter "Remote Tower Experimental System with Augmented Vision Videopanorama" and served for testing the long distance WAN-connection between Braunschweig and Erfurt.

Fig. 3 Prototype RTO-Video panorama with basic controller console in the DLR Tower-Lab (2013) reproducing the out-of-windows view at Braunscheig airport. 360° horizontal FOV with 6 vertical displays for the 228° view towards the runway in northern direction. Two horizontal displays covering the southern 136° viewing sector with the same resolution

Figure 3 depicts the extended RTO-CWP showing the 360° panorama of Braunschweig airport: six portrait orientation displays with 68° vertical and 228° horizontal FOV are complemented by two horizontal displays attached to the left and right side covering the southern direction. The PTZ display with camera interaction controls is integrated in the tilted touch-input display at the operator console which in addition contains a second touch input display with electronic flight strips.

This experimental RTO-CWP at the DLR-Towerlab (see chapter "Multiple Remote Tower Simulation Environment") was used as development environment. It demonstrated the excellent data transmission performance of the WAN-connection between Braunschweig and Erfurt, i.e. the flawless transmission of the video streams over several hundred kilometers (see Sect. 3.3). The experience gained with the experimental system combined with the results from the design workshops yielded an operational RTO-CWP prototype design derived from the latest DFS controller console with three levels of information or surveillance and control input respectively as depicted in Fig. 4.

The realisation based on this design and the installation of the RTO-CWP in a control room adjacent to the Erfurt airport tower building was realized by DFS engineers. This final version as depicted in Fig. 3 was used for the validation experiments described in the following chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" and "Model Based Analysis of Two-Alternative Decision

Fig. 4 Implementation of the R-CWP at Erfurt tower control room without direct airport view. Three levels of information with the video panorama in the back, the row of four flight information displays, operator console with two large touch-input displays and a small one at the right side for radio communication, are clearly structured (for details see text)

Errors in a Videopanorama-Based Remote Tower Work Position". In addition to the video panorama and the touch-input control-and-PTZ-display integrated in the operator console (for details see Sect. 3.4) the work-position also includes the standard displays for the essential operational ATM data (from left to right): IDVS (weather data; see also chapter "Assessing Operational Validity… "), (approach) radar, flight data (electronic flight strips). The latter could be moved to the large touch-input display on the right side of the operator's console.

The described camera system together with the RTO-CWP in Erfurt was the technical basis for the validation tests described in following chapters. This initial quasi-operational validation by means of so called passive shadow mode tests (no communication between RTO-controllers and pilots) serves as the basis of larger scale validation exercises within the a European ATM-research context (SESAR Work Package 6.8.4).

3.3   High-Bandwidth Wide-Area Network

One goal of the second DLR Remote Tower project was the concept development, implementation and the verification of a high bandwidth long distance connection between the RTO-CWP in the DLR Tower-Lab (see chapter "Multiple Remote Tower Simulation Environment") and a remote airport for testing the performance of the data transfer of the video streams from the advanced prototype sensor system. This long-distance connection over several hundred km was realized with the advanced video panorama system described above, between the camera system on the Erfurt tower and the experimental RTO-CWP at the DLR-tower lab. The live video panorama could be displayed simultaneously at the latter CWP and at the operational RTO-CWP at the ground level of the Erfurt tower building for the validation experiments without direct far view. For testing the live transmission and verifying its performance (see Sect. 5) a specific secure high-bandwidth WAN-connection was set up. A peer-to-peer fiber-optic connection between the RTO-CWP and Erfurt tower was realized by an external provider with a minimum bandwidth of 50 Mbit/s (optionally to be increased up to 100 Mbit/s) and the router/switching devices at both endpoints. The network was decoupled from the DLR-domain as well as from the internal DFS-network so that data security at both sides of the connection was established.

3.4   RTO Software and Human–Machine Interaction

The advanced RTO software for the prototype RTO-system to be used for the shadow mode validation experiment (see chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" and "Model Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position") was
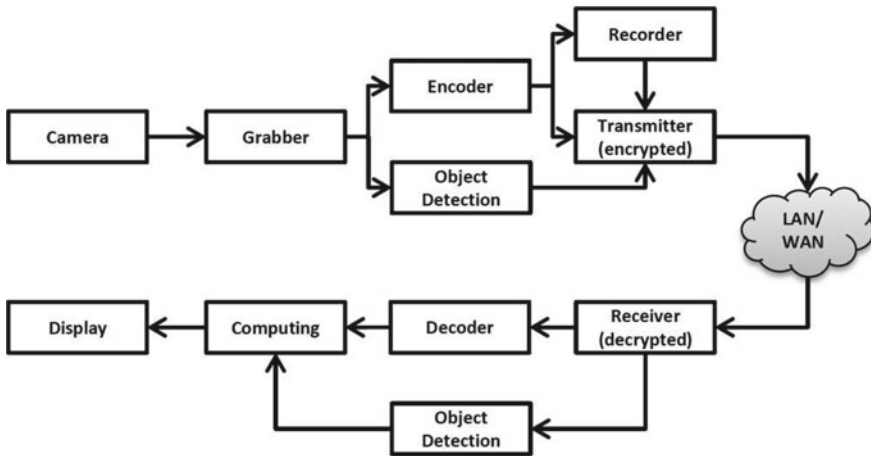
Fig. 5 Data path of the video stream

based on the initial experimental RTO system at Braunschweig research airport (DLR project RapTOr, see chapter "Remote Tower Experimental System with Augmented Vision Videopanorama"). Besides improved image processing and quality (data compression (CoDec)), and Bayer color format interpolation, see below) it provided more flexibility for integration into other hard- and software environments.

The general data path of a video stream in the remote tower software is shown in Fig. 5.

The stream of the captured raw images is transferred to the grabber who converts it to RGB images. According to the basic concept (Fürstenau et al., 2008a, b) in the following step the output of the grabber is split into two paths. The first one feeds the uncompressed data into the movement and object detection for achieving minimum delay and maximum image processing quality. The second stream passes an image compression stage, is recorded for playback and transmitted to the remote location. For the initial experimental system (see previous chapter "Remote Tower Experimental System with Augmented Vision Videopanorama") an MJPEG encoder was used because of limited computational power and the unlimited bandwidth in the local (GBit-Ethernet) network. For the functional verification of the prototype system with long-distance transmission, in contrast a H.264 encoder was employed to transfer the high definition video stream to a remote location with a limited network bandwidth of 50 Mbit/s. To ensure data security the stream was AES-256 encrypted by the transmitter. After decryption the receiver forwards the stream to the video decoder which in turn forwards the decoded video to the display.

The initial experimental system provided the possibility to use any kind of grabber that delivers Microsoft DirectShow drivers and industrial CameraLink grabbers. For the prototype this support was extended to industrial GigE Vision grabbers and GigE Vision cameras connected with a standard network interface card. This ensured to integration of a wide range of GigE Vision cameras. Furthermore a software-grabber

allowed for integration of simulated cameras of the DLR tower simulator. In this way the same RTO-software could be used for the real world live stream and for the synthetic DLR simulation environment. The possibility to use network interface cards instead of special grabber cards resulted in a significant cost reduction.

For the majority of high quality industry cameras the captured images are provided in the Bayer color format. The grabber has to interpolate a RGB color image from the raw picture with an adequate algorithm to provide the best quality possible with preferably low delay. The demosaicing process usually is done onboard in hardware if industrial products are used. In the case of standard network interface cards, however the raw Bayer images are provided and the images have to be converted with custom made algorithms. For the RTO prototype an adaptive homogeneity-directed demosaicing algorithm was implemented because it delivers high quality images in a reasonable time. The interpolation was done in OpenCL on a high end graphics card resulting in a very low computation time. The maximum video framerate (see chapter "Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration During Simulated Aircraft Landing") was limited to ca. 33 Hz by the GigE Vision IP based camera interface standard.

For manual control of the pan tilt zoom camera (PTZ) a specific display was developed that offers several possibilities to navigate the camera, based on pen-touch input functionality. Figure 6 depicts a photo of this advanced HMI version which represented an advanced version derived from the initial experimental one described in chapter "Remote Tower Experimental System with Augmented Vision Videopanorama".

On the top right side a number of preset buttons and buttons for static commands like move, zoom or (window) clean is located. Below a kind of wind rose can be seen. The inner circle serves as "virtual joystick" where a seamless movement of the camera in a specified (tilt) direction is possible with specified speed. The outer ring serves for commanding the desired horizontal (pan) position. The actual position and field of view of the camera is highlighted there with yellow color. On the left side of the ring a corresponding vertical scale is integrated for setting the tilt position. Outside of the ring are the fields to control predefined zoom factors, Z = 2, 4, 8, 16. At the bottom left a reduced version of the video panorama can be seen. A click inside this sector moves the camera viewing direction to the corresponding pan-tilt angles. The position of the camera is shown in the video panorama by a yellow frame. Usability trials with operators showed that this feature supports the orientation when users manually control the camera.

Structured interviews of controllers during design workshops and RTO-simulator experiments (see Sect. 2 of the present chapter and chapter "Assessing Operational Validity of Remote Tower Control in High-Fidelity Simulation") as well as during the shadow mode validation experiments (chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" and "The Advanced Remote Tower System and Its Validation") showed that automated tracking of the pan tilt zoom camera would be very helpful. The remote tower software offers the automatic tracking of aircraft by multilateration data via Mode-S transponder (see chapter "Remote Tower

Fig. 6 Pan tilt zoom camera-control pen-touch input display with viewing direction indicator (virtual joystick, bottom right) including Zoom factor selection (Z = 2, 4, 8, 16), and mini-panorama. For details see text

Experimental System with Augmented Vision Videopanorama" and Chap. 8) and by using vehicle positions from automatic movement detection (see Sect. 4). This functionality was already demonstrated in the first RTO project RApTOR with a more basic approach (see previous chapter). It was not intended to activate the automatic movement detection and tracking functions within the validation experiments due to limited reliability that was not sufficient for operational testing. The results of the validation experiment however show, that automation features of this kind are probably required in order to rise the RTO-system performance and usability to the operational level (see chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" and "Model Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position").

4 Perspectives of Automatic Movement and Object Detection

The basic automatic movement detection and PTZ-tracking functions were demonstrated already wit the initial experimental RTO system of the DLR-project RApTOR, together with augmented vision features. This concerned in particular dynamic Mode-S transponder information overlay (A/C identification and altitude, see previous chapter "Remote Tower Experimental System with Augmented Vision Videopanorama"). The importance of a certain degree of automation by using data fusion of e.g. Mode-S transponder information and/or movement detection with PTZ-object tracking was derived from the performance deficits of the basic prototype systems as quantified in the initial validation experiments (see chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation", "Model Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position", and "The Advanced Remote Tower System and Its Validation").

Here we present some additional research, mainly done by experts at the DLR-unit "Optical Information Systems"/Berlin-Adlershof (a cooperation partner within the DLR-RTO projects). They were presented at the RaiCe Final Workshop (Remote Airport Traffic Control Center (RAiCe), 2013). Although these developments were not implemented for the validation experiments due to limited operational stability they indicate promising directions for future RTO automation which will be particularly useful for the reduction of operator's workload and increase of usability under multiple airport control within a Remote Tower Center (RTC, see chapter "Multiple Remote Tower Simulation Environment", and part IV of this volume). For example a selective marker in the images of the real time view (as augmented vision element (see chapters "Introduction: Basics, History, and Overview", "Remote Tower Experimental System with Augmented Vision Videopanorama", "The Advanced Remote Tower System and Its Validation", and Chap. 15) could help the (remote) tower operator to focus the attention to moving objects and potentially critical situations.

For this purpose an experimental version of automatic scene analysis by means of image processing in combination with an object tracker was investigated. In the system two different image processing approaches were realized in order to investigate the specific advantages and drawbacks. The algorithms were running in parallel and on the one hand employed optical flow analysis and on the other hand region tracking with background estimation. The combination of both was visualized as superimposed information within the video panorama (augmented vision). These approaches are suitable for moving object detection and are briefly described in Sects. 4.1, 4.2 and 4.3.

A specific problem of movement and object detection in the RTO-environment in contrast to exclusive ground traffic that has to be solved by all of these approaches, are moving objects like birds and clouds, without relevance for the RTO-tasks. It

turned out that this can be solved only by using different algorithms for the image sections below and above the horizon.

## 4.1 Movement Detection via Optical Flow Analysis

Optical flow analysis is comparable to the human peripheral vision. It detects objects due to their motion in a series of subsequent images. Following (Shi & Tomasi, 1994) in a first step the corners in the image are selected as features. In a second step these features are tracked with the KLT-Tracker (Kanade et al., 1991) through a block of eight subsequent images. Because both feature detector and tracker have to be very sensitive in order to support the detection of small (and distant) objects as well, the majority of the tracked features will be wrong ones. Two techniques are used in order to disclose and erase wrongly tracked features.

First, the features are redundantly tracked back from the last image of a block back to the first image (Wohlfeil & Börner, 2010). Erroneously tracked features can be determined by comparing the initial and final position of a tracked feature. Figure 7 depicts an example of optical flow analysis with a scene of the east-viewing camera on the tower of Erfurt airport.

Features moving less than a pixel during the block of eight images are regarded as features of static objects and ignored (red dots in the figure). All remaining features are displayed in the figure as red crosses with a red line showing their current motion. In a second step additional irrelevant features are erased by assuming that relevant objects move almost linearly within the short periods of eight frames.

Another problem is the motion of the clouds and their shadows. They move linearly, and without a human understanding of the scene they have all attributes of real moving objects in the sky or on the ground, respectively. Anyway, by means of the tracked features in the sky, the mean cloud motion in object space can be determined by assuming that all clouds are in a height of 1000 m and move in almost the same direction during several minutes. By knowing the mean cloud motion in object space, the mean cloud motion in image space is calculated (blue lines in the upper part of Fig. 7). Features in the sky, which move in the same way as the clouds, are regarded as features of clouds (dark red in the figure). Features which move in a different direction or with different speed are features of objects.

Finally, the remaining features are clustered to objects (yellow circles in the figure) and tracked through multiple blocks of images, assuming almost linear motion of the objects also during longer periods of time.
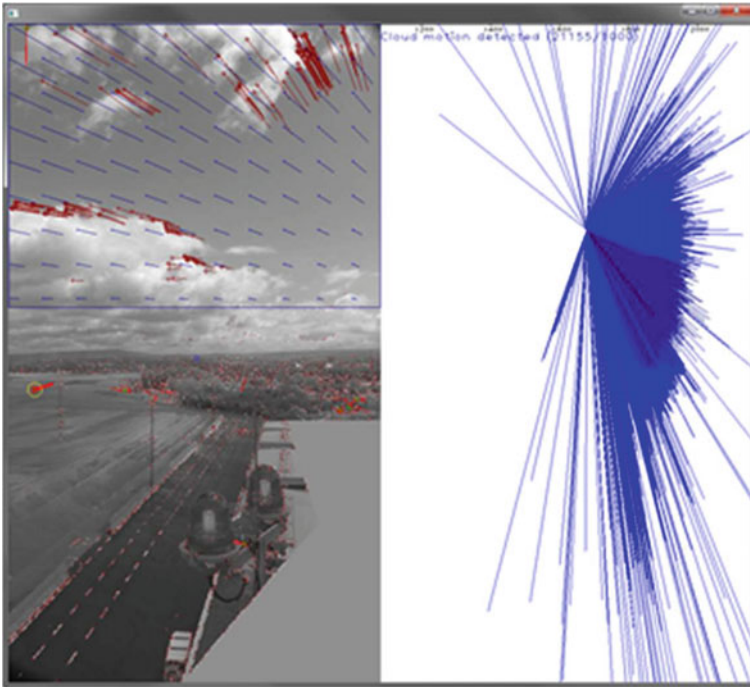
Fig. 7 Example of Object tracking with the optical flow method (scene from Erfurt airport tower). Filtering out of static objects (red dots) and clouds (dark red and blue lines indicating average movement direction), with remaining real (traffic) object recognition (yellow circle with bright red line). For details see text

4.2 Region Tracking Algorithm Based on Background Estimation

Static background subtraction was successfully used for detection of moving cars and other objects in general ground traffic. With the present application however, many false positive candidates especially for moving clouds in the sky were generated. On the other hand the implemented feature tracking algorithm based on optical flow as described above, also exhibits deficits. E.g., it does not provide a size or shape estimation for the object candidates. That is why experiments were performed with both algorithms used in parallel. An example result is depicted in Fig. 8.

In the first step sufficiently reliable candidate objects are detected with both approaches which are then fused on the object level. Only candidates detected and tracked with both approaches in a series of subsequent images are sent to the server as detected objects.

Fig. 8 Information Fusion.
Combination of optical flow
and background estimation
algorithms: detected objects
in the image colored green
and blue

4.3 Object Classification

The previous two algorithms were specifically optimized for combined detection of
moving aircraft. Especially the optical flow algorithm will lose sight of the detected
moving object immediately upon stop of movement. Also the object detection with
the background estimation is not able to detect the tracked object for an extended
time span. After a small period (depending on a "forgetting-factor") the object will
become background itself.

That is why additional algorithmic approaches for image interpretation of static
objects are needed, especially for aircraft not moving. For this task classification-
algorithms were developed. Before realizing the classification itself, it is important to
find linearly independent and robust features which can separate the class "aircraft"
from other object classes. For this purpose different assumptions were considered.
E.g., most aircraft exhibit bright or white shapes on the surface that in general exhibits
homogenously grey values. Moreover A/C are mostly brighter than the background

(soil, grass, apron or runway. However, above the horizon this feature can be inverted: A/C mostly appear darker than the bright sky as background. For this reason each image is split between above and below the horizon. An automatically calculated binary mask uses different algorithms for the feature generation. In a first step the total raw image was transformed from the RGB image into the YUV color space. Thereafter a new color feature was created: Feature_4 = (abs(yuv_image(2,*,*)* Maske(*,*) − yuv_image(1,*,*)* Maske(*,*))) (see also Fig. 9 below).

For increasing the rubustness of the analysis with regard to variations of the illumination in the images (e.g. induced by variable shading through clouds) big clusters were searched which exhibit corresponding similar grey values (these clusters are mostly the aprons surfaces or the surrounding vegetation (large grass areas). Here the mean illumination can be directly detected and will normalize the classification feature. An example of automatic color feature extraction and overlay is shown in the following Fig. 9.

Fig. 9 Bright color feature characterizing aircraft (upper b/w-image, Braunschweig panorama camera 3). Lower image: colored overlay results depict detected objects including one false positive on the left side, indicating the necessity of additional features for discrimination, e.g. area restrictions

4.4 Thermal Imaging

A detailed investigation in the potential of thermal imaging is reported in chapter "Designing and Evaluating a Fusion of Visible and Infrared Spectrum Video Streams for Remote Tower Operations". Here we provide a brief introduction based on a thermal camera integrated in the DLR prototype system. The standard CCD-camera chip exhibits a significant sensitivity in the near infrared spectrum (around 1 μm wavelength). In the usual configuration this part of the spectrum is filtered out by means of an infrared filter because it induces some imbalance in the RGB-color fidelity. For low visibility conditions and during nighttime however, this sensitivity may be utilized by removing the filter if this feature is included in the camera design.

A significantly larger step towards low-visibility object detection improvement can be achieved by including a high-resolution thermal imager. The relevant IR-spectral ranges are the atmospheric water windows around 5 and 12 m. For investigating and demonstrating the thermal imaging potential a 640×512 pixel cooled thermal imager for the 5–7 m midwave (MWIR) range with three discrete zoom steps (FOV: 1.2°, 4.7°, 15.3°) was included in the advanced video panorama setup at Braunschweig airport. It was integrated in a common housing together with the visual PTZ (see Fig. 10) Although for the experimental purpose this relatively high cost military thermal imager (FLIR) was integrated into the system, also a lower cost system is expected to significantly improve the reconstructed far view and to increase the situational awareness of controllers on low traffic airports without any other electronic (SME, multilateration, see Chap. 8) surveillance systems.

The following figure depicts an example taken with the FLIR PTZ-thermal imager. In this example it is applied for imaging the thermal aircraft signature during landing. Reverse thrust detection and limitation of usage may be required for certain airports due to noise protection regulations. Experiments were performed with the DLR Airbus A320 in order to discriminate thermal signatures for activated/non-activated reverse thrust after touchdown. The upper and lower pairs of images in Fig. 11 depict

Fig. 10 FLIR thermal imager (large circular Ge-window) as part of the DLR-RTO system at Braunschweig airport, integrated in a remotely controlled housing together with visual PTZ camera

Fig. 11 Thermal Image of turbine exhaust without (upper images) and with reverse thrust activated (lower images)

examples for landings with reverse thrust enabled (lower pair) and not enabled (upper pair) respectively.

5   Functional Tests and Verification

5.1   Measuring Camera-Display Latency

The operational requirement for the upper limit of delay between the situation at the remote airport and the video reconstruction at the controller working position is usually given as $\Delta t < 0.5$ s by the domain experts. It seems worth mentioning that the radar update time typically is 4 s due to rotation time. Also this fact underlines the relevance of the visual information if quick reaction of controllers is required.

5.1.1 LAN-Delay in the Braunschweig Airport Experimental System

For the experimental system at Braunschweig Research Airport a 500 m cable with multiple single mode optical fibers had been installed to realize a GBit LAN, in order to avoid any transmission bandwidth limitations. With this system it was possible to send light from the display side up to the camera position and illuminate the lens. With a light pulse sent up to the camera, a shutter signal indicating the send time could be used to measure the time difference between the signal arriving at the display and the start time defined by the shutter at the source position. The signals were measured by means of a dual trace oscilloscope yielding a camera—display latency of 230–270 ms.

5.1.2 WAN-Delay of the Long-Distance Transmission

For the long-distance transmission-time the measurement was somewhat more demanding. Of course in this case no light could be sent directly from the display position to the cameras. The total camera—display latency Erfurt-Braunschweig was measured by using a small procedure, which sends a defined command (e.g. "Close shutter") to one of the panoramic cameras and measures the time until the answer. Therefore it is necessary to check at first the transmission time by sending a "ping" from Braunschweig to Erfurt to determine the latency without any video-image processing. The next step is to monitor the transmission start time of the command sent to the remote camera in Erfurt and stop the instant when the effect of the command ("Close shutter" means black display) arrived on the display in Braunschweig. Finally the real video latency between the camera site and the display site (Braunschweig) is the measured (stop–start) time minus the half command-transmission time (for considering the runtime of the command from Braunschweig to Erfurt).

The results of this test procedure are delays between 270 and 330 ms. Considering the fact that the command to close the shutter could reach the camera at the beginning, in the middle or at the end of one frame, with one frame interval ca. 33 ms at a framerate of 30 f/s, a mean delay of 300 ms is obtained. This is in rough agreement with the latency measured optically with the fiber-optic LAN of the Braunschweig experimental system.

5.2 Electromagnetic Compatibility

Within the safety critical aviation environment new electronic systems for operational environments have to fulfill strict requirements concerning electromagnetic compatibility in order to guarantee zero electromagnetic interference with other electronic equipment. That is why electromagnetic emission of the prototype RTO system had

to be measured with certified equipment before installation in the operational environment for validation experiments (between antennas at the tower roof). The aim of the measurements was to show that the radio frequencies used at the airport will not be disturbed in a worst case scenario. These frequencies at the Erfurt airport cover a range between 140 and 160 MHz.

The electromagnetic compatibility tests were performed in cooperation with the StudING-UG, a company of the Institute for Electromagnetic Compatibility of Braunschweig Technical University that is certified for this kind of measurement.

All measurements were made with a Rohde & Schwarz ESCS 30 Spectrum Analyzer (Frequency range 9 kHz bis 2.75 GHz) and a Rohde & Schwarz HL023 logarithmic-periodic antenna (Frequency range 80–1300 MHz). The complete camera system (cameras, PTZ-camera, air blast cleaner and electronic equipment) was scanned from all four sides for its electromagnetic emission. The distance from the top of the antenna to the middle of the camera system was always 1 m. The antenna was placed at a height of 1.5 m. The variation of the measuring direction was realized by turning of the camera system in steps of 90° so that a reasonably consistent background radiation can be assumed. Each side of the camera system was measured both with horizontal and vertical polarization of the antenna.

Figure 12 depicts the experimental setup for measuring the spectrum of electromagnetic emissions of the camera system with visual PTZ and air blast cleaner before installation on the Erfurt tower.

The measurements were carried out with the following parameters:

- Frequency range 100–400 MHz
- Filter bandwidth 9 kHz
- Scanning steps 5 kHz
- Scanning time per frequency 1 ms
- Scan of background radiation with deactivated camera system

A scan of the background radiation with turned off camera system was carried out prior to each single measurement and was compared directly with the measurement of the radiation emission of the activated system.

Due to the long scanning time of nearly 15 min for every scan (more than 8 h in total) it was not possible to measure the radiation of the working PTZ-camera and the air blast cleaner without damaging both devices. That is why fast scans were carried out with a different filter bandwidth (120 kHz) and scanning steps of 60 kHz to obtain the emission of PTZ and air blast cleaner.

A typical result of background und emission scan is depicted in Fig. 13, with logarithmic radiation power plotted versus frequency.

The characteristic emission band at 170 MHz occurred more or less clearly in all measurements. After some shut-off tests of single components, the effect could be attributed to the network switch in the electronic box.

The measurements demonstrated that strong interferences can occur using various electronic devices (particularly network components) even if these devices are installed in metallic electronic boxes. But the measured interference frequencies in this case were not within critical operational frequency bands of radio frequencies

Fig. 12 Experimental setup for measuring electromagnetic emission
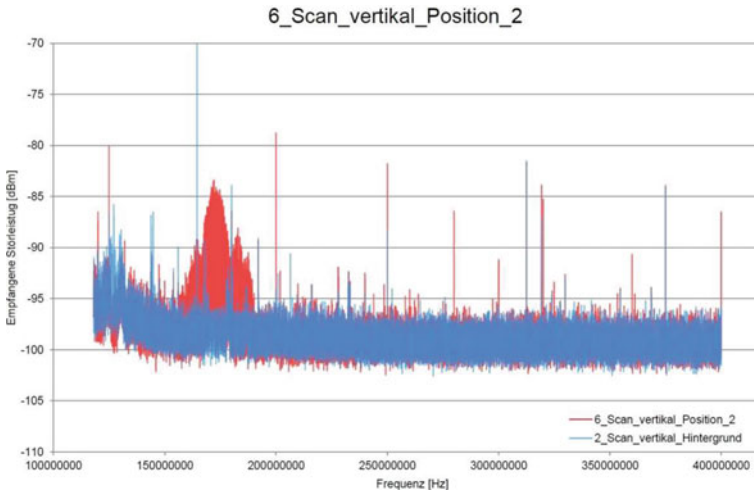

Fig. 13 Example of electromagnetic emission measurement. Horizontal axis: Frequency/Hz; vertical axis: radiation power/dBm. Detected emission band (red) around 170 MHz

at Erfurt airport. Therefore the camera system could be used without any restrictions on top of the tower roof close to the antennas of the radio communication system.

5.3 Image Optimization

After initial tests of the redesigned prototype RTO system at the remote airport as a first measure some public areas (streets, parking) had to be blanked out for reasons of privacy.

For the design and setup of the initial experimental system described in the previous chapter "Remote Tower Experimental System with Augmented Vision Videopanorama" the theoretical performance prediction and the field tests for verifying the corresponding data had focused on the expected pixel resolution. It was mentioned before that the pixel resolution represents only a very optimistic idealized value, the Nyquist limit in terms of spatial frequency (see Appendix A). It was evident from the beginning, however, that the optimization of contrast and dynamic range would play a crucial role in approaching the required system performance. The realization of the prototype system involved several measures addressing these questions with the goal of systematic improvement of the respective system parameters.

For the panorama optimization the following display-parameters have to be optimized:

1. working distance (defined by the special requirements of console and panorama display)
2. display size
3. pixel resolution (fixed by initial constraints: HD $1920 \times (1080 \times 6)$)
4. Luminance
5. static contrast
6. Gamma value

Some high-end displays for professional operation offer presets for playback in aligned video modes, where quite a few of these parameters are already optimized.

On the camera side the image quality is influenced by

1. sensor chip/pixel size
2. minimum aperture to avoid diffraction effects (dependent on illumination in automatic mode)
3. dynamic range (nominally 8 bit)
4. Gamma value
5. Bayer demosaicing algorithm
6. Codec/Decodec—methods

As alternative for using industrial cameras there are so-called smart cameras available, generally installed in security areas, that are optimized for video recording even under adverse environmental conditions (low contrast, sunlight, twilight etc.).

These cameras contain internal image data processing. The intrinsic latency, however easily exceeds an unacceptable level of 0.5 s.

In what follows we will address two aspects which improve the usability of the panorama system with regard to visual object detection: Gamma adjustment and local contrast enhancement.

5.3.1 Gamma Adjustment

In this subsection we will focus on the Gamma value which should be matched between camera and display for optimum image reconstruction (for details see Appendix A at the end of this volume). Due to the nonlinear input–output characteristics of cameras, displays, and the human perception the optimization of the systems Gamma value is of particular importance. Besides the usual display controls which include Gamma adjustment, in the experimental system a control menu is available which includes Gamma adjustment for each camera separately.

Usually the effective Computer-display Gamma characteristic (output luminance $\sim$ (input signal)$^\gamma$, see Appendix A) can be adjusted via settings on the graphics card. Typically the display exhibits a characteristic with $\gamma = 2 \ldots 2.5$. This results in low sensitivity with regard to luminance change in the dark range.

With sub-optimal γ-value small luminance differences vanish, dependent on illumination of the scenery, contrast and dynamic range. This is critical due to the limited dynamical range (8 bit) of the system corresponding to 255 discrete luminance intervals. That is why the camera γ-value should be adjustable to take account of specific local conditions. If the goal is a video reconstruction with 1 to 1 correspondence of the natural impression (=visual impression of real tower view with $\gamma_{\text{camera}} \approx 0.45$) this should be the setting of the camera. For a linear system input–output relationship this requires the display setting $\gamma_{\text{Display}} \approx 2.2$.

A practical example of the effect of gamma-adjustment is depicted by the following two photos of Fig. 14 with a high luminance display surrounded by persons in a dark environment.

Fig. 14 Effect of decrease of display γ to 2.5 for increasing luminance difference in dark area while keeping contrast in high luminance display area in an acceptable range (see also Appendix A)

5.3.2 Local Contrast Enhancement

During the initial field tests some observability problems became evident which were due to the limited dynamic range and contrast. Specifically some problematic movement areas with dark asphalt (e.g. heliport) were identified. If e.g. a dark helicopter was inside these areas, the contrast of the captured images was not sufficient for the controller to recognize the objects in the RTO-CWP. This problem was solved by locally enhancing the contrast at these specific areas to a level where it is possible to detect these objects. As an example Fig. 15a shows the original image of the heliport with a black helicopter that was not discriminable from the dark background, and Fig. 15b with the contrast enhanced image.

Fig. 15 Helicopter on heliport, **a** (top) original with helicopter not discriminable, and **b** (bottom) enhanced contrast with helicopter as black dot on gray background
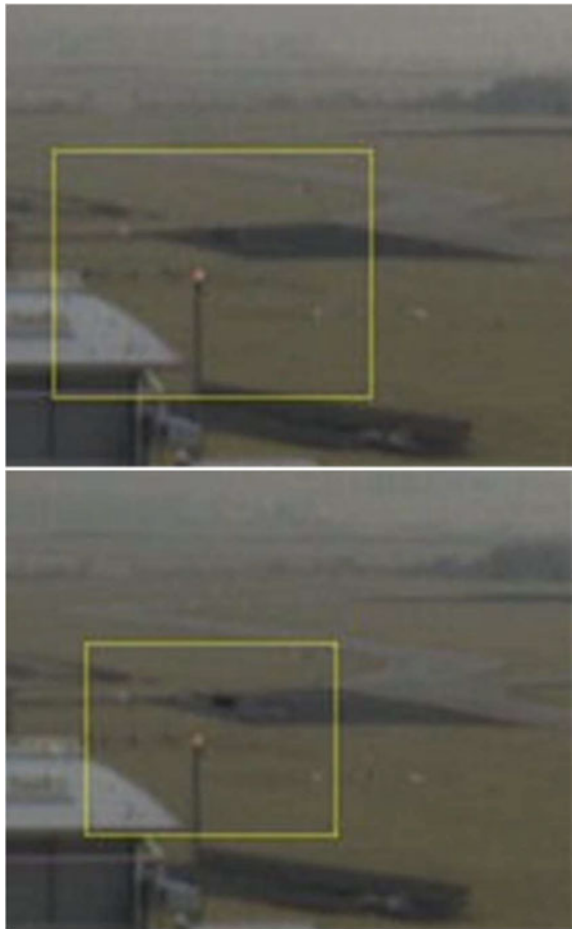
6 Conclusion

Important aspects of the design, development and verification of improved technical features of the Remote Tower prototype system is described that was used for the initial validation experiments at Erfurt airport in 2012 (Friedrich & Möhlenbrink, 2013; Fürstenau et al., 2014), and the following chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" and "Model Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position"). A comparable and extended RTO-system replaced the initial experimental one at Braunschweig research airport. The development was accompanied by a detailed work and task analysis including design workshops with domain experts in cooperation with DFS (the German ANSP). This resulted in an extended requirement list.

The prototype development included all basic RTO features for which at the time of the validation experiments stable and reliable functioning could be expected, such as long-distance high bandwidth live video panorama transmission, remote PTZ-control, local contrast enhancement. Advanced live raw data processing with new hard- and software allowed for 30 Hz video panorama framerate, approaching the value required for minimizing visual movement-discrimination errors (see chapter "Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration During Simulated Aircraft Landing" and Chap. 17). Due to their lower development status, more advanced features such as automatic movement detection and object tracking were investigated only with the local system at Braunschweig airport and not included in the remote one for shadow mode experiments.

The functional verification confirmed the fulfillment of EMC requirements for camera installation at the operational tower, the latency requirement, and specific local contrast requirements, partly achieved with some modifications after initial testing. The predicted limited visual resolution and contrast of the video panorama system made it clear that for the operational tests (chapters "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" and "Model Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position") the PT-Zoom camera and its usability would play an important role for (visual) situation awareness, certainly exceeding the use of binoculars in the standard tower environment.

and particularly for granting a half year extension of the project schedule in order to successfully finish the delayed validation activities.

References

Friedrich, M., & Möhlenbrink, C. (2013). Which data provide the best insight? A field trial for validating a remote tower operation concept. In *Proceedings of the 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM 2013).*

Fürstenau, N. (Ed.). (2013). Remote Airport Traffic Control Center (RAiCe). *DLR IB-112-2013/20.* DLR.

Fürstenau, N., Mittendorf, M., & Friedrich, M. (2014). Discriminability of flight maneuvers and risk of false decisions derived from dual choice decision errors in a video panorama-based remote tower work Position. *Lecture Notes in Artificial Intelligence (LNAI), 8020*, 105–114.

Fürstenau, N., Rudolph, M., Schmidt, M., Werther, B., Hetzheim, H., Halle, W., et al. (2008a). Flugverkehr-Leiteinrichtung (Virtueller Tower). *Europäisches Patent EP1791364.*

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, M., & Werther, B. (2007). Development of an augmented vision video panorama human-machine interface for remote airport tower operation. In *Proceedings of the 6th Eurocontrol Innovative Research Workshop* (pp. 125–132). Eurocontrol.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., & Halle, W. (2008b, September 14–19). Augmented vision video panorama system for remote airport tower operation. In I. Grant (Ed.), *Proceedings of the ICAS 2008, 26th International Congress of the Aeronautical Sciences.*

ICAO. (2013). *European (EUR) Regional Supplementary Procedures to Annex 2, 6, 10, 11, 15.* PANS-ATM (DOC 4444) & PANS-OPS (DOC 8168), ICAO.

Kanade, Lucas, & Tomasi. (1991). Detection and tracking of poijnt features.

Papenfuss, A., & Möhlenbrink, C. (2009). *RAiCe:Kognitive Arbeitsanalyse Lotsenarbeitsplatz, Internal Report DLR-IB 112-2009/20.*

Schmidt, M., Rudolph, M., Werther, B., Möhlenbrink, C., & Fürstenau, N. (2007). Development of an augmented vision video panorama human-machine interface for remote airport tower operation. In M. J. Smith & G. Salvendy (Eds.), *Lecture Notes Computer Science: Human Interface II* (Vol. 4558, pp. 1119–1128). Springer.

Shi, J., & Tomasi, C. (1994). Good features to track. In *Proceedings of the IEEE CVPR '94 Computer Vision and Pattern Recognition* (pp. 593–600).

Wohlfeil, J., & Börner, A. (2010). Optical orientation measurementfor remote sensing systems with small auxiliary image sensors. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS XXXVIII(1).*

Integration of (Surveillance) Multilateration Sensor Data into a Remote Tower System

Satoru Inoue, Mark Brown, and Yasuyuki Kakubari

Abstract Information from a surveillance sensor can be integrated into a Remote Tower System. Surveillance sensors provide position and altitude information that can be used to supplement optical sensor tracking, and a cooperative surveillance sensor further provides aircraft identification that can be correlated with flight information. In the Remote Tower research system developed by the Electronic Navigation Research Institute (ENRI), integration with a multilateral sensor allows integrated flight information to be displayed in a label (tag) associated with each aircraft, and surveillance position is combined with optical sensor information to assist automatic object-following by a PTZ camera function, which improves its performance particularly in low visibility conditions. To integrate surveillance information with optical target tracking accurately, calibration or correction techniques such as georeferencing (3D-2D mapping) are required. This chapter gives details on the following specific topics: (1) correlation of optically detected and tracked objects with surveillance sensor targets and labelling, (2) PTZ object following, and (3) mapping between two-dimensional screen coordinates and real-world three-dimensional PTZ coordinates.

Keywords Visual supporting · Multilateration surveillance sensor · Object following · Target tracking

1 ENRI's Remote Tower Research Project

The Electronic Navigation Research Institute (ENRI) has been conducting a remote tower research project to support expanding the capabilities of Japan's Airport Remote Mobile Communication Service. As mentioned in the chapter "Introduction of Remote AFIS in Japan", this service has for over 40 years been providing Aerodrome Flight Information Service (AFIS) to small airports with low levels of scheduled air traffic from Remote Air Ground communication (RAG) facilities (Remotely

S. Inoue (✉) · M. Brown · Y. Kakubari
Electronic Navigation Research Institute, National Institute of Maritime, Port and Aviation Technology, 7-42-23, Jidaiji-higashi-machi, Chofu, Tokyo 118-0012, Japan
e-mail: s-inoue@enri.go.jp

Operated Aerodrome Flight Information Service, 2012). Several RAGs are typically operated at a single Flight Service Centre (FSC) facility. The current RAG system uses ITV (Industrial Television) cameras for remote visual aerodrome observation since it has not been economically feasible to install surface movement detection equipment or terminal radar at such small airports. The initial phase of ENRI's remote tower project therefore dealt with purely optical surveillance. However, another ENRI project to develop a low-cost multilateration (MLAT) surveillance system called OCTPASS(Kakubari et al., 2014) has increased the economic feasibility of radar-like surveillance of the surface and surrounding airspace of smaller airports, while the Japan Civil Aviation Bureau (JCAB) has been considering upgrading the RAG from an advisory to an aerodrome air traffic control service. The scope of our project was therefore expanded to integrate surveillance sensors such as MLAT into the remote tower.

With the Airport Remote Mobile Communication Service, video images from ITV cameras and data from sensors such as surface wind sensors at the airport are transmitted to the AFIS operator at the RAG facility over a data communication network, and the operator provides necessary information to pilots through air-ground voice communication. Video images are considered as supplementary information to support operations and are not essential, although they increase efficiency by removing the need for some procedural steps when available. Although the system has limitations, it is adequate for airports with low scheduled traffic volumes that cannot economically justify a control tower. In the future, JCAB is planning to introduce remote AFIS and ATC services to airports with higher levels of traffic. This will require an expansion of the remote surveillance capability of the RAG to a level comparable to that of a conventional tower. ENRI's current research focus is therefore to support JCAB by developing and demonstrating a remote tower system concept and technologies, and formulating technical requirements for future systems, aimed at aerodrome ATC as well as AFIS. The technologies may also be used to supplement existing control towers.

ENRI's research on remote towers grew initially from investigations into applying new technological and ergonomic concepts to air traffic control, such as large-area displays with a unified human–machine interface. As part of these investigations, we looked at work in Europe, particularly at DLR and in Sweden, on the research and development of Remote Towers using a video panorama environment and supporting technologies such as information overlays (Fürstenau et al., 2016; Schaik et al., 2016; Schmidt et al., 2016), and began developments in a similar direction recognising the possibility of application to RAG airports.

This chapter consists of four sections. Section 2 gives an overview of the ENRI remote tower system—the system structure, specifications, functions and performance of the video panorama display, PTZ camera, and use of MLAT as a surveillance sensor. Section 3 explains about object following techniques such as detection and recognition based on image processing. In Sect. 4 we describe the integration of MLAT surveillance information for the tracking of aircraft and vehicles and their linking with flight information, and to improve the automatic PTZ following of moving objects. Section 5 summarises the chapter.

2 System Functions and Components

2.1 System Overview

ENRI's remote tower research system, shown in Fig. 1, comprises an operator console at its main laboratory in Tokyo and remote sensors at its Iwanuma branch office adjacent to Sendai Airport, approximately 300 km away. Sendai airport is a towered regional airport with many scheduled airline movements per day (over 50,000 annual movements in 2015) that hosts a regional airline base and two flying schools, one that operates twin-engined piston light aircraft and another that operates mostly helicopters. The variety of traffic, ranging from heavy jet transport aircraft to general aviation light aircraft and helicopters, some flying under instrument flight rules and others flying visually, makes it useful for remote tower testing. ENRI's research aircraft, a Beech King Air 350, is also based at Sendai airport and is available for remote tower testing and calibration.

The optical sensors consist of 12 fixed cameras mounted around a common vertical axis to give 360-degree coverage and a PTZ head with a camera and light gun (Fig. 2). These are installed on a 30-m-high tower immediately adjacent to the Sendai airport perimeter fence with views of both runways and most of the aprons and taxiways except for part of the main apron, which is partly obscured by the terminal building. In addition, there is a MLAT system with receivers at various points around the airport to give whole surface coverage, and an anemometer. Images from the fixed cameras are "stitched" and presented on 12 display screens at the Tokyo site to provide an out-of-the-window (OTW) panoramic view. The functions of ENRI's test system are described in more detail below.
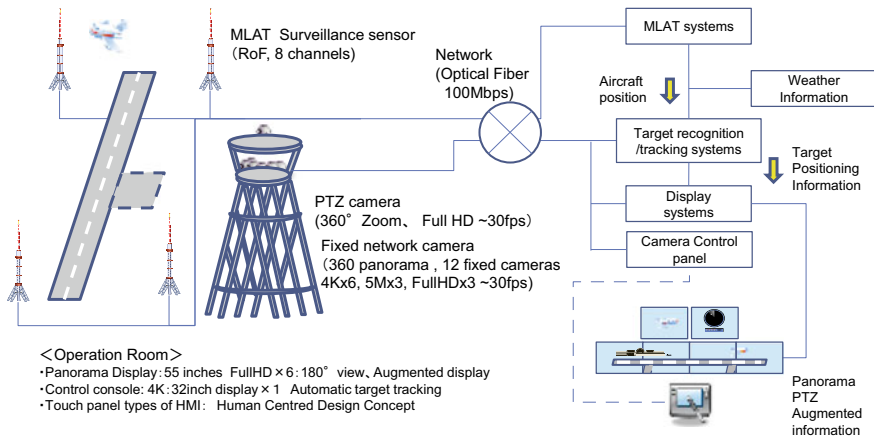

Fig. 1 Overview of ENRI's RT research system

Fig. 2 OTW cameras and PTZ camera system at Sendai airport

2.2 The Out of the Window (OTW) View System

2.2.1 Fixed Panorama Cameras

As described in Sect. 1, the system was initially developed targeting remote AFIS, so the initial panoramic OTW was a 180° horizontal field-of-view (HFOV) system (Fig. 3a) using six "Full HD" resolution cameras. This was further developed into a system also targeting ATC which uses 12 cameras to give a 360° HFOV (Fig. 3b). Several different types of camera are used to evaluate the effects of resolution on image quality and image recognition function performance. Of the 12 fixed cameras, six have 4 K resolution and are used for the 180° hemisphere facing the airport and covering the runways, aprons and taxiways. To cover the other hemisphere facing away from the airport and covering part of the visual traffic circuit, three 5 MP (2992 × 1620) cameras and three "Full HD" cameras are used. The data transfer rate per camera is 6 Mbps or less. The camera specification are summarised in Table 1.

The camera video update rate was reduced from the "native" 30–25 fps in consideration of flicker. The AC power supply frequency in eastern Japan is 50 Hz, and flicker might occur in images taken under artificial lighting at night depending on the camera's sensitivity (gain) setting. Flicker was avoided by reducing the video update rate to half the AC supply frequency, 25 fps.

The images from the panorama cameras are "stitched" at the display. Lens distortion correction applied during stitching will be described in detail in the next subsection.

(a)

(b)

Fig. 3 **a** OTW interface for remote AFIS **b** OTW interface for RT

2.2.2 Display Devices

Image data from each fixed camera is processed by a dedicated computer at the operator console to correct distortion and stitch the images together into a seamless panoramic image. The 360° panorama is currently displayed on 12 monitors, which is the same as the number of cameras but does not have to be. Each monitor shows

Table 1 Specifications of fixed cameras for OTW view

Cameras and image sizes (pixels)	3 × Full HD (1920 × 1080),3 × 5MP (2992 × 1620), 6 × 4 K (3840 × 2160)
Refresh frame rate	25 or 30fps
HFOV	33°/1 camera
CODEC	H.264
Latency (camera to display)	<0.7 s (including network)

an image with an HFOV of 30°, but the images from adjacent cameras must overlap to be stitched. Therefore, the actual image HFOV of each camera is 33°, and a 30° area is cut out from each image and stitched.

When combining images from multiple fixed cameras to make a single seamless panorama, it is not possible to simply abut the images because the edges will not match perfectly (Szeliski, 1996). Reasons for this include lens distortion and camera alignment errors including slight unintentional tilts. For stitching, it is necessary first to correct lens radial distortion (barrel or pincushion distortion, where straight lines are bent into curves), then to correct for misalignment between adjacent images. For each camera image, the parameters to be adjusted are the camera image's position in the overall panorama, horizontal and vertical scaling, rotation, and the quadratic coefficients of radial distortion (Jeught et al., 2012; Smith et al., 1992).

Correction for lens distortion is accomplished by correcting images of a calibration target with parallel horizontal and vertical lines, such as a chess board, prior to installing the camera in the field. Fig. 4 shows an example of image correction. The left picture shows an image before calibration and scaling, and the right picture is corrected for distortion and adjusted by fixed scaling after calibration. The red line is a vertical reference and meets the edge of a building and radar antenna tower at the top of the tower, indicated by the upper white dot on the line. In the left image in Fig. 4, radial distortion by the lens means that the bottom of the building does not meet the vertical line even though the camera not tilted (see the lower dots). In the right image, the radial distortion has been corrected by calibration from a test target image.

2.2.3 PTZ (Pan-Tilt-Zoom) Camera

The PTZ camera provides the same function as binoculars in a conventional control tower, allowing the operator to get a close-up view of a place or object. At the most basic level, the PTZ camera can be manually steered and zoomed by the operator to point to and magnify a desired object such as an aircraft or vehicle. To reduce operator workload, an object can also be automatically "captured" and then followed (tracked) when it moves based on image processing-based motion recognition supplemented by object position information from a surveillance sensor (the OCTPASS MLAT

Fig. 4 Comparison between "before calibration processing (left)" and "after calibration processing (right)"

system). Position information from the MLAT sensor improves the reliability of the PTZ to continuously following not only with vehicles and aircraft moving on the airport surface but also aircraft during takeoff and landing and in the traffic circuit. The automatic tracking function is described later in the section on target tracking.

The PTZ head also has a signal-light-gun pointing in the same direction as the PTZ camera, allowing the operator to optically signal aircraft without radio communication.

2.2.4   Network

The cameras are installed next to Sendai Airport approximately 300 km from the operator console in Tokyo. The remote site and console are connected by a commercially provided optical network virtual circuit with a guaranteed bandwidth of 100 Mbps. The network delay between the network interface units (CSU and DSU) at each site is about 30 ms.

2.3   The OCTPASS Surveillance Sensor

2.3.1   Overview

A multilateration surveillance sensor calculates the positions of transponder-equipped aircraft and vehicles by measuring the time differences between the reception of transponder response signals at multiple receivers in different locations. Transponder responses also include a four octal digit Mode A code or 24-bit ICAO aircraft address that can be correlated with flight plan information and allow "identification" of aircraft (in the air traffic control sense). MLAT sensors have a high update rate (typically around 1 Hz) compared to radar as well as high spatial resolution (less than 10 m position accuracy) are often used for airport surface movement surveillance.

ENRI has developed a low-cost MLAT sensor called OCTPASS (Optically Connected Passive Surveillance System). The advantage of OCTPASS is that receiver station sites can be designed simply. In mainstream multilateration, each receiver station site is provided with a signal processing unit for complex calculations, which requires a large housing. On the other hand, in OCTPASS, a receiver station has only to transmit received signals after optical conversion, which in essence requires just an antenna and radio-over-fibre (RoF) transmitter. These features result in compact size, light weight, low power consumption, and easy maintenance. Furthermore, the small receiver stations can be installed across an airport with more flexible layout. This system can surveil aircraft that are flying in the vicinity of an airport as well as on its surface, and has installed a research system at Sendai airport, although it is only used as a supplemental information device, and it's not used operationally. We are therefore using the Sendai airport OCTPASS system in ENRI remote tower research system as an external surveillance sensor instead of a surface movement radar and secondary surveillance radar. The Sendai airport OCTPASS system is shown in Fig. 5. It consists of a single transponder interrogator transmitter installed


Fig. 5 Locations of the MLAT receivers at Sendai airport

at ENRI's Iwanuma branch office and an 8-channel MLAT receiver connected to antennas installed at various locations within the airport perimeter (indicated by RX# in the Fig. 5) by optical fibre connections.

The horizontal position accuracy of the Sendai airport OCTPASS system is 5 m or less on the runways and taxiways, and the maximum error within the airport perimeter is no worse than 10 m. The position update rate is currently 1–2 Hz. Hyperbolic positioning calculation is used mainly inside and near the airport within the area of bounded by the sensors, and elliptical positioning is used away from the airport, with a maximum effective range of about 20 NM demonstrated by evaluation flights using ENRI's research aircraft, as shown in Fig. 6.

The OCTPASS receiver can also receive and process ADS-B signals, in which aircraft broadcast their positions from their navigation systems. This gives OCTPASS a hybrid positioning capability that allows surveillance of traffic even beyond MLAT coverage, as shown in Fig. 7. OCTPASS therefore combines high accuracy, high update rate surface movement, terminal area and en-route surveillance in a single low-cost system.

OCTPASS position information is also used in the remote tower system in a runway incursion warning function that alerts the operator if an aircraft or vehicle moves to within a certain distance of the primary runway.
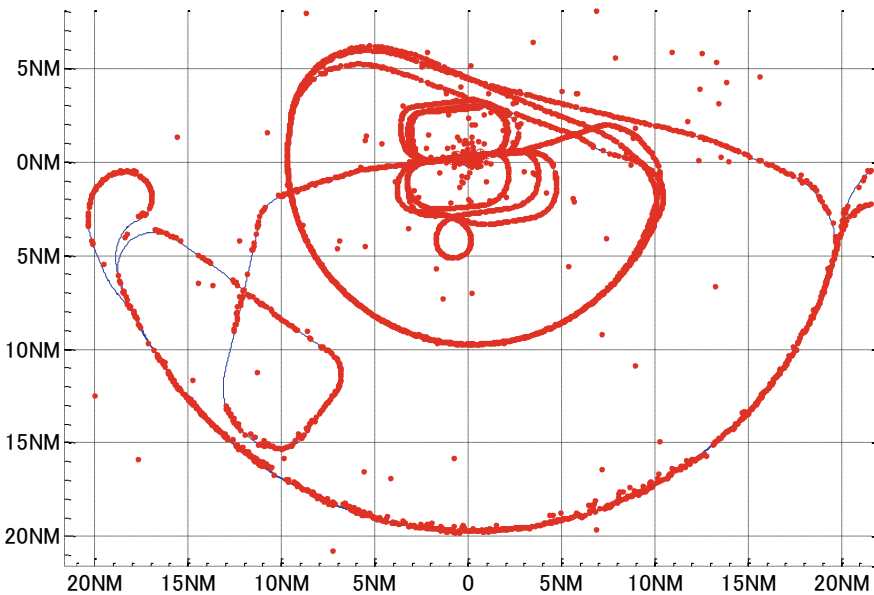
Fig. 6 Tracking of an aircraft in flight by elliptical positioning. Red dots indicate OCTPASS position samples. Blue lines indicate the flight path obtained from an on-board GPS receiver. The origin of the plot (0, 0) is at the ENRI Iwanuma branch office
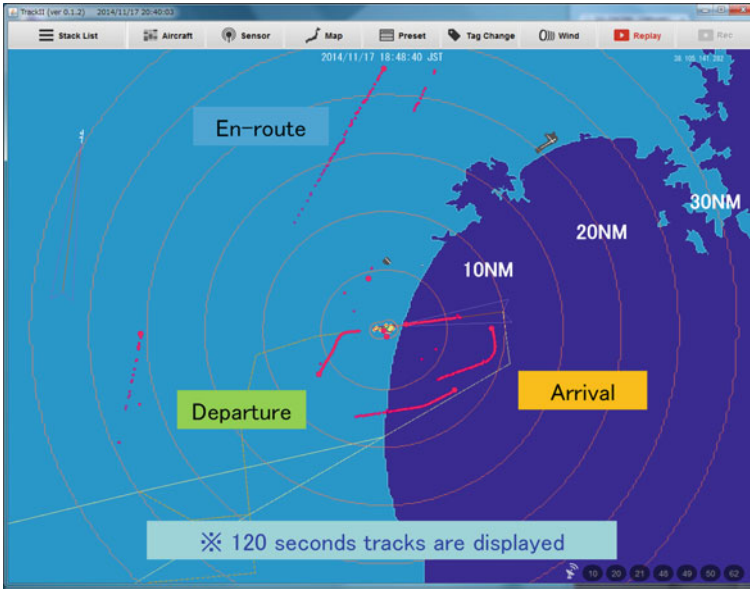
Fig. 7 Sample OCTPASS tracks of aircraft operating in the vicinity of Sendai airport. As well as "local" traffic in the terminal area (within 10 NM), traffic overflying at high altitude can be observed out to 20 NM or more by ADS-B

2.3.2 A Difficulty of Using MLAT

The difficulty with using MLAT position information in a remote tower system is that the receiving stations are all on the same ground plane, so the calculated altitude of airborne targets has poor accuracy. In the ENRI remote tower research system, we instead use the barometric altitude information encoded in the transponder response signal (Mode C or Mode S) corrected by local atmospheric pressure (aerodrome QNH). This information is only received about once every 2 s, however, which is lower than position update rate. Furthermore, the granularity of the altitude data may be as poor as 100 ft. If the altitude information is used "as is" with the horizontal position information, for example to display a symbol on the OTW panorama view or as an input to a tracking function such as PTZ object following, the vertical movement of an airborne object's symbol will appear stepped. To smooth out this effect, we therefore estimate the altitude for alternate position updates where no altitude information is available from previous altitude data, and interpolate for each image frame.

Fig. 8 shows a sample comparison of predicted altitudes (green plus symbols) with barometric altitude received from ADS-B (red circles) and Mode S transponder responses (blue circles). The granularity of the barometric altitude is 10 ft in this case. When the climb/descent rate is close to constant, the altitude estimate is good and the technique works well in giving a smooth symbol display or tracking. However,
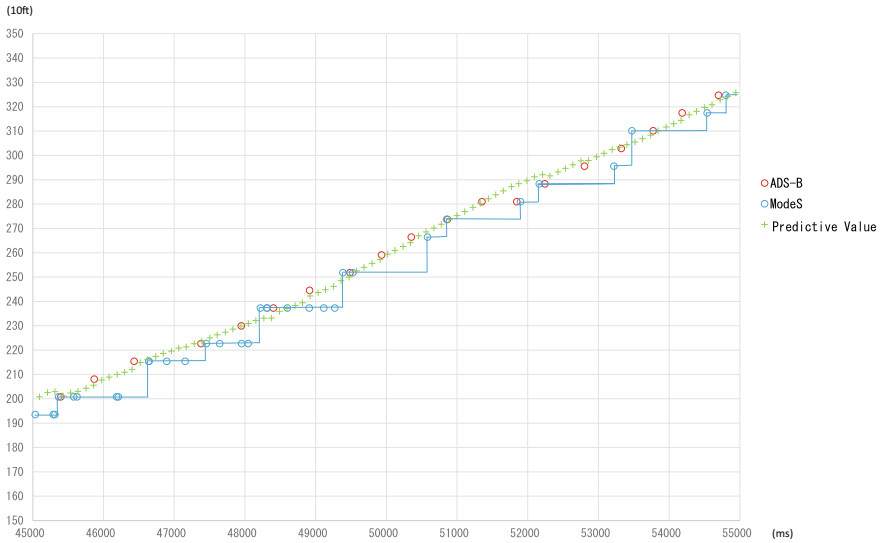
Fig. 8 Position prediction of altitude for value interval compensation

when vertical speed changes rapidly, such as when an aircraft lifts off from the runway and begins to climb, simple extrapolation from previous altitude data may be inadequate and the vertical tracking performance of an object-following function may be degraded in such cases.

3 Object Following Techniques

ENRI's remote tower system uses three mechanisms for object following (target tracking).

- *Video-based object detection and following*. This subsystem which is sometimes called a "video tracker" operates on the video streams from the fixed cameras and the PTZ. A hierarchical system of processing is applied. At the lowest level, foreground "objects" are discriminated from a static background by edge detection (Shrivakshan & Chandrasekar, 2012) and background subtraction (Lu et al., 2012; Matsuyama et al., 2006), detected as clusters of pixels. Temporal continuity is then applied via optical flow (Okada, Shirai, & Miura, 1996) or Kalman filter (Scott et al., 2009; Mandellos et al., 2011) to associate moving clusters of pixels with objects and improve the detection reliability.
- *Automatic object recognition*. The ability for the remote tower system to "recognize" objects is a useful function, allowing the system to reduce "nuisance" tracking of unwanted objects and other applications. Below, we describe tests we carried out of object classification using two techniques: an image-based

"deep learning" technique that attempts to recognise objects from their pixels, and knowledge-based classification that attempts to classify objects by their behaviours.

• *Position information from surveillance sensor*. Surveillance information from the OCTPASS sensor is used to augment visual object following.

3.1   Video-Based Object Detection and Following

Video-based object detection works by analysing pixels in both individual image "frames" (that is, images in a video stream) and changes between successive frames using classical image processing techniques of background subtraction and optical flow. Background subtraction assumes that foreground objects (objects of interest to the operator such as aircraft and vehicles) move against a background which is static and. Image frames are binarized (thresholded) to detect edges, and successive frames are then compared. The process is simple and since at every update only one frame needs to be thresholded and compared against the already thresholded previous frame, it can be carried out at high speed even when the number of pixels per frame is large. The performance of background subtraction depends on factors such as the brightness and contrast of foreground objects against to the background, and its sensitivity can be globally adjusted by varying the sensor gain and contrast. If the sensitivity is too high, "false" detections often occur due to changes in the image other than moving objects, such as brightness fluctuations and even image noise in low light conditions.

Background subtraction alone is insufficient for reliable object detection since it only detects clusters of moving pixels between the previous and current frames. This means that it cannot handle situations such as two moving objects that overlap as they pass each other, or moving objects partly occluded by static obstructions closer to the sensor. Dealing with such situations requires looking at the temporal continuity of clusters of pixels over a longer time span than two frames. Optical flow and the Kalman filter are effective for this. Optical flow can detect the motion vectors of clusters of pixels in the image over successive frames and so can detect moving objects; clusters of pixels detected as "objects" by background subtraction that do not have temporal persistence can be rejected as noise. The Kalman filter can further estimate the future position of a cluster of moving pixels and so can monitor the continuity of moving objects even when they are partly occluded or moving past another object. Fig. 9 shows an example of moving object detection using optical flow. The green square patch indicates an area of the image in which the motion vector of a cluster of pixels is detected; in this case, an aircraft on final approach for landing.

To summarize, edge detection and background subtraction are simple and quick but are easily prone to false object detection and cannot handle cases of partial occlusion or objects moving past each other. Optical flow is more reliable because it uses temporal continuity, but calculating motion vectors between successive frames

Fig. 9 Moving object detection by optical flow. The above image is a still video frame showing an aircraft on final approach and static aircraft parked on the apron the bottom image shows the optical flows: red areas indicate left-to-right and green indicate right-to-left. The approaching aircraft (moving left) has been successfully detected

has a higher computational load and is therefore more time consuming. The ENRI remote tower system employs features of both techniques, and can give sufficiently reliable object detection at frame rates in excess of 30 fps when tuned. Clusters of pixels that have temporal persistence are determined to be objects and assigned a tracking identifier (ID) by the system further reduce the likelihood of detecting a moving cluster of pixels as two separate objects.

3.2 Automatic Object Recognition

The ENRI remote tower system uses two methods for object classification and discrimination (that is, determining what an object is); one image-based and one knowledge-based. The image-based method uses "deep learning" with an artificial neural network called a Convolutional Neural Network (CNN) to identify the class of an object (aircraft, vehicle, person etc.) from its pixels (Krizhevsky et al., 2012; Jia et al., 2014). The knowledge-based recognition method uses knowledge of the expected position and behaviour of aircraft to discriminate airborne aircraft from other moving clusters of pixels even when their image are too small for image-based classification.

3.2.1 Convolutional Neural Network Image Recognition

Object Recognition

A CNN is hierarchy of layered artificial neural networks that can classify objects from their images. The process of machine learning-based object recognition using a CNN is shown in Fig. 10. The network extracts feature points from input images, and "learns" to associate the those features with certain object classes by being presented with many samples of "training data"—images that are labelled as to whether or not they correspond to the object class to be recognized. Once trained, the CNN can then be used to classify arbitrary images containing those objects; when presented a
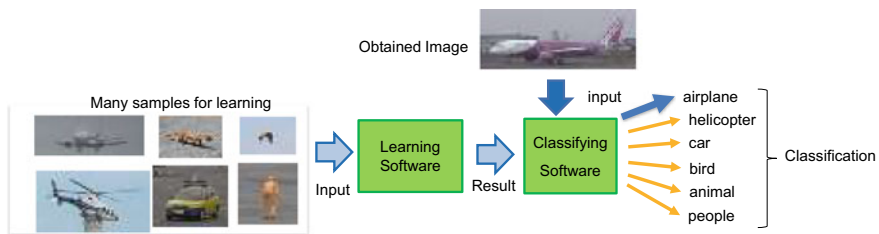
Fig. 10 Machine learning-based image recognition system

Table 2 Example of image recognition result by using CNN base algorithm

Input image data								
recognition rate	0.9999	1.0000	0.9997	0.9998	0.9987	0.9987	0.9997	0.9975	0.9975

stimulus image, it outputs a confidence from 0–1 that image matches an object class it learned (Fig. 11).

CNNs tend to require a large amount of training data required to avoid the phenomenon of overfitting: with too little training data the classification can be too specific to the training set and fail to classify more general images. Another problem is that it is possible to unintentionally introduce hidden biases in training. For example, if many vehicle images in the learning data have a grass background while aircraft are mostly of aircraft in flight with a sky background, the CNN might learn the background instead of the object so that after training, when presented with an image of aircraft on a runway, the grass background behind the runway may cause the CNN to misidentified it as a car. Care is therefore needed in the selection of the training data.

Using CNNs for real-time classification applications requires large amounts of processing and can therefore take time. However, since the processing can be extensively parallelised, the use of Graphics Processing Units (GPUs) can improve the speed considerably.

Classification Reliability Experiments

ENRI evaluated image classification by a CNN-based framework in 2018. After training with 9,000 sample images of six types (aircraft, automobiles, helicopters, humans, animals), images of airborne aircraft could be classified with an output confidence of 99.9% or greater. Table 2 shows examples of images of aircraft in flight extracted from the OTW panorama view and the output confidence. The images are 120 pixels wide and 80 pixels or less tall.

However, when classifying an aircraft on the ground, some mis-recognitions occurred; that is, there were cases where the confidence of matching another object class was greater than that of an aircraft. Table 3 shows some examples. In each case, the output confidence of the "car" object class is higher than that of "airplane". (It should also be noted that the confidences of "people", "animal" and "helicopter" were all below 1%.) Analysis of these mis-recognitions show that if the aircraft's image partly blends into the background so that the features that distinguish it as an aircraft are not prominent, the similarity with images of a vehicle against a like background is high, increasing the likelihood of misidentification of the aircraft as a car. We are currently investigating whether the reliability of recognition can be improved by increasing the variety of training data.

Another drawback of image-based recognition is that a certain minimum number of pixels of an object is required to give sufficient feature points for classification. If

Table 3 Example of mis-recognition cases in the test

0.5677 − "car"	0.6608 − "car"	0.5442 − "car"	0.5124 − "car"
0.4313 − "airplane"	0.3388 − "airplane"	0.4520 − "airplane"	0.4856 − "airplane"
0.0009 − "helicopter"	0.0002 − "animal"	0.0026 − "people"	0.0012 − "people"
0.0000 − "people"	0.0001 − "people"	0.0009 − "animal"	0.0006 − "animal"
0.0000 − "animal"	0.0000 − "helicopter"	0.0002 − "helicopter"	0.0001 − "helicopter"

an object's image is too small, the confidence and reliability of the recognition tend to decrease. Although image recognition could be used for reliable object classification in areas where the images of objects are sufficiently large, such as for aircraft and vehicles manoeuvring on aprons or the runway, it is less able to recognise small and/or distant objects.

Object Search

Although it is difficult to compare processing speed independently of hardware, an off-the-shelf CNN software package required 300 ms for detection and classification of an image of a single object against a background using an Nvidia Tesla K20 general-purpose GPU (GPGPU). (Processing time was largely independent of the number of pixels in the input image since the matching process dominates the processing time.) However, an OTW panorama image may contain many objects. If object search is to be added to basic recognition and multiple objects are to be recognized, the processing time increases. CNN-based algorithms have been developed that extract targets from an image as well as recognise them, for example the well-known R-CNN (Girshick et al., 2014) and its derivatives Fast R-CNN (Girshick & Fast, 2015) and Faster R-CNN (Ren et al., 2017). However, R-CNN-based algorithms require a large amount of calculation, so they are difficult to apply when real-time performance is desired.

Instead of R-CNN, therefore, ENRI evaluated target extraction and recognition using the CNN-based YOLO (You Only Look Once) framework (Redmon et al., 2016) which can perform target extraction as well as recognition in real time (Fig. 11). In an initial test using same hardware as the CNN test described above (Nvidia Tesla K20 GPGPU), the YOLO V3 (Redmon & Farhadi, 2018) algorithm detected and recognised multiple targets an image in around 100 ms. Moreover, YOLO is still evolving, with newer versions of the YOLO algorithm now available, V4 (Bochkovskiy et al., 2020) and V5, which are being continuously developed. Processing time appeared to be independent of the number of targets, at least up to a few tens of targets, but the target extraction rate and recognition rate varied depending on the location and the size of the target, and is difficult to give a single value that can represent expected real-world performance. It is necessary to thoroughly consider
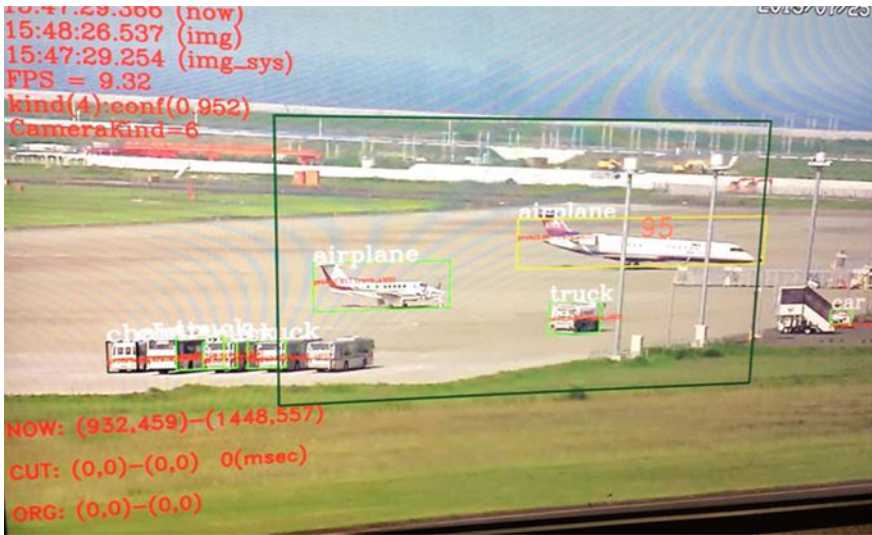
Fig. 11 Example of YOLO detection

the performance such as the application area and the size of the target to be extracted in the system design.

Summary

Several high-speed processing algorithms for target extraction and recognition based on CNN have been developed, and we are still investigating which algorithm is appropriate. For practical application, it is necessary to determine the appropriate trade-off between recognition reliability and real-time performance from the results of case studies.

3.2.2   Knowledge-Based Processing

Object recognition using machine vision is only effective when there are enough pixels of an object for the system to detect its distinguishing features. The OTW image of the ENRI system has at least 1,080 pixels per 30° HFOV, and image-based object recognition can be used on the runway and on taxiways and aprons even though the image sensor is over 1 km from the main apron. However, airborne aircraft are resolved only as small dots until they are close to the airport, and image-based classification of objects as fixed-wing aircraft, helicopters or birds will either at least unreliable if not impossible. However, in actual operations, operators may "recognize" an object as an aircraft or a helicopter even if it appears only as a distant moving point. It is thought that this recognition is based on the operator's knowledge and expectation of aircraft operations (Kang et al., 2014) at the aerodrome. We therefore experimented with knowledge-based processing to discriminate distant

moving objects as aircraft or not, even if they are displayed as little more than a dot, with the aim of improving the reliability of airborne object detection and tracking.

Specifically, the system applies knowledge such as information on the areas in which aircraft normally fly and their typical trajectories to the detected speed and direction of object motion to reject objects whose position and motion are not consistent with these criteria, thereby reducing "false targets". However, it is necessary to create a knowledge database such as an ontology network which is specific to each airport, and the accuracy of the false target rejection depends how detailed the described rules are. Theoretically, more detailed descriptions increase the processing burden, requiring faster, more expensive hardware if real-time performance is required (although persistence of discrimination could be applied to reduce the demanded processing speed). This gives a cost–benefit trade-off. ENRI is still conducting ongoing studies and evaluations, but so far knowledge-based processing has been effective in eliminating clusters of moving pixels associated with clouds from object following.

3.2.3 Summary

The ENRI remote tower research system applies video-based object detection combined with knowledge-based discrimination to detect and follow moving objects in OTW panorama images, even if the images of the objects are occluded when moving past a closer static object or when passing and overlapping another moving object. Image-based object classification is used mainly by the PTZ camera object-following function, and is also used recognize aircraft displayed on the OTW panorama in certain areas where they are sufficiently large, such as aprons, runways and taxiways.

Since the PTZ gives a magnified view, there tend to be fewer objects visible since the field of view is narrow, and more pixels for each visible object compared to the OTW view. At the same time the image pixel dimensions are smaller than the OTW image. These characteristics make the PTZ image more amenable to image-based object recognition, with an accuracy of 99% achieved in discriminating between aircraft and vehicles, enabling highly reliable PTZ object following.

4　MLAT Integration

4.1　Target Tracking on OTW (Video + MLAT)

A surveillance sensor can give position and possibly altitude information on aircraft (and some airport vehicles) operating on the aerodrome surface and in its immediate vicinity. In addition, Mode A code or 24-bit aircraft addresses obtained from transponder responses or ADS-B transmissions can be correlated with flight data

processing systems to give callsign, aircraft type and other information such as destination (of a departure) or allocated stand (of an arrival). By mapping the surveillance sensor-derived position of an aircraft onto the two-dimensional OTW screen, a symbol may be drawn to show the aircraft's location and textual information can be shown in a label close to it. This can improve the operator's situational awareness, is convenient, and helps the operator maintain a continuous visual watch of the aerodrome rather than having to go "heads down" to look down for information on separate displays. The ENRI remote tower research system allows selective labelling of objects with information required for AFIS and ATC purposes.

To realize this feature, an object detected by the remote tower's video tracker must be linked to data on the same object detected by MLAT or another radar-like surveillance system. However, the two corresponding pieces of information are almost never exactly the same, particularly since the sensor update rates are different (e.g. ~25 Hz for the optical surveillance sensor versus 1–2 Hz for an MLAT system such as OCTPASS). It is necessary to have a technique to combine the information from multiple sources into a single object for display and other purposes, a process which is sometimes called "sensor fusion". This and other technical issues related to this problem are described below.

It is requisite that 3D positions (latitude, longitude and altitude) from a surveillance sensor can be accurately mapped to pixel locations on the OTW image (for objects on the ground, height is implicitly zero although the airport surface may have slight undulations). Normally, if the image point corresponding to such a surveillance position is close to where an object is detected by image processing on the OTW image, the system "connects" the object's video tracker ID with the MLAT target ID and labels it with MLAT-derived information (Fig. 12).

However, if the position indicated by MLAT is not close to an object detected by the video tracker, it is not straightforward to determine which object is associated with the MLAT information, and heuristic rule-based methods must be used. One method is to make persistent a previous connection between an object's MLAT target ID and its video tracker ID when their respective mapped and screen locations were close together. With this method, even if the detection boxes for other IDs come into close proximity later, it is unlikely that the labels will be "swapped" between them. The "connection" (association) status is indicated on the object's label by a "C" (see Fig. 12).

If multiple objects are detected by the video tracker near an MLAT display position, it is not possible to judge which should be associated with the MLAT target from this information alone. In this case, a "W" is displayed on the label to indicate that the box indicator from the image processing object detection and the label derived from the MLAT target data are not associated by the system. The system independently displays the detected positions (the box and the label are separated) of the target on the image and the target of MLAT. When several labels and moving objects (airplane, car, etc.) must be linked, the logic that determines whether to the link a detection box and MLAT information uses parameters such as the distance between the object's MLAT and video tracker positions, movement vector etc. Tuning the

Fig. 12 Examples of object detected by image processing labelled with MLAT information

logic and parameters that determine the labelling association may improve the accuracy of association; however, such tuning is site environment-specific, depending on the layout of the airport and the characteristics of the images obtained from the optical sensors.

4.2 Example of a Phenomenon Caused by MLAT Issues

As described earlier, the update rate of MLAT altitude information is much lower than the optical object detection refresh rate, so the ability of an MLAT position/altitude-derived symbol (e.g. label box or leader line) to smoothly follow an aircraft's image on the OTW presentation will be poor if the aircraft has a large rate of change of altitude, such as when taking off. Even if the horizontal position of an object is well-estimated from MLAT information, a vertical error is inevitable in such situations. In practice, such an error is greatest when the aircraft starts to climb away from the runway after its takeoff run. However, as the climb progresses the altitude estimation error decreases and the aircraft's distance from the optical sensor increases, so the mismatch between the aircraft's image and its label decreases.

## 4.3 PTZ Camera Automatic Tracking (Video + Monitoring Sensor)

The ENRI remote tower system can automatically follow an aircraft with the PTZ camera when its flight strip is selected (Fig. 13) by linking identification information from the MLAT surveillance sensor with flight information from the flight data processing system.

When an aircraft to be followed is designated by the operator selecting its flight strip, its latitude, longitude, and altitude information are obtained from the MLAT surveillance information and the PTZ camera is pointed in that direction. The PTZ image is then analysed using machine learning-based recognition to distinguish aircraft from other objects, and the camera is controlled to centre the desired aircraft in the image (Fig. 14).

During continuous PTZ tracking, the aircraft's positions from the video tracker and the MLAT sensor are checked for consistency each frame. If a large discrepancy arises, the aircraft will be searched for again by analysis of the PTZ camera image. If tracking data from either image analysis or MLAT are lost, the PTZ camera will continue following using the remaining position source. Even in situations where a video tracker has difficulty, such as in poor visibility or when the aircraft is momentarily obscured by an obstacle or cloud, it is possible to continuously track it by

Fig. 13 Interface of electric flight strips for automatic target tracking function
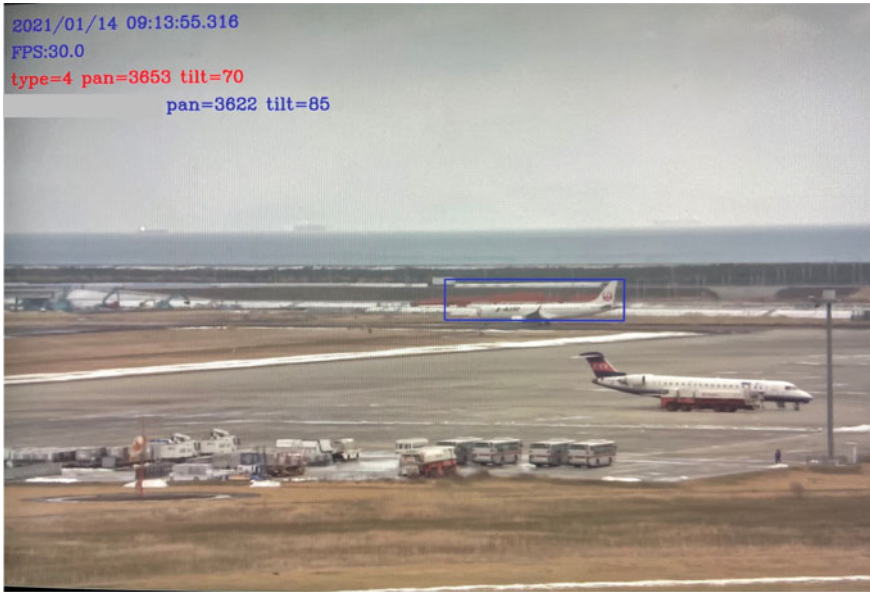
Fig. 14 Selected target image recognition based on image learning (YOLO)

pointing the PTZ towards the MLAT position. This automatic tracking function has proved reliable and reduces operator workload.

An automatic tracking function using surveillance sensor positions can also capture and follow an aircraft at distances greater than the panorama camera or PTZ camera detection ranges, as might occur with aircraft approaching the aerodrome for landing.

4.4 Integrating Panoramic Image, PTZ and Surveillance Information

It is necessary to perform calibration in order to integrate 2D video images and 3D surveillance sensor information. Concretely, a mapping between 2D display coordinates and 3D world coordinates is required to display position and altitude information obtained from a surveillance sensor such as MLAT at a location or an object's position on the panoramic OTW image. This mapping is determined by the following three steps:

1. Obtain a mapping between points on the OTW image and PTZ camera pointing direction (pan/tilt angles).
2. Obtain a mapping between PTZ camera pan/tilt angles and real-world 3D vectors.

3. Obtain a mapping between 3D real-world positions and 2D OTW display pixels through the PTZ camera axis coordinate system.

This method converts between the OTW's 2D image coordinates and 3D real-world coordinates using the PTZ camera's orientation as a common coordinate system that connects the two. Errors are corrected by calibrating the conversion between the OTW's 2D plane coordinates and PTZ camera's spatial coordinates, and by calibrating the conversion between the PTZ camera's spatial coordinates and 3D real-world coordinates. Finally, the error between the 3D coordinates and the 2D plane coordinates is checked. Details are described below.

4.4.1 OTW-PTZ Pan/Tilt Angle Mapping

The pan angle and tilt angle values of the PTZ camera when commanded to point to a number of calibration points displayed on the OTW screen are measured, allowing the conversion between PTZ camera pan/tilt angles and OTW coordinates to be calculated via a transformation function that is optimized to minimize error. The following conditions are set as prerequisites.

- OTW camera images are corrected for lens distortion
- OTW camera images are corrected for rotation error (i.e. horizontals and verticals in the image are true).

The transformation function parameters are obtained as the solution of a nonlinear optimization problem that minimizes errors. In other words, we calibrate the mapping between the 2D OTW image coordinates and PTZ camera pan/tilt angles by estimating the parameters of a PTZ camera pan-tilt calculation model of the mapping that minimize errors. The transformation parameters to be optimized are given in Table 4.

The (X, Y, Z) coordinate system is as shown in Fig. 15. The input information to this coordinate conversion are OTW image coordinates (u, v) and the pan and tilt angles of the PTZ camera (φ, θ) as shown in Table 5.
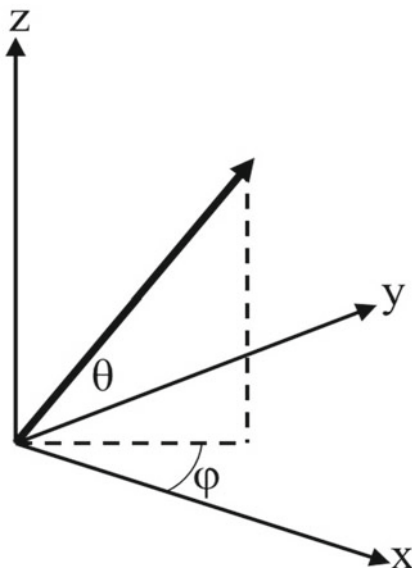
Coordinate system

The XYZ coordinate system is a right-handed coordinate system with XY plane horizontal and the Z-axis pointing vertically upwards. The origin is on the ground

Table 4 Transformation function parameters for optimization	1. Parameters for each OTW image sensor: Sensor position (X, Y, Z) Sensor pointing direction (azimuth) Horizontal and vertical field of view angles
	2. PTZ platform parameters: PTZ platform orientation (inclination from horizontal) Z position

Fig. 15 PTZ camera
pointing direction (blue
arrow) in OTW-PTZ
mapping coordinate system

Table 5 Input measurements
for calibration

• Spatial positions (latitude, longitude) of three surface points that define a ground reference plane
• Projected position of the PTZ camera onto the ground reference plane (this is always assumed to be the origin)
• Projected positions of each OTW image sensor onto the ground reference plane (with the PTZ camera at the origin)
• OTW image (u, v) coordinates of calibration point landmarks
• PTZ camera pan and tilt angles (φ, θ) when the calibration point is centred in the PTZ camera image

level immediately below the centre of the PTZ camera head; that is, the PTZ camera head is along the Z-axis with (X, Y) = (0, 0). The X axis is arbitrary but would typically be chosen in the direction of North or a prominent fixed feature. In Fig. 15 below, the blue arrow shows the optical axis (pointing direction) of the PTZ camera in the coordinate frame.

- The PTZ camera pan angle is defined to be 0 degrees along the positive X-axis
- Pan (azimuth) and tilt (elevation) angles are defined as follows as shown in the diagram
- Tilt angle is the angle θ between the axis of the PTZ camera and the XY plane
- Pan angle (azimuth) is the angle φ from the X-axis to the projection of the PTZ axis onto the XY plane.

We used Simulated Annealing (Kirkpatrick et al., 1983) and the Nelder-Mead method (Nelder & Mead, 1965) to search for the optimum values of the parameters

in Table 4. Calibration was performed on each of the 12 fixed OTW cameras and the PTZ camera of the ENRI system.

Fig. 16 shows an example of errors at various image points after calibration was performed using reference points on the airport surface and also on distant mountains that were above the airport ground plane. In the example shown, errors at points on the airport surface points near the mountains are within 3 pixels. When positions on clouds were used to verify the calibration in sky portion of the image, the accuracy was not stable because the clouds were moving, but there were no significant errors observed when tracking aircraft taking off.
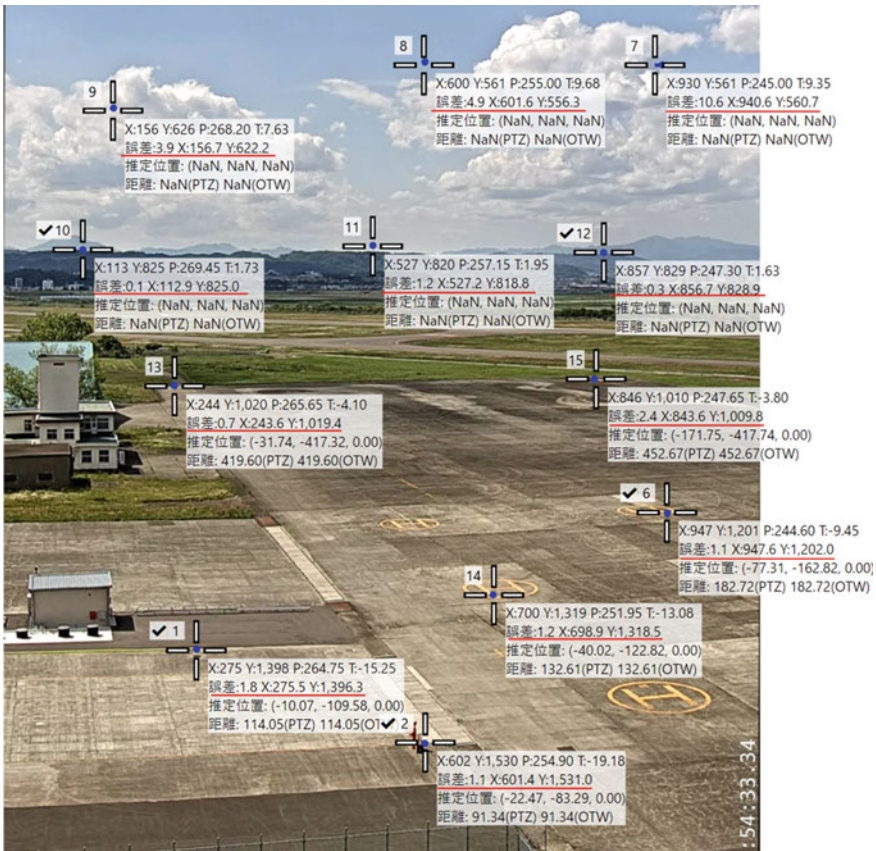

Fig. 16 Example of Calibration result of OTW-PTZ translation (Red lines indicate error values)

4.4.2   Mapping Between PTZ Pointing Direction and Real-World Position

In order to use a sensor such as MLAT to cue the PTZ to point to an object, a transformation function is needed between the object's latitude, longitude and altitude obtained from the sensor and the PTZ head pan angle and tilt angles required to point PTZ camera at that position. The transformation uses a model that calculates the real-world coordinates (latitude, longitude, altitude) of a target point from the PTZ camera's real-world location and the pan and tilt angles of its axis.

Similarly, with the OTW-PTZ pan/tilt angle mapping, calibration is achieved by optimizing the transformation model parameters to minimize the error, using the sum of the squared norms of the error of the calculated coordinates as the objective function. The parameters are the PTZ camera's installed position and its inclination to the horizontal. The calibration input information are the three-dimensional coordinates (latitude, longitude, altitude) of some chosen calibration points and the PTZ camera pan and tilt angles when those points are centred in the PTZ camera image. A search method such as the Nelder-Mead method is used for the optimization as in the case of OTW-PTZ calibration. Multiple optimum solutions that correspond to local minima may exist, but it is necessary only to search those parts of the parameter space that correspond to the actual ranges of the parameter values.

The calibration is verified by measuring the errors for some arbitrary real-world points (landmarks) that were not used for calibration. The landmark coordinates are transformed by the calibrated transformation function to obtain PTZ pan and tilt angles and the PTZ camera is moved to those angles. The error is the offset of the landmark's position in the PTZ image from the centre pixel, and if the error value is smaller than before calibration and is within the permitted tolerance, the calibration is considered as verified.

Table 6 shows results from sample calibration verification. Error values were 0.2° or less at points less than 1,000 m from the PTZ camera. On the runway about

Table 6 Example of error gap result of calibration from 3D to PTZ			
Point No	Error of pan (°)	Error of tilt (°)	Distance (km)
1	0.043632	−0.137233	1680.635
2	−0.084611	−0.101805	950.340
3	−0.23887	−0.085208	182.613
4	0.034572	−0.081441	113.151
5	−0.08842	−0.003554	483.058
6	−0.192249	0.055744	104.066
7	−0.127563	0.100433	371.682
8	0.233032	0.197523	492.137
9	0.050956	0.241507	128.132
10	0.153926	0.243569	181.377
11	0.095249	0.274863	115.464

Fig. 17 Example of
Calibration result of
coordinate transformation
from 3 to 2D (almost 1.5 km
from the camera position to a
reference point on the airport
control tower

1,000 m from the PTZ camera, it was confirmed that the errors of pan and the tilt
were corrected to within 0.1°.

4.4.3 OTW-Real World Position Mapping

Finally, OTW-3D Real World mappings are calculated with minimal error by
combining the above two calibrated coordinate transformation functions. In theory,
there is no need for calibration, but the error must be verified.

To verify the accuracy of the mapping, the measured 3D positions of arbitrary
reference points visible on the OTW display are converted to OTW display coor-
dinates via the above two calibrated transformation functions, and the calculated
pixel coordinates are compared with the actual pixel coordinates at which they are
displayed to verify that the errors are within a specified tolerance. By this method, it
has been confirmed that the error of ENRI's remote tower system does not exceed its
specified tolerance of 8 pixels, so no further correction is required. In the example
shown in Fig. 17, the point shown in red on the control tower is a reference point, the
yellow dot shows its location on the OTW image calculated from PTZ coordinates
using a model before calibration, and the blue dot shows its screen location calculated
after calibration.

4.4.4 Further Remarks

A great practical advantage of this method of mapping between pixels on the 2D
OTW image and 3D real-world positions via the PTZ camera's pan/tilt coordinate

system is that it is not necessary to measure precisely the position and inclination of the PTZ camera head. This is because the PTZ position and inclination from the horizontal used in the coordinate conversion are adjusted to minimize the error of the objective function during calibration. This can reduce the installation effort and hence cost.

The calibration described above uses a number of static features that are visible on the OTW image and measures their geographical and 2D display positions. The accuracy of the calibration increases with the number of points used, but in practice it is sufficient to use only as many points as needed to reduce the error to within a certain tolerance. To calibrate the ENRI remote tower system, we used as calibration points three or more locations within the airport perimeter; features such as the corners of buildings, runway and taxiway markings, radio navaid antennas and lighting installations. The calibration error using just three points was within 10 pixels on the screen in the horizontal direction. A problem is that when there are few features above the plane of the aerodrome surface, it is difficult to verify the height error above the runway with this method.

5   Summary

With a Remote Tower, it is possible to introduce support functions that can reduce operator workload and improve situational awareness such as object tracking, object recognition, and surveillance sensor integration. On the other hand, there are challenges due to the current limitations of surveillance sensor capability and accuracy. This chapter introduced two functions of ENRI's remote tower research system that were technically challenging to develop; object labelling on the display and object tracking. Object labelling displays target information obtained from a surveillance sensor on the OTW panorama view, and requires image correction, surveillance sensor integration, and coordinate conversion. Object tracking combines image processing-based target detection and motion extraction with object recognition and surveillance sensor integration, and also requires coordinate conversion and calibration to assist automatic PTZ cueing. Here, we first introduced the image correction and surveillance sensor technologies that are the basic components for these functions. We also introduced object recognition using neural networks and knowledge-based techniques and their issues. In addition, we presented our efforts on coordinate transformation and calibration that are needed for system integration.

In our verification efforts so far, image-based target detection and object recognition have obtained highly accurate results under good conditions, but issues such as misrecognition require further efforts to address, and we will continue to consider measures such as improving the algorithms and learning including training set construction. Greater object recognition accuracy can improve tracking-related functions and is also expected to be applied in other support functions in the future. System integration-related techniques of coordinate transformation and calibration are important for support systems such as labelling and tracking that are considered

to be key enhancements for remote and digital towers. The calibration technique introduced here was developed by ENRI specifically for its remote tower system, but there are various alternative methods and it is important to consider a method that suits each system. We will continue to evaluate and work to improve the issues.

References

Bochkovskiy, A., Wang, C., & Liao, H. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection.* arXiv:2004.10934 [cs.CV], https://arxiv.org/pdf/2004.10934.pdf.

Fürstenau, N., & Schmidt, M. (2016). Remote tower experimental system with augmented vision videopanorama. In N. Fürstenau (Ed.), *Virtual and remote control tower, research topics in aerospace* (pp. 163–192). Springer International Publishing.

Girshick, R. (2015). Fast R-CNN. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 1440–1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition.*

Jia, Y., & Shelhamer, E., et.al. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM International Conference on Multimedia.* arXiv:1408.5093 [cs.CV], https://arxiv.org/pdf/1408.5093.pdf.

Kakubari, Y., Koga, T., & Miyazaki, H., et al. (2014). Enhancement of passive surveillance system for airport surface movement. *Electronics and Communications in Japan, 97*(3), 24–30.

Kang, Z., Bass, E. J., & Lee, D. W. (2014). Air traffic controllers' visual scanning, aircraft selection, and comparison strategies in support of conflict detection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 77–81.

Kirkpatrick, S., Gelatt. C. D. Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*(4598), 671–680.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems* (Vol. 1, pp. 1097–1105).

Lu, X., Izumi, T., Teng, L., & Horie, T. (2012). A novel background subtraction method for moving vehicle detection. *IEEJ Transactions on Fundamentals and Materials, 132*(10), 857–863.

Mandellos, N. A., Keramitsoglou, I., & Kiranoudis, C. T. (2011). A background subtraction XE "background subtraction" algorithm for detecting and tracking vehicles. *Expert Systems with Applications, 38*, 1619–1631.

Matsuyama, T., Wada, T., Habe, H., & Tanahashi, K. (2006). Background subtraction XE "Background subtraction" under varying illumination. *Systems and Computers in Japan, 37*(4), 77–88.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7*, 308–313.

Okada, R., Shirai, Y., & Miura, J. (1996). Object tracking based on optical flow and depth. In *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems.*

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp.779–788).

Remotely Operated Aerodrome Flight Information Service. (2012). AN-Conf/12-WP/130 23/10/12. In *ICAO Working Paper, TWELFTH AIR NAVIGATION CONFERENCE Montréal.*

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1137–1149.

Redmon, J., & Farhadi, A. (2018). *Yolov3: An incremental improvement*. arXiv:1804.02767 [cs.CV], https://arxiv.org/pdf/1804.02767.pdf.

Schmidt, M., Rudolph, M., & Fürstenau, N. (2016). Remote tower prototype system and automation perspectives. In N. Fürstenau (Ed.), *Virtual and remote control tower, research topics in aerospace* (pp. 193–220). Springer International Publishing.

Scott, J., Pusateri, M. A., & Cornish, D. (2009). Kalman filter based video background estimation. In *Applied Imagery Pattern Recognition Workshop (AIPRW)*. IEEE.

Shrivakshan, G. T., & Chandrasekar, C. (2012). A Comparison of various edge detection techniques used in image processing. *International Journal of Computer Science Issues, 9*(5), 269–276.

Smith, W. E., Vakil, N., & Maislin, S. A. (1992). Correction of distortion in endoscope images. *IEEE Transactions on Medical Imaging, 11*, 117–122.

Szeliski, R. (1996). Video mosaics for virtual environments. *IEEE Computer Graphics and Application, 16*(2), 22–30.

Van der Jeught, S., Buytaert, J., & Dirckx, J. J. J. (2012). Real-time geometric lens distortion correction using a graphics processing unit. *Optical Engineering, 51*(2), 027002.

van Schaik, F. J., Roessingh, J. J. M., Bengtsson, J., Lindqvist, G., & Fält, K. (2016). The advanced remote tower system and its validation. In N. Fürstenau (Ed.), *Virtual and remote control tower, research topics in aerospace* (pp. 263–278). Springer International Publishing.

Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation

Maik Friedrich

Abstract This chapter describes the metrics for the validation of a Remote Tower Control workplace. The study shows how Air Traffic Control Officers (ATCOs) observe traffic from a Tower Control Working Position at Airport Erfurt-Weimar in comparison to a Remote Controller Working Position. Shadow-mode trials were used to cover perceptual, operational, and human factors aspects of a Remote Tower System, including a live video panorama and a research aircraft. The aircraft was used to fly different maneuvers within the aerodrome. These maneuvers allow insights on the detectability of an aircraft within different distances from the tower and the gathering of operation information about aircraft status. In addition, a vehicle was used to position static objects on the airfield to determine the detectability of these objects for different distances to the Control Tower (RTO-camera system). Eight ATCOs from the DFS participated in the validation exercise. Time-synchronized questionnaires for the controller working position remote (CWP-remote) and the controller working position tower (CWP-tower) were applied, addressing operationally relevant questions to the ATCOs. The validation exercise targets the evaluation of metrics that could help standardize the process of testing Remote Controller Working Positions. The results consider expense of realization, comparability and feasibility as major classifications for the used metrics. Further an approach for combining the classification into one score is presented, to rank the metrics in relation to each other.

1 Introduction

Future remote control of low traffic airports (Remote Tower Operation, RTO) will rely on the replacement of the conventional Air Traffic Control (ATC) workplace (CWP-tower) by a remote controller working position (CWP-remote). For short- and midterm realization of a CWP-remote the out the window (OTW) view will be a digitally reconstructed panoramic view using high resolution video cameras. The

M. Friedrich (✉)
German Aerospace Center, Institute of Flight Guidance, D-38104 Braunschweig, Germany
e-mail: Maik.Friedrich@dlr.de

DLR-internal project RapTOr (Remote Airport Tower Operation Research, 2005–2007) focused on remote tower control of single airports, while the project RAiCe (Remote Airport traffic control center, 2008–2012) focused on the RTO-prototype development and the idea to control multiple small airports from one remote center (Fürstenau et al., 2008a, b, 2009; Möhlenbrink et al., 2009, 2012; Schmidt et al., 2007).

In parallel to these projects, remote tower operation was pushed forward by a joint venture project of the Swedish Civil Aviation Administration (LFV) and SAAB, called ROT (Remotely Operated Towers, 2006–2008) (Saab Security, 2008). SAAB also coordinated the EU-Project ART (Advanced Remote Tower, 2007–2009) (van Schaik et al., 2010) focusing on single remote tower control (see chapter "The Advanced Remote Tower System and Its Validation"). Further, the German Aviation Research Program iPort funded the ViCTOR project (Virtual Control Tower Research Studies, 2009–2012), which was led by DFS and addressed new concepts of remote operation, team work, as well as visualization aspects.

From an American perspective there is a strong motivation to work out operational and functional requirements (Ellis & Liston, 2010), technical/system requirements and the integration of concepts (Hannon et al., 2008), to ensure the safety when applying RTO. In the US concepts on staffed NextGen Tower also explore alternative surveillance systems for the OTW (Friedman-Berg, 2012) (see also chapter "Remote Tower Research in the United States". The same perspective applies for Europe, especially within the Single European Sky ATM Research Program (SESAR). There, remote tower is addressed under a separate Operational Focus Area (OFA, 06.03.01) (Committee Sesar Program, 2010). This OFA comprises the different Remote Tower Activities assigned in the Operational Projects.

To test the feasibility of the RTO concept, human-in-the-loop studies have been completed addressing research questions for single remote tower (European Organisation for the Safety of Air Navigation, 2010). To complete the analysis of feasibility, research prototypes are tested also within field trials. In 2007, field trials with the first experimental RTO system, consisting of four cameras for reconstructing a panoramic view were completed at Braunschweig Airport (see chapter "Remote Tower Experimental System with Augmented Vision Videopanorama" and e.g. Schmidt et al. (2007)). The data of the field trials have been used to quantify the effective resolution of that video panorama (Fürstenau et al., 2009). Within the ART-Project, van Schaik et al. (2010) assessed the importance of visual cues for remote tower operations and suggested a formula for calculating the required resolution for either detection or recognition of each cue. While we agree that a definition of minimum resolution requirements for RTO is one important issue, it remains unclear whether the calculated minimum resolution requirement can be empirically validated by Air Traffic Control Officers' (ATCO) detection and recognition rates of such items under daylight and good visibility conditions. The problem of visual resolution and the existence of different prototypes developed by different institutions and companies leads the authors to believe that a structured validation concept is needed to enable quantitative comparison.

Considering the different CWA-remote projects and their prototypes we need metrics that measure the discrepancy between out the window (OTW) view and RTO. This way the different technical solutions could be validated with the same remote tower metrics (RTMs) and made comparable to the ATCOs. This chapter presents a set of RTMs that were evaluated in a validation exercise (Friedrich & Möhlenbrink, 2013), which was completed under the scope of the OFA Remote Tower. Two Remote Tower validation exercises under this scope were already completed in Sweden (Mullan et al., 2012a, b). All three validation exercises contribute to the transition from feasibility to pre-industrial development and integration. Therefore, the remote tower operations concept descriptions and the functional/operational requirements have been defined in the Operational Service and Environment Description (OSED) for Remote Provision of Air Traffic Services to Aerodromes (Mullan et al., 2012a, b). The functional/operational requirements define what the user (here: ATCO) of the system wants the system to do. It is important to note that the functional requirements are independent from the technical solution. Complementary to the functional specification, technical system requirements define whether a specific technical system can provide specific information to the user.

Within this chapter RTMs for a CWP-remote validation are presented and combined with the results from the third validation exercise. This helps not only to evaluate the CWP-remote itself but also identify RTMs that are essential for a validation. We used a prototype developed by DFS and DLR in 2012 and explained in detail below.

First, an extended schematic will be introduced to improve the metrics evaluation process. Second, the list of RTMs for the CWP-remote is presented. Third, the method section covers the experimental setup. Fourth, the results are presented. Fifth, the contribution provided by the RTMs will be discussed. In addition the methods for validating a remote tower system will be discussed. Sixth, data and methods are summarized as appropriate to judge which RTMs cover the important parts of the ATCOs work.

2 Extended Field Trial Infrastructure

The goal of this chapter is to elaborate and discuss RTMs for a single RTO concept. Previous validation exercises concerning remote towers mainly focused on analyses of subjective data such as questionnaires, interviews, observations and ATCOs' feedback (Ellis et al., 2011; Mullan et al., 2012a, b). However, an extended infrastructure for field trials with RTMs can provide additional objective data to support the development and consolidation of specifications for future RTO.

Figure 1 shows the field trial schematics extended with the RTMs that are developed for CWP-remote. This schematic description shows the stepwise validation within the project. For (1) Providing Air Traffic Service (ATS) (from tower or remote) the (2) functional requirements have been elaborated (Mullan et al., 2012a, b). For setting up field trials, a validation plan is written and it is defined, whether
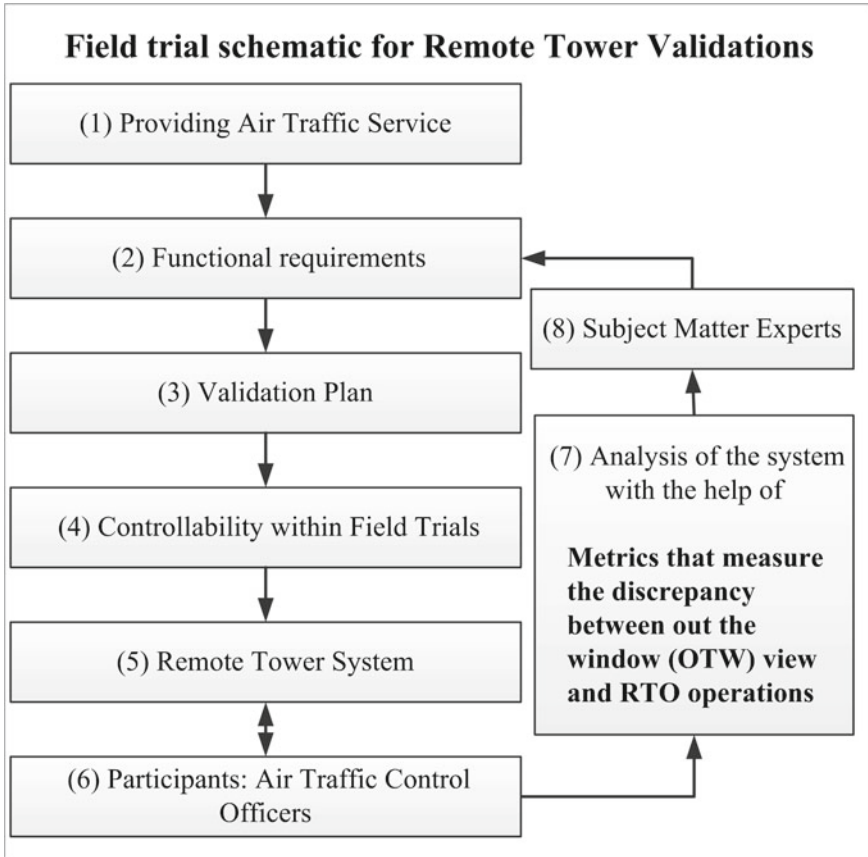
Fig. 1 Field trial schematics for development and consolidation of specifications for future remote tower operations

the (3) experimental design includes a control condition beside the experimental condition(s). The control condition is important to have a baseline or reference, to evaluate the results of the experimental condition(s). Within field trials a baseline cannot always be provided for several reasons. However, RTO allows a comparison between CWP-tower and CWP-remote. In addition, the (4) controllability within a field trial is usually limited: The amount of traffic and flight manoeuvers is not under the experimenters control and the accessibility of operational data thereby limited. Such limitation can be overcome by using a research aircraft which is under the control of the experimenter. Thereby, the experimenter can define the traffic patterns and number of iterations for certain flight manoeuvers for a systematic analysis.

Then, the (5) Remote Tower System is tested within field trials and data is collected from (6) ATCOs'. For the (7) analysis of the system, or to be more precise, for the analysis whether the system provides the functional requirements, different kind of data can be analyzed.

3   Remote Tower Metrics

The RTMs are identified by refinement and consolidation of the functional require-ments of a CWP-remote and in cooperation with ATCOs that serve as system matter experts. The following RTMs have been identified and therefore provide the base for the validation exercise introduced in this chapter. The main difference between CWP-tower and CWP-remote is the visual presentation of the OTW. Therefore the RTMs focus on the aspect of visual perception of static and dynamic objects.

The process of identifying the RTMs was performed in three steps. First, types of visual tasks were identified that the ATCOs have to perform via the OTW. These types of visual task are related to the moving aircraft within the control zone (Aircraft) and objects on the apron (Apron Objects). Second, a workshop was conducted with ATCOs to determine specific tasks for each type of visual task. Third, the tasks were transformed into RTMs to allow performance measurement. The eight RTMs were separated into types of visual tasks as followed:

- Aircraft (5 RTMs)
- Objects at the Apron (3 RTMs)

To be consistent with the infrastructure proclaimed in Fig. 1 the RTMs were all used within one validation exercise. For more details on the influence that the RTMs have on the validation objectives and success criteria see (Friedrich & Möhlenbrink, 2013). For the purpose of this chapter we will not concentrate on the results of the

Table 1   The RTMs divided into types of visual tasks

Types of tasks	Remote tower metric	Description
Aircraft	Dutch roll	The ATCO has to identify if the aircraft performs a Dutch roll
	Route	The ATCO has to identify if the aircraft follows a specific route
	Decline	The ATCO has to identify if the aircraft performers a decline maneuver
	Landing light	The ATCO has to identify the status of the aircraft landing lights
	Flight path	The ATCO has to identify if the aircraft is on or above the flight path
	Gear status	The ATCO has to identify the status of the landing gear
Apron objects	Static objects	The ATCO has to identify specified signs in different distances to the tower
	Runway status lights	The ATCO has to identify the status of runway status lights
	Taxi ways/holding points	The ATCO has to identify the status of taxi ways and holding points

validation exercise itself other than to evaluate the RTMs for further use. Table 1 contains a list of all RTMs and a description.

The Aircraft tasks were discussed and specified with pilots to ensure their feasibility. The Apron Objects types of visual tasks were defined without any help of system matter experts.

4 Method

4.1 Participants

Eight Air Traffic Control Officers (ATCOs) employed by the DFS, participated in the validation exercise. The average participants' age was 30 years with a SD of 11.5. The average work experience was 10 years with a SD of 9. All participants worked at local or regional sized airports. 50% of the participants claimed that they had known the project in advance of the validation exercise. The participants received no additional payment and participated during typical working hours.

4.2 Apparatus

The experimental setup to analyze the RTMs consists of the technical setup (CWP-remote) and the experimental vehicles (car and aircraft) used for the validation exercise. The technical setup presents an overview of the CWP-remote and available information systems. The most important change to the CWP-tower is the visual reproduction of the out the window (OTW) view (Fürstenau et al., 2008a, b, 2009; Schmidt et al., 2007). A camera platform with five HD cameras (1920×1200 pixel) in separate temperature controlled housings, each equipped with a $2/3''$—CCD sensor and f = 8 mm lens was used (see chapters "Remote Tower Experimental System with Augmented Vision Videopanorama" and "Remote Tower Prototype System and Automation Perspectives"). The visual resolution of the sensor can be approximated by using the fundamental relationship.

$$\frac{G}{B} = \left(\frac{g}{f} - 1 \right) \approx \frac{g}{f} \tag{1}$$

with f = focal length = 8 mm, g = object distance, G = object size, B = image size, and CCD pixel size of p = 5.5 μm. This leads to a vertical object size by g = 1000 m distance corresponding to 1 Pixel is

$$\frac{G}{B} \approx \frac{1000 \text{ m}}{0.008 \text{ m}} \Rightarrow \frac{0.68 \text{ m}}{5.5 \ \mu m} \Rightarrow \frac{0.68 \text{ m}}{1 \text{ Pixel}} \tag{2}$$

vertical, or ca. 2 arcmin angular resolution. This approximate value is valid under ideal illumination (i.e. contrast) conditions (see chapter "Remote Tower Experimental System with Augmented Vision Videopanorama" and Appendix A).

In addition to the panorama-camera system, a pan-tilt zoom (PTZ) camera was mounted on the top, to allow a detailed look into participant guided areas. The PTZ camera was moveable within the full 360° viewing range and had 12 pre-sets in the range $1 \leq Z \leq 23$ (fixed positions and zoom values) for fast responses. The optical specifications of the PTZ camera is approximated by

$$\alpha_Z = \frac{p_H}{Z f_0} \tag{3}$$

yielding for a zoom setting of e.g. $Z = 4$ an ideal pixel resolution $\alpha_Z = 1$ arcmin (with $f_0 = 3.6$ mm, $p_H = 4.4$ mm, viewing angle $2\theta = 15°$, see also chapters "Remote Tower Experimental System with Augmented Vision Videopanorama" and "Remote Tower Prototype System and Automation Perspectives" and Appendix A). The PTZ control and video stream was presented via a separate monitor within the CWP-remote (Fig. 2). That is why it was expected that due to the limited resolution of the panorama controllers in the RTO-CWP would make more use of the PTZ than

Fig. 2 CWP-remote at the airport Erfurt

controllers in the Tower-CWP make use of the binoculars for supporting decision making.

The visual reproduction from the five cameras, situated on top of the Erfurt-Weimar tower was displayed on five 40″ LCD monitors arranged in a "broken circle" around the CWP-remote (Fig. 2), providing a 200-degree field of view.

A microbus (VW bus T4) and an the DLR aircraft (Dornier Do 228–101 twin turboprop engine test aircraft; length 15.03 m, body height × width 1.8 × 1.6 m, wing span 16.97 m, wheel diameter 0.65 m) were used as research vehicles to perform the Aircraft type of tasks. The bus was used to position static objects in predefined distances (250, 500 and 1000 m) to perform the Apron Objects tasks. The static objects had a diameter of 0.6 m and could be a circle or a cross mounted in the center of a square signage with an edge length of 0.7 m.

The participants were placed about 1.8 m from the monitors. Besides the panorama as reproduction of the OTW view the participants were provided with the following additional sources of information:

- Videopanorama
- PTZ Camera (Controlled via pen-input)
- Air Situation Display
- Flight Plan Data
- Weather Information System.

The RTMs were measured by synchronized questioning of two ATCOs working either CWP-tower or CWP-remote. This increases the RTMs significance for the purpose of comparing both workplaces. The survey software "Controlsurvey" was used to question the participants during the trials. Controlsurvey was developed by the DLR for the purpose of synchronized questioning and with the flexibility of reacting to minor deviations from planned scenarios.

Table 2 Connection between the remote tower metrics and their implementation as a question

Remote tower metric	Traffic pattern	Questions to capture the metrics	
Dutch roll	A	"Did the aircraft wag its wings?"	
Route	BC, and EF	"When is the aircraft turning?"	
Decline	D	"Did the aircraft decline?"	
Landing light	G	"Are the landing lights off?"	
Flight path	$G_{1,2,3}$	"Is the aircraft on the flight path?"	
Gear status	H1–H3	"Is the landing gear pulled up?"	
Static objects		"Which symbol can you see next to the car?"	
Runway status lights		"Are the runway status lights on?"	
Taxi ways/holding points		"Which holding point are you not able to see?"	

The RTMs transform into questions that were used during the validation exercise. Table 2 shows the implementation of RTMs in questions. Each RTM connected to the Aircraft tasks is connected to a point or position within the traffic pattern (Fig. 4).

4.3   Design

The validation exercise was completed as a passive shadow-mode close loop field trial. The experimental design is based on the direct comparison between the CWP-tower and CWP-remote. The workplace of the participating ATCO within a trial is the independent variable that is measured with the RTMs. Through the comparison of both workplaces and the synchronized questioning (Fig. 3), the effect of the confounding variables: unforeseen traffic events, meteorological conditions and time of day were reduced.
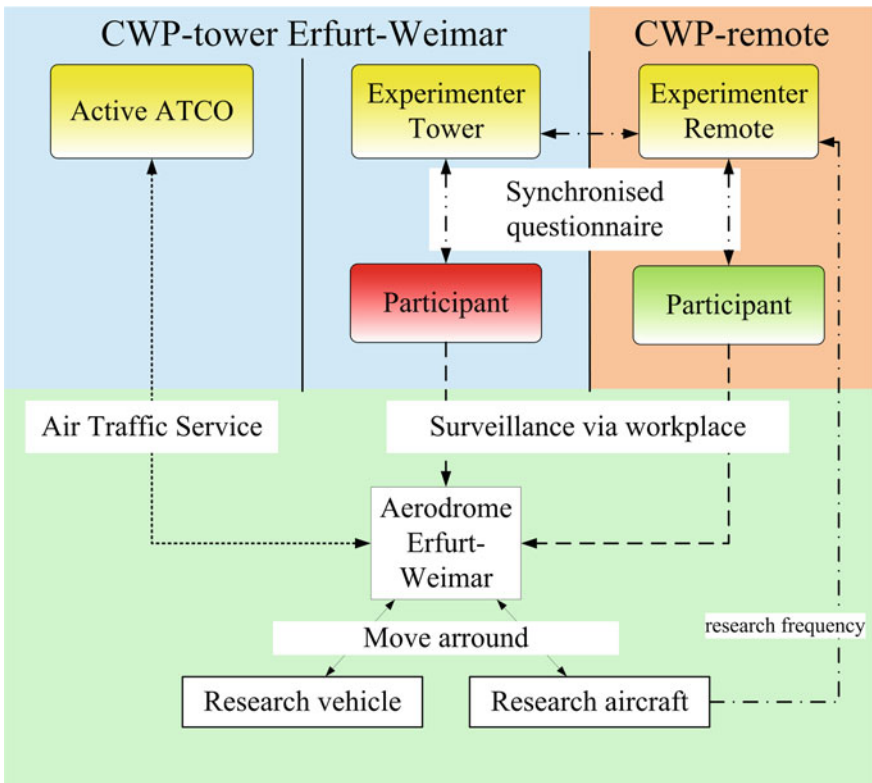
Fig. 3 Experimental procedure for comparing CWP-tower and CWP-remote

Fig. 4 Traffic pattern of the reseach aircraft within the aerodrome Erfurt-Weimar
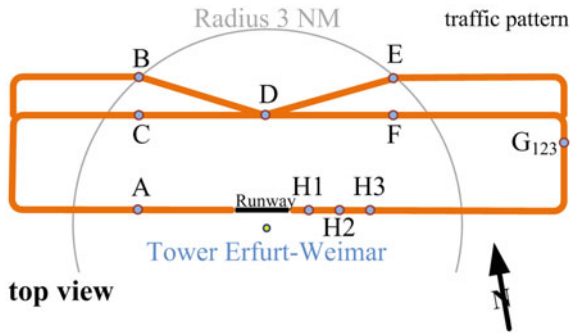
The research aircraft flew predefined scenarios within the aerodrome to create authentic monitoring situations. Two mirrored scenarios were defined and switched between the runs. Each scenario was varied by the order of events that the aircraft should perform while flying the traffic pattern (Fig. 4) 14 times. The aircraft was also equipped with an additional radio to communicate with the experimenter remotely via a research frequency (Fig. 3), to account for unforeseen situations. Besides the research aircraft traffic, additional unplanned traffic could arrive throughout each validation run. This allowed a mixture of scheduled and unscheduled traffic and increases the external validity of the validation exercise.

The design enables the RTMs (Table 1) to work as objective performance measurements in the terms of comparing two workplaces against each other. Since the CWP-tower is state of the art this comparison is necessary to judge the influence of CWP-remote on visibility and safety. The RTMs questions can be analyzed for correctness of the given answer and response times. The answers to the RTMs are always unambiguously, whereas the response times depend on the current attention of the participant.

An additional question concerning the used sources of information was used, to evaluate the RTMs validity for comparing the two workplaces. The used sources of information were subdivided into the panorama (OTW view or Video panorama), the magnification (binoculars or PTZ-camera),and the air situational display (Radar), weather information system (WIS). The participants were instructed to name only the system that they used to make their final decision. This means e.g. if they used the video panorama to position the PTZ-camera and then used the PTZ video for their answer the used source of information was the PTZ-camera.

The feasibility of the RTMs were covered by a debriefing questionnaire. The debriefing questionnaire used a 6-point Likert Scale (1 = totally disagree; 6 = totally agree; average of 3.5) to judge each RTM for its feasibility. For each RTM one question was formulated in the following style: "Did you find the questions concerning the Dutch Roll feasible?".

4.4 Procedure

The participants were randomly divided into four groups (two per group). The validation exercise took place from 17th of July until the 20th of July 2012. Every day a different group took part in the exercise. Each group had to complete two trials. For the first trial it was randomly decided which participant worked at the CWP-tower and CWP-remote. Within the second trial the group members always switched workplaces. Besides the two participants an active ATCO was needed for every validation run to ensure the provision of ATS (Fig. 3). This was necessary because air traffic safety regulations did not allow active control by the participants of any traffic within the aerodrome.

Within the validation exercise, the procedures for every day were equal. A briefing of the new group was performed and they were instructed about the project and the validation exercise. That was followed by assigning the ATCOs to the different workplaces (Fig. 3). Afterwards a 30 min PTZ camera training was conducted. Then the first validation run was performed with duration of 140 min. After that the participants switched workplaces and the second validation run was completed. At the end, a 60 min debriefing with a debriefing questionnaire was performed with both participants.

Every validation run started with the research aircraft's first movement away from its apron parking position. The aircraft followed a predefined scenario, while the participants on both workplaces had to answer the same questions addressing the different RTMs (Table 2). All questions, regardless of the type, occurred synchronized to generate two comparable sets of answers that differ only in the used workplace. Every question was placed in a dialog between the participant and the particular experimenter. The experimenters read the questions to the participants. The participants used their workplace to collect the answer. Then they replied the collected answer as fast as possible to the experimenter and added their used source of information. The answers from both CWP were combined into question pairs. Question pairs were generated if both participants answered. In addition to this conservative analysis Fürstenau et al. (2013) performed a different analysis using signal detection theory and time-pressure theory (Fürstenau et al., 2014), and included also the answers that were not provided (non-answers) as false answers (see chapter "Model based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position"). The questions concerning the aircraft manoeuvers were asked at predefined points within a standardized traffic pattern (Fig. 4, A to H).

5 Results

This section is divided into two parts. In the first part we show the basic analysis method applied on safety related metrics to give an example of the RTM potential. The Decline, Landing Lights and Gear Status are the most safety related and will
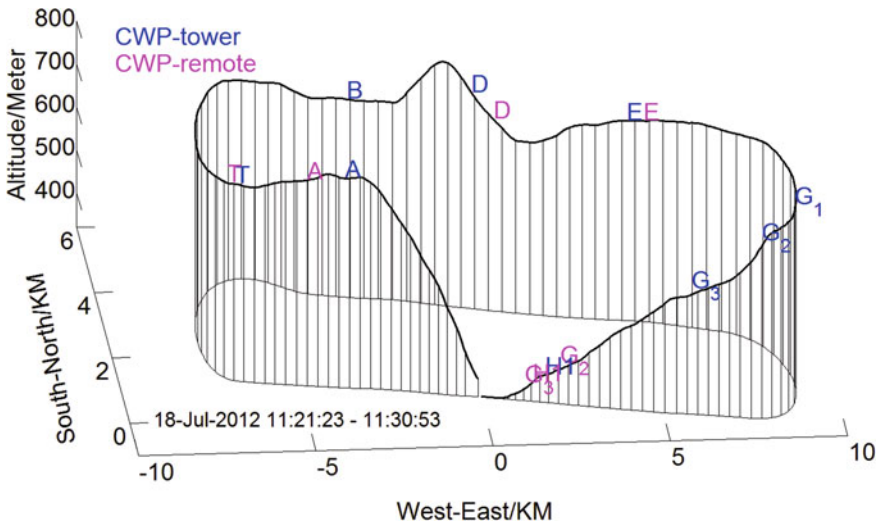
Fig. 5 One traffic circle performed by the research aircraft with Erfurt-Weimar tower as origin

therefore be presented in detail. The second part of the section contains a rating for the proposed RTMs in terms of expense of realization, comparability and feasibility. As mentioned above, this chapter does not focus on the results of the validation but on the RTMs. A complete list of all results from the SESAR- JU D36 Project can be found in (Friedrich et al., 2012).

Throughout the validation the RTM lead to a total number of 1326 question pairs (CWP-tower and CWP-remote). 936 Aircraft RTM questions pairs lead to an average of 117 per trial and an average of 12 completed traffic patterns per run. Figure 5 shows one traffic circle with the corresponding answer times from both workplaces. The letters in Fig. 5 are similar to those in Fig. 4. They do not show the position of planned maneuver but the position of the research aircraft when the participant answered the question related to the maneuver. Therefore every letter comes in pairs. Except for B were no answer was given on the CWP-remote.

5.1 Basic Analysis of Safety Related Metrics

Decline, Landing Lights, and Gear Status are the most safety related types of visual tasks. For each of the three RTMs an average of 12 question pairs per trial were collected. These values were used as a direct comparison between the CWP-tower and CWP-remote. The results for correct answers are presented in Table 3.

As Table 3 shows, the participants' answers given concerning safety related RTMs are not significant degraded for Decline and Gear Status and are significantly degraded for Landing Lights. Comments from both workplaces indicated that the

Table 3 Answers for the safety related manoeuvers (% correct answers (standard deviation) from provided answers)

RTM	Mean correct answers (SD) CWP-tower	Mean correct answers (SD) CWP-remote	Significant difference (F-test)
Decline	86.1% (34.9)	82.4% (38.3)	$F(1, 7) = 1.62$, n.s
Landing lights	83.33% (37.0)	44.3% (49.37)	$F(1, 7) = 40.45$, $p < 0.05*$
Gear status	94.32% (23.2)	94.52% (22.2)	$F(1,7) = 0.96$, n.s

position of the landing lights at the research aircraft wasn't easy to identify. This comments are used in the second part of result section to judge the feasibility of the RTM Landing Lights.

The design of the synchronized questioning allows not only for an analysis of the paired answers but also the reaction times. The reaction time within the validation depended on the time an experimenter needed to read the question out loud and the participant to answer it. Due to reading training before the validation the influence of the experimenters were reduced to a minimum. The reaction times therefore are mainly influenced by the performance of the participants. As a second basic analysis the reaction times for Gear Status were analyses for three distances (H_1: H_2: 0.5 NM, H_3: 1.0 NM, 1.5 NM) to the tower when the questions were asked. Figure 6 shows the results separated by workplace. The analysis allows a detailed view on the reaction times and the influence the different workplaces have on them. This analysis is possible for all RTMs, if synchronized capturing is used.

By looking at Landing Lights we find a metric that shows a decrease in performance. For Landing Light the correct answers given are significantly lower for CWP-remote than for CWP-tower.[1] The response times for the Gear Status for CWP-remote are also higher at a distance of 0.5 than on the CWP-tower. Because safety is always to be the first priority this results lead to a bad grading of the CWP-remote. However, Decline, Landing Lights, and Gear Status are not equal RTMs which will be presented in the next section.

5.2 Evaluation for the RTMs

The evaluation of the RTMs is based on the expense of realization (ER), the comparability (C) and feasibility (F). In this section ER, C, and F are explained and their connection to the validation exercise are presented. This leads to a comprehensive view of the RTMs connected to the CWP-remote. The evaluation of the RTMs results in an overall ranking by combining ER, C, and F into a single score.

---

[1] However, due to the problematic interpretation of the Landing Lights RTM the %-correct analysis in Table 3 shows now significant difference altogether. An extended analysis is discussed in Chapter "Multiple Remote Tower Simulation Environment (S. Schier)".
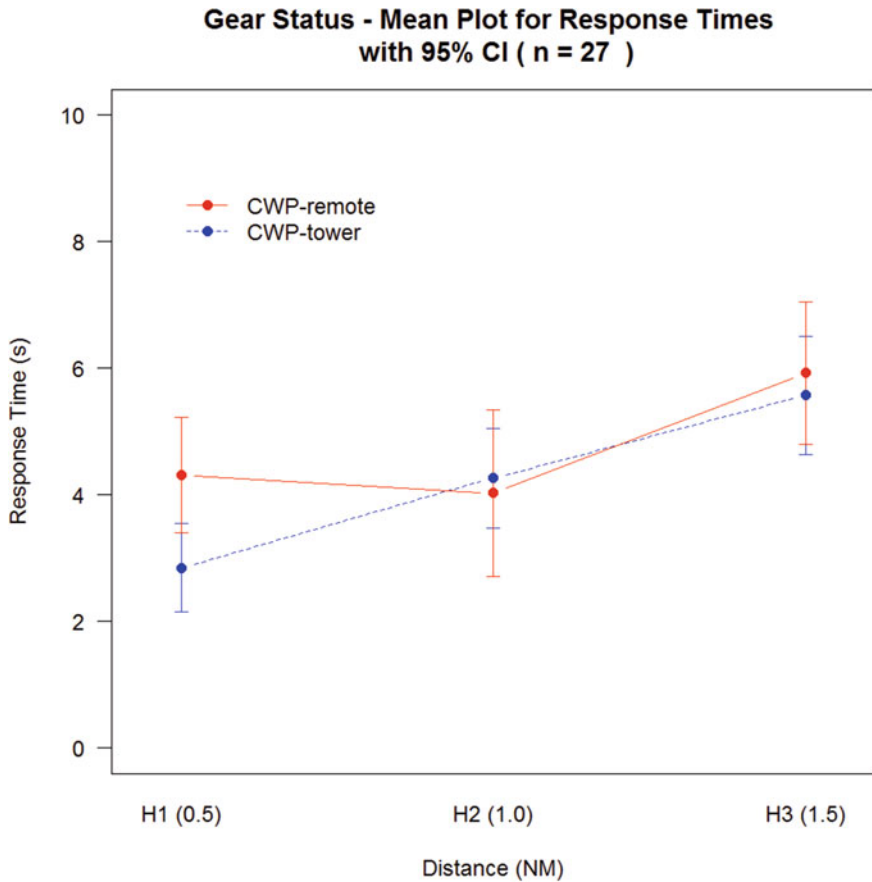
Fig. 6 Gear status—mean plot for response times with 95% confidence intervall (N = 27)

ER evaluates the procedure to set up the validation environment and enable situations for the RTMs to be tested. Some RTMs even require an aircraft that follows a defined scenario. ER classifies the RTMs into low, medium or high expenses to realize the situation. Low means that no special equipment is needed throughout the validation. Medium means that special equipment is needed but the cost are beneath 1000€ per day. High means that special equipment is needed and the cost are above 1000€ per day. The ER results are summarized in the level of types of visual task. Table 4 presents the types of visual tasks with a detailed explanation on the ER results.

Since the objective data arises from the direct comparison of CWP-tower and CWP-remote the C of the RTMs has to be evaluated. C in this case can be defined as the amount of questions answered with the same sourced of information. C is high if the participants use the same source of information and therefore the difference between the two CWP is under test. If C is low, it indicates switching to source

Table 4 Expense of realization (ER) results for the RTMs, separated by type of visual tasks

Types of visual tasks	Expense of realization (ER)	Detailed explanation
Aircraft	High	The Aircraft type of task need an aircraft that follows a scenario. Therefor the cost per day are above € 1000
Apron objects	Medium	For static objects the ER is medium, because due to the two possible answers (cross or circle) Static objects metric requires that the objects are disclosed until shortly before questioning. The consequence is that staff and a research vehicle are needed
	Low	Low is true for runway status lights and taxi ways/holding points because the RTMs do not need any special equipment or staff

of information and therefore a direct comparison between CWP-remote and CWP-tower is less significant. The correlation between the used sources of information bases on the amount of usage per source. We also identify a major switching tendency if more than 25% usages switched category from CWP-tower to CWP-remote. Table 5 shows the results of the analysis sorted from high to low.

F presents the feasibility of each RTMs during the validation. The rating of F depends on the debriefing questionnaire (6-point Likert Scale, (1 = totally not feasible; 6 = totally feasible)) and comments of the experiments concerning the feasibility during the validation trails. Table 6 shows the results from the debriefing questionnaire. All RTMs are above the scale average of 3.5, except the Landing Light.

The final step to summarize the evaluation is to rate the RTMs depending on their ER, C, and F results. Because ER, C, and F have different dimensions we decided to standardize them. Because ER, C, and F are equally important to judge a metric we

Table 5 Comparability (C) results for the RTMs, ranked from high to low

Remote tower metric	C	Major switching tendency
Taxi ways/holding points	0.991	None
Landing light	0.878	From OTW to magnification
Route	0.830	From OTW to Radar
Flight path	0.710	From OTW to Radar
Gear status	0.708	From OTW to magnification
Runway status lights	0.685	None
Dutch roll	0.643	From OTW to magnification
Static objects	0.616	From OTW to magnification
Decline	0.155	From OTW to Radar

Table 6 Results from the debriefing questionnaire and the experimenter rating concerning F

Remote tower metric	Average F (N = 8)	Comments by the experimenter
Dutch roll	5.63 (sd = 0.52)	
Route	5.50 (sd = 0.53)	The aircraft was not always visible in the panorama due to the distance to the camera system
Decline	5.63 (sd = 0.52)	
Landing light	2.88 (sd = 0.83)	The position of the aircraft landing light strongly dependent on the type of aircraft. The position of the landing lights has strong influence on detectability
Flight path	5.38 (sd = 0.52)	
Gear status	4.00 (sd = 0.53)	The focusing of the PTZ-camera was too long
Static objects	5.25 (sd = 0.71)	
Runway status lights	4.38 (sd = 0.74)	The runway status lights had to be always on and therefore variation in the condition was impossible
Taxi ways/holding points	5.25 (sd = 0.71)	
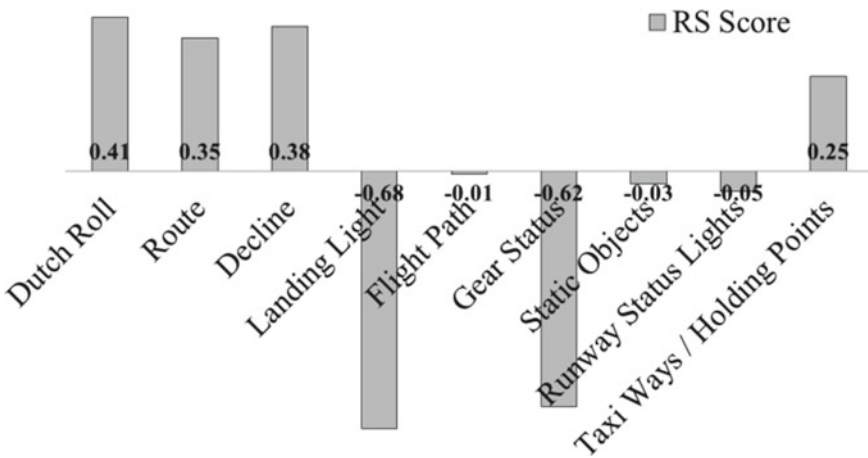
Fig. 7 The ranking of all RTMs in relation to their RS score

decided merging the standard scores using an equal distribution equation

$$RS = \frac{1}{3}Z_{ER} + \frac{1}{3}Z_C + \frac{1}{3}Z_F \qquad (4)$$

The result of the merging is defined as RTMs Score (RS). Figure 7 shows all R
TMs and their RS. The results show that each RTM has a different RS score and that
there are differences in the quality of a metric.

6 Discussion and Conclusion

In the result section, two different approaches for analyzing the RTMs were presented. The results show that the RTMs are able to distinguish the ATCO s' performances data depending on the used workplace. They allow capturing of the objective data in relation to the workplace, which allows insight on how the ATCOs perceived the different working environments. It has to be discussed in what way the RTMs can be a significant help for analyzing the CWP-remote and how the RS can support this task.

6.1 Basic Analysis of Safety Related Metrics

The basic analysis of safety related metrics gives an example of the opportunities provided by the correct measurement for performance and it also shows the difficulties. The metrics allow insights on the performance of the ATCOs by defining tasks that are necessary for ATC. The results are not only dependent on the performance of the participants but also on the selection of ATC task that were chosen to define the metrics. This is especially important for safety related issues.

The results for Decline, Landing Light, and Gear Status show that there are significant differences between CWP-tower and CWP-remote. The significant differences are in correct answers (Landing Lights) and also in reaction times (Gear Status). The topic of reaction times is specifically addressed in the following chapter "Model based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position" within the Time Pressure Theory based data analysis (Fürstenau et al., 2014). Initial interpretation of the results could suggest that the CWP-remote is not as safe as the CWP-tower. As mentioned above, this chapter does not focus on the implications for remote tower operations, but on the RTMs to measure the difference between two workplaces. For completion of the interpretation of our results we need to evaluate the RTMs.

6.2 Evaluation for the RTMs

As mentioned in the previous section, sometimes different metrics connected to the same domain contradict each other. This leads to inaccurate results because an aggregation of different metrics is almost impossible. Therefore, the evaluation of the RTMs is important to increase and order the validity of complex RTO studies. Only then it becomes possible to measure the differences that exist between the CWP-tower and CWP-remote. This does not only apply for the CWP-remote presented in this chapter but should apply for all CWP-remote systems that will be designed and tested in the future.

The ER classification of metrics shows a connection to the validation budged, but also to the validity of the validation. Of course the cost for a research aircraft is immense, but the task of an ATCO officer is to control flying aircrafts and therefore an aircraft and metrics to capture the performances is needed. The ER for the Aircraft visual tasks could be reduced to "low", if for example the usual traffic on the remotely controlled airport is used. This would lead to a series of adaptation to the metrics and would reduce the between-subject comparability of the results.

The C classification could be interpreted in two directions. The first is to analyze the change in used source of information as an indicator for the major difference of the two CWP. For example, the data of the compatibility tests revealed that the participants moved away from the panorama to the PTZ-camera or the radar to gather their information. The second is the demand to compare the RTMs results from each workspace on an equal basis. Therefore the similarity between used sources of information was analyzed. Stronger variations in the used sources of information are considered as less comparable because the performances do not base on the same systems.

More than ER and C, F quantifies how well the RTM fits into the procedure of the validation. This is shown by the subjective rating of the ATCOs but also by the comments of the experimenters. We consider the F classification also as a learning indictor for further changes to the RTM before finalization.

The RS scores show that there are differences between the RTMs in quality and validity. This leads to the proposal of standard metrics that need to be defined for the evaluation of remote tower operations. An individual definition of metrics is misleading, because the system developer might have a narrow perspective on their prototype. The authors propose to use the RTMs presented in this chapter for testing them with different prototypes and determine the agreement of the quantitative results.

7 Outlook

In line with the results of the accomplished validation exercises under the operational focus area "Remote Tower", the evaluation of RTMs within this validation exercise provide an additional step for the remote tower concept validation, based on a live-video panorama. The chapter focused on the RTMs rather than the results of the validation exercise itself. Those are reported in separate publications (Friedrich & Möhlenbrink, 2013; Fürstenau et al., 2013, 2014) and chapter "Model based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position". The validation also shows that metrics can be judged differently depending on their quality to distinguish between different systems. After addressing the feasibility of the concept within this exercise, validation activities centre on system integration, for which the consolidation of the operational concept and the prototype system is the main goal.

References

Committee Sesar Program. (2010). *Overview and Release 1 Plan v1.0*. Brussel. https://extranet.ses arju.eu

Ellis, S. R. & Liston, D. B. (2010). Visual features involving motion seen from airport control towers. In *11th IFAC/IFIP/IFORS/IEA*.

Ellis, S. R., Fürstenau, N., & Mittendorf, M. (2011). Determination of frame rate requirements for video-panorama-based virtual towers using visual discrimination of deceleration during simulated aircraft landing: Alternative analysis. In *Tagungsband 9. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 519–524). VDI-Verlag.

European Organisation for the Safety of Air Navigation. (2010). *E-OCVM v3.* Vol. II. http://www.eurocontrol.int/valfor/gallery/content/public/docs/E-OCVM3 Vol. I WebRelease.pdf

Friedman-Berg, F. (2012). *Staffed Next GenTower Human-in-the-Loop Camera Integration Evaluation.*

Friedrich, M., & Möhlenbrink, C. (2013). Which data provide the best insight? A field trial for validating a remote tower operation concept. In *Tenth USA/Europe Air Traffic Management Research and Development Seminar* (p. 10).

Friedrich, M., Möhlenbrink, C., & Carstengerdes, N. (2012). *SESAR-JU D86 Project 06.08.04—Single Remote TWR Ph1 V2—Validation Report* (00.01.00 ed.).

Fürstenau, N., Schmidt, M., Halle, W., & Rudolph, M. (2008a). *Internal Report—RAiCe Projektplan, IB 112–2008/09.*

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., & Halle, W. (2008b). Augmented vision videopanorama system for remote airport tower operation. In I. Grant (Ed.), *Proceedings of the ICAS 2008, 26th International Congress of the Aeronautical Sciences.*

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Papenfuss, A., & Kaltenhäuser, S. (2009). Steps towards the virtual tower: remote airport traffic control center (RAiCe). In *ENRI Int. Workshop* (Vol. 1, p. 14).

Fürstenau, N., Friedrich, M., Mittendorf, M., Schmidt, M., & Rudolph, M. (2013). Discriminability of flight maneuvers and risk of false decisions derived from dual choice decision errors in a videopanorama-based remote tower work position. In *Engineering Psychology and Cognitive Ergonomics* (Vol. 8020, pp. 105–114). Springer.

Fürstenau, N., Mittendorf, M., & Friedrich, M. (2014). Model based analysis of two-alternative decision errors in a videopanorama-based Remote Tower work position. In *HCII2014/EPCE-11.* Lecture Notes Computer Science LNCS/LNAI.

Hannon, D., Lee, J., Geyer, M., Mackey, S., Sheridan, T., Francis, M., et al. (2008). Feasibility evaluation of a staffed virtual tower. *Journal of the Air Traffic Control Association*, 27–39.

Möhlenbrink, C., Papenfuss, A., & Jakobi, J. (2012). The role of workload for work organisation in a remote tower control center. *Air Traffic Control Quarterly, 20*(1), 5.

Möhlenbrink, C., Friedrich, M. B., & Papenfuss, A. (2009). RemoteCenter: Eine Mikrowelt zur Analyse der mentalen Repräsentation von zwei Flughäfen während einer Lotsentätigkeitsaufgabe. In *Tagungsband 8. Berliner Werkstatt Mensch-Maschine-Systeme* (p. 6). Berlin-Brandenburgische Akademie der Wissenschaften.

Mullan, C., Lindqvist, G., & Svensson, T. (2012a). *SESAR-JU Project 06.09.03—Remote provision of ATS to a single aerodrome—Validation plan* (00.01.02 ed.). SESAR. www.sesarju.eu

Mullan, C., Lindqvist, G., Svensson, T., Abel, M., & Ankartun, P. (2012b). *OSED for remote provision of air traffic services to aerodromes, including functional specification for single & multiple aerodromes.*

Saab Security. (2008). *Remotely operated tower—The future of being present*.

Van Schaik, F. J., Lindqvist, G., & Rössingh, H. J. M. (2010). Assessment of visual cues by tower controllers. In *11th IFAC/IFIP/IFORS/IEA* (p. 19). NLR-TP-2010-592.

Schmidt, M., Rudolph, M., Werther, B., Möhlenbrink, C., & Fürstenau, N. (2007). Development of an augmented vision video panorama human-machine interface for remote airport tower operation. In M. J. Smith & G. Salvendy (Eds.), *Human interface and the management of information. interacting in information environments* (44th ed., pp. 1119–1128). Springer. https://doi.org/10.1007/978-3-540-73354-6_122

Model Based Analysis of Two-Alternative Decision Errors in a Videopanorama-Based Remote Tower Work Position

Norbert Fürstenau

Abstract   Initial analysis of Remote Control Tower (RTO) field test with a prototype videopanorama system under quasi operational conditions (Friedrich and Möhlenbrink in ATM, 2013) has shown performance deficits quantified by two-alternative aircraft maneuver discrimination tasks (Fürstenau et al. in EPCE/HCII 2013, Part II, Lecture Notes in Artificial Intelligence (LNAI), pp. 105–114, 2013). Here we present the quantitative analysis of these results using the complementary methods of Bayes inference, signal detection theory (SDT) with parametric and non-parametric discriminabilities d' and A, and application of time pressure theory (Fürstenau et al. in EPCE/HCII 2013, Lecture Notes in Artificial Intelligence (LNAI), pp. 143–154, 2014). RTO-controller working position (CWP-) performance was directly compared with that one of the conventional tower-CWP with direct out-of-windows view by means of simultaneous aircraft maneuver observations within the control zone at both operator positions. For this analysis we took into account correct (Hit rate), incorrect (False Alarms, FA) answers to discrimination tasks, and we took into account non-answers for a pessimistic quantification of RTO-performance. As initial working hypothesis this lead to the concept of time pressure (TP) as one major source of the measured response errors. A fit of experimental error rates with an error function derived from the Hendy et al. information processing (IP/TP)-hypothesis (Hendy et al. in Human Factors 39:30–47, 1997) provides some evidence in support of this model. We expect the RTO-performance deficits to decrease with the introduction of certain automation features to reduce time pressure and improve the usability of the videopanorama system.

Keywords Remote tower · Videopanorama · Field testing · Flight maneuvers · Two-alternative decisions · Signal detection theory · Information processing theory · Time pressure

N. Fürstenau (✉)
German Aerospace Center (DLR), Inst. of Flight Guidance, Braunschweig, Germany
e-mail: Norbert.fuerstenau@dlr.de; nfuerstenau@tonline.de

1 Introduction

The present chapter is based on results presented on the HCII conferences in Las Vegas (2013) and Crete (2014), published in Fürstenau et al. (2013) and Fürstenau et al. (2014). It extends the discussion of the RTO validation experiments in the previous chapter "Which metrics provide the insight needed? A selection of remote tower evaluation metrics to support a remote tower operation concept validation", performed within a DFS-DLR cooperation as the final work package of the DLR project RaiCE (see also Friedrich and Möhlenbrink (2013)).

Since more than ten years remote control of low traffic airports (Remote Tower Operation, RTO) has emerged as a new paradigm to reduce cost of air traffic control (Fürstenau et al., 2009; Hannon et al., 2008; Schmidt et al., 2007). It was suggested that technology may remove the need for local control towers. Controllers could visually supervise airports from remote locations by videolinks, allowing them to monitor many airports from a remote tower center (RTC) (Fürstenau et al., 2009). It is clear from controller interviews that usually numerous out-the-window visual features are used for control purposes (Ellis & Liston, 2010). In fact, these visual features go beyond those required by regulators and ANSP's (air navigation service providers) which typically include only aircraft detection, recognition, and identification (Schaik et al., 2010). Potentially important additional visual features identified by controllers in interviews involve subtle aircraft motion. In fact, the dynamic visual requirements for many aerospace tasks have been studied, but most attention has been paid to pilot vision (e.g. Watson et al. (2009)). In this work we investigate a group of visual cues derived from flight maneuvers within the range of observability in the control zone. They might be indicative of aircraft status and pilots situational awareness which is important with the higher volume of VFR traffic in the vicinity of small airports, the target application of RTO/RTC.

These considerations led to the design of the present validation experiment within the DLR project RAiCe (Remote Airport traffic Control Center, 2008–2012). The field test was realized within a DLR-DFS (German ANSP) Remote Airport Cooperation. Specifically dual-choice decision tasks (the subset of "Safety related maneuvers" in Friedrich and Möhlenbrink (2013)) were used for quantifying the performance difference between the standard control tower work environment (TWR-CWP) and the new RTO controller working position (RTO-CWP) based on objective measures from signal detection theory (SDT; parametric and non-parametric discriminability d′ and A respectively) (MacMillan & Creelman, 2005) and Bayes inference (Fürstenau et al., 2013, 2014) (a brief summary of the three methods is presented in Appendix A2). These analyses are complemented by an error model derived from the information processing/time pressure (IP/TP-) hypothesis of Hendy et al. (Hendy et al., 1997) for comparing the measured performance deficit of the RTO-CWP as compared with the Tower-CWP.

Experimental methods are reviewed in Sect. 2 followed by the results in Sect. 3 (response times, Hit and False Alarm rates, and non-answers). Using these data in Sect. 4 we present the analysis with the three complementary Bayes and SDT

methods, and we introduce a modified time-pressure based error function for fitting the measured error rates. We finish with a conclusion and outlook in Sect. 5.

2 Methods

In what follows we briefly review the experimental design for two-alternative decision tasks as part of the remote tower validation experiment and present additional details relevant for the IP-theory based analysis. Further details of the full passive shadow mode validation trial are reported in Friedrich and Möhlenbrink (2013) and chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation".

2.1 Participants

Eight tower controllers (ATCO's) from the German air navigation service provider DFS were recruited as volunteer participants for the experiment. The average age was 30 (stdev 12) years with 10 (stdev. 10) years of work experience, and they came from different small and medium airports. They took part at the experiment during normal working hours and received no extra payment. They were divided into 4 experimental pairs for simultaneously staffing the control tower (TWR-CWP) and the RTO-CWP.

2.2 Experimental Environment and Conditions

The experiment was performed as passive shadow mode test under quasi operational conditions on the four days July 17–20 2012. The remote tower system was located at the DFS-operated Erfurt-Weimar (EDDE) control tower. It was an improved version of the initial RTO-experimental testbed at Braunschweig airport which was in use since 2004 for initial verification and validation trials (Fürstenau et al., 2009; Schmidt et al., 2007) (see chapter "Remote Tower Experimental System with Augmented Vision Videopanorama"), and it corresponded to the prototype system described in chapter "Remote Tower Prototype System and Automation Perspectives". A comparable advanced design was located at DLR facilities in Braunschweig with RTO-CWP in the Tower-Lab simulation environment that was used for verification of system functions during the setup of the quasi-operational system (see chapters "Remote Tower Prototype System and Automation Perspectives" and "Multiple Remote Tower Simulation Environment (S. Schier)").

Figure 1 show the sensor system and the RTO-CWP with ca. 200°—videopanorama and operator console based on a reconstructed far view with five

Fig. 1 Remote tower installation with 200°—panorama and pan-tilt zoom camera sensor system at the tower roof (left photo), and operators workplace with 40″-HD-format displays (right). Reprinted from Fürstenau et al. (2013), with permission

HD-format 40″-displays (892 × 504 mm, 1920 × 1080 pixel, pixel distance = 0.47 mm). A separate monitor (left console side) displays the pan-tilt zoom camera which is controlled via a pen-input interaction display with virtual joystick. Twelve preset viewing directions and four zoom factors Z = 2, 4, 8, 16 (viewing angles 26°– 3°) could be selected. Additional monitors include (middle row from right to left): flight plan data, approach radar and weather display. In contrast to the experimental TWR-CWP the RTO-CWP was on the TWR ground floor in a separate room without visual contact to the airfield. The TWR-CWP was located close to the operational ATCO, but they were instructed not to communicate with each other.

Pre-defined flight maneuvers were generated with a DLR DO228 twin turboprop engine test aircraft (D-CODE, length 15.03 m, body height × width 1.8 × 1.6 m, wing span 16.97 m, wheel diam. 0.65 m) performing aerodrome circling.

The set of five well defined flight-maneuvers as stimuli for decision tasks at given positions within the EDDE control zone is indicated in Fig. 2 with a 3D plot of the
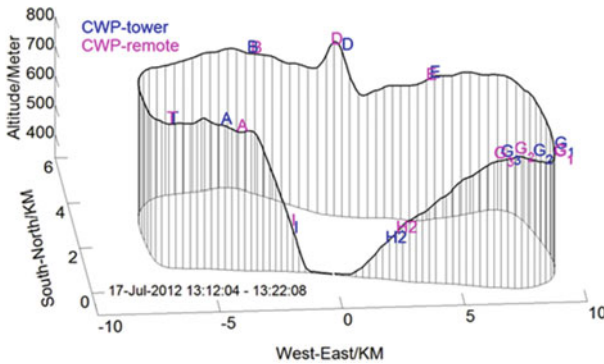
Fig. 2 DCODE trajectory measured with on-board satellite navigation. Letters indicate event positions with logged answers of TWR-CWP (blue) and RTO-CWP (red) operators to simultaneous task oriented on-line questioning. Distance between vertical lines = 5 s, projected to trajectory altitude minimum at ca. 350 m (sea level). Tower position (0, 0), height = 30 m. Reprinted from Fürstenau et al. (2013), with permission

logged on board GPS trajectory. Trajectory minimum altitude represents a runway overflight at about 30 m above ground. The two types of maneuver-stimuli at the respective positions (S1/S2 = maneuver/no maneuver) could be observed either visually-only (e.g. landing gear down) or visually and by radar (altitude change). During the experiment sometimes additional low volume normal traffic took place which now and then lead to delays in the traffic circle. Average duration of a full circle (= one run) was ca. 10 min yielding typically 140 min of experiment duration per participant pair for the nominally 14 full circles.

Radio communication between D-CODE pilots and flight engineer and the experimenter at the tower was realized with a separate research frequency in addition to the standard A/C–TWR radio channel. The available time for participant's responses to decision tasks was limited so that correct, incorrect and non-answers were possible.

2.3 Experimental Design and Task

Based on the fixed set of evaluated two-alternative events (A, D, G_1, G_2, $H_{1,2,3}$) at fixed positions during a single circle the concrete event situation (stimulus alternative S1 = maneuver/event, S2 = no maneuver/no event) for decision making were switched statistically between runs according to two mirrored scenarios with different task sequences. In this way during the 10–14 circles per experiment / participant for each event with two alternative stimuli (S_1, S_2, in random succession) per task 5–7 answers per event alternative and per participant were obtained for the analysis (TWR and RTO CWP condition as independent variables). The within subjects design (TWR vs. RTO-CWP) yields N = 40–56 answers (correct and incorrect, and non-answers for averaging with the 8 participants.

During one aerodrome circling the two participants at the TWR and RTO-CWP had to simultaneously answer 19 different types of questions concerning the D-CODE maneuvers (events), object detection, and weather status. The following subset of 5 of the 9 event related questions is evaluated with regard to hit- and false alarm rates using Bayes inference, signal detection theory (discriminability indices d' and A) and IP/TP theory based error model (in brackets: maximum response time Ta = interval until next task/question): 1. Does A/C perform repeated bank angle changes? (event position A; Ta = 20 s), 2. altitude variation ? (by 300 ft, event pos. D; 20 s), 3. landing light-off? Report status, event pos. G_1; 180 $_s$: switching-off not observable), 4. A/C on glide path?, event pos. G_2; 90 $_s$; 5. Landing gear-down?; Report during final approach; event pos. $H_{1,2,3}$, distance 1.5, 1, 0.5 km; 10 s). A subjective certainty rating on a 5-point scale was not evaluated for the present analysis.

Every pair of participants had to complete two experimental trials. For the first trial (duration approximately 140 min) in the morning till noon, the participants were randomly assigned to one of the two CWP's. Positions were changed for the second trial in the afternoon. The present data analysis was focused on deriving objective measures for the two-alternative decision tasks. Additional data evaluation was presented in Friedrich and Möhlenbrink (2013) (see also previous

chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation") addressing performance (answers given, response times, and sources of information) and subjective measures (debriefing, questionnaires).

3 Results

The response matrices of the measured estimates of conditional probabilities $p(y|S_1) =$ hit rate H, $p(n|S_1) =$ misses M, $p(n|S_2) =$ correct rejections CR, $p(y|S_2) =$ false alarms FA, for the two alternative situations (stimuli), S1, S2, structure the results of each of the five events. The evaluation of the answers on the five decision tasks of the eight participants, i.e. the percentage correct analysis in the previous Chap. 9 and (Friedrich & Möhlenbrink 2013) $(H + CR)/(p(S1) + p(S2))$ with neglection of non-answers (i.e. no decision during the available time Ta) had suggested no significant performance difference between TWR-CWP and RTO-CWP. A closer look into the statistics of the non-answers, however, revealed a significant increase under RTO-CWP as compared to TWR-CWP conditions, as shown in Fig. 3. It depicts the relative frequency of non-answers, separated for the TWR-CWP and RTO-CWP condition.
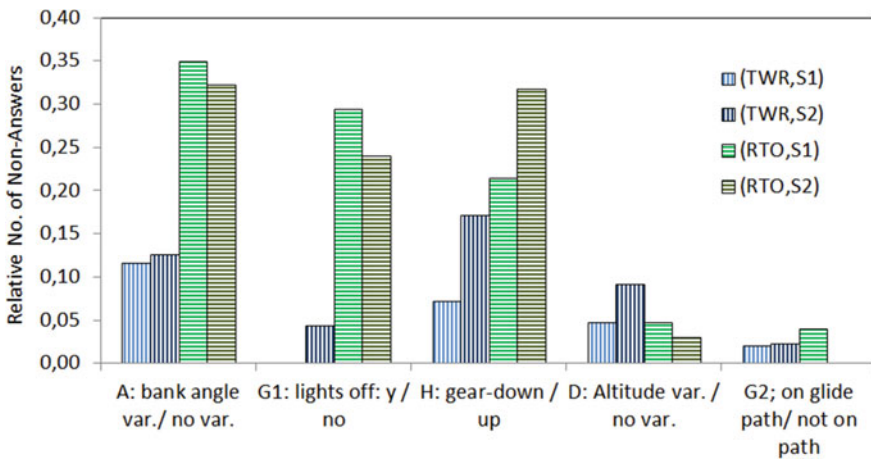
Fig. 3 Relative number of non-answers (included in the set of false answers (M(S1), FA(S2) for the pessimistic analysis = maximum errors) for the five analyzed decision tasks, separated for the two conditions TWR-CWP (left two columns, blue, vertical lines), RTO-CWP (right columns, green, horizontal lines), normalized with regard to the two respective alternative situations S1 (flight maneuver @ stimulus position, light colour), S2 (no flight maneuver @ stimulus position, dark colour)). Reprinted from Fürstenau et al. (2014), with permission

Table 1 Measured hit and false alarm rates (H = p(y|S1), FA = p(y|S2), ± stddev from Binomial distribution according to MacMillan and Creelman (2005) for five events and two conditions (TWR, RTO-CWP) with (a) non-answers excluded and (b) non-answers added to error rates FA and M. Ta = available decision time, Tr required average decision time with stderror of mean/seconds. Reprinted from Fürstenau et al. (2014), with permission

Event with alternatives S_1/S_2 (Ta/s)	Tr/s ± stderr	CWP	(a) Non-answers excluded		(b) Non-answers included	
			$p(y|S_1)$	$p(y|S_2)$	$p(y|S_1)$	$p(y|S_2)$
A: bank angle var.: y/n (20)	13.8 ± 1.7	TWR	0.92 ± 0.04	0.08 ± 0.04	0.81 ± 0.06	0.20 ± 0.05
	14.0 ± 1.1	RTO	0.93 ± 0.05	0.11 ± 0.05	0.60 ± 0.07	0.39 ± 0.07
D: Altitude var.: y/n (20)	8.8 ± 1.4	TWR	0.80 ± 0.06	0.03 ± 0.03	0.77 ± 0.06	0.12 ± 0.06
	12.4 ± 1.5	RTO	0.73 ± 0.07	0.03 ± 0.03	0.70 ± 0.07	0.06 ± 0.04
G1: lights off: y/n (180)	27.0 ± 6.6	TWR	0.94 ± 0.04	0.25 ± 0.07	0.94 ± 0.04	0.28 ± 0.07
	95.4 ± 7.4	RTO	0.92 ± 0.06	0.63 ± 0.08	0.65 ± 0.08	0.72 ± 0.07
G2: Glidepath y/n (90)	21.6 ± 6.4	TWR	0.90 ± 0.04	0.32 ± 0.07	0.88 ± 0.05	0.33 ± 0.07
	34.2 ± 8.1	RTO	0.92 ± 0.04	0.22 ± 0.06	0.88 ± 0.05	0.22 ± 0.06
H: gear-down: y/n (10)	8.1 ± 0.9	TWR	0.98 ± 0.02	0.06 ± 0.04	0.91 ± 0.04	0.22 ± 0.06
	9.2 ± 0.5	RTO	0.98 ± 0.02	0.07 ± 0.05	0.77 ± 0.06	0.37 ± 0.08

This result suggested to analyse two types of response matrices: (a) (optimistic) neglecting non-answers, (b) (pessimistic) interpreting non-answers as false decisions (M or FA). In this way we obtain for each of the decision tasks an optimistic and a pessimistic estimate with regard to decision errors. The interpretation of the non-answers as erroneous responses appears to be justified due to increased uncertainty about the correct answer resulting in hesitation to respond at all because tower controllers work ethics requires decision making with high certainty.

Table 1 lists the measured hit and false alarm rates (H, FA ± standard deviations derived from binomial distributions) for the five events to be analysed, together with the average response times Tr and available response times Ta. In addition to H and FA, the rate of misses M = 1 − H is required for calculating the total number of errors to be compared with the formal time-pressure error model in Sect. 4.3.

Comparing the measured hit and false alarm rates for all five events under TWR and RTO conditions with non-answers not considered (optimistic case (a): left two data columns), the RTO-CWP exhibits no significant difference as compared to the TWR-CWP. If however, the non-answers are interpreted as erroneous responses and correspondingly attributed to rates FA and M (pessimistic case (b): right two data columns), significant differences TWR versus RTO are obtained (smaller H(RTO), larger FA(RTO)) for event/task A (bank angle variation?), H (gear down?), G1 (lights off?), whereas for event/tasks D and G2 responses again exhibit no significant difference. The latter two tasks reflect the fact that altitude information could be read directly from the radar display and operators were free to select their appropriate

information source. An extremely high FA difference TWR versus RTO is observed for both case (a) and (b) for the "lights-off" event which is reflected also in a large difference of decision distance (correlated with response time). This was already reported in the previous chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation" where percentage correct analysis for the optimistic case (a) analysis (without non-answers) was evaluated.

4   Data Analysis and Discussion

4.1   Technical Limitations

Technical parameters of the reconstructed far view with videopanorama and PTZ (Fürstenau et al., 2009; Schmidt et al., 2007) leads to predictions concerning performance differences under the two conditions TWR and RTO-CWP. The measured performance also depends on the usage of the different available information sources, in particular videopanorama, PTZ, and approach radar, and the general system usability. The relevance of the used RTO metrics is discussed in the previous chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation".

The visibility limitations of the videopanorama are quantified by the modulation transfer characteristic (MTF, see Appendix A), with the digital (pixel) camera resolution providing the basic limit (Nyquist criterion) for detectable objects and maneuvers: angular resolution was estimated as $\delta\alpha \approx 2$ arc min $\approx 1/30° \approx 0.6$ m object size/km distance per pixel under maximum visibility and contrast (about half as good as the human eye (1 arcmin)). Reduced contrast of course reduces the discriminability according to the MTF and the question arises how the discriminability difference TWR versus RTO-CWP is affected. The gear-down situation at positions H1–H3 with wheel diameter 0.65 m, e.g. can certainly not be detected before the wheel occupies, say, 4 pixels which for the 40″ display (0.55 mm pixel size) means a viewing angle of ca 1 mm/2 m ≈ 0.5 mrad corresponding to the visual resolution of the eye (1 arcmin) under optimum contrast. This estimate results in a panorama based gear-down detectability distance of <500 m. It means that under RTO conditions this task requires usage of PTZ in any case in order to allow for a decision. The same argument is valid for the detection of bank angle changes at position A following the overflight of the runway because it requires optical resolution of the A/C-wings. The "lights-off?"-decision (G1) has a somewhat different character because in situation S_1 (lights-off, answer "yes" = hit) observers usually wait until they actually detect the A/C whereas situation S_2 can be recognized at a larger A/C distance due to the higher contrast ratio of landing-light-on/background luminance.

## *4.2   Bayes Inference: Risk of Unexpected World State*

A brief overview of this method is provided in Appendix B of the present volume. Measured rates of hits, misses, correct rejections, and false alarms (H, M, CR, FA) are estimates of conditional probabilities p(di|Sj) (i $\neq$ j) which by means of the Bayes theorem are used after the measurement by multiplying with the a-priori knowledge p(S$_i$) for calculating the inverse probabilities, i.e. risk of an actual situation contradicting the decision based on the perceived evidence:

$$p(S_i|d_j) \;=\; p(d_j|S_i)\, p(S_i) \,/\, p(d_j) \tag{1}$$

with responses d$_i$, i $=$ 1, 2, d$_1$ $=$ yes, d$_2$ $=$ no, p(y) $+$ p(n) $=$ 1 for a given situation S$_i$, and the probability of any of the two possible situations (world states) p(S$_1$) $+$ p(S$_2$) $=$ 1, under TWR and RTO conditions of the experiment. Of particular interest are the two probabilities for the risk of a situation contradicting the decision based on the observed evidence on the nature of the observed event. p(S$_1$|n) is the probability of e.g. the aircraft with bank angle variation (situation S$_1$, e.g. signaling some special situation during radio interruption) conditional the case that no variation is perceived (i.e. a Miss). P(S$_2$|y) is the probability for a situation with a/c not performing bank angle variation conditional on the false response "variation perceived"(i.e. a False Alarm). The following Figs. 4, 5 depict the corresponding Bayes inference results (risk) for the five events for analysis case (b), i.e. non-answers treated as errors (S$_1$: non-answer $=$ M; S2: non-answer $=$ FA). It clearly shows that the risks for world states not corresponding the observed evidence (decision, averaged over the eight participants and seven decisions per situation S$_i$) are at least two times as high for the RTO-CWP as compared to TWR- CWP, with the exception of the events D and G2 (altitude variation and deviation from glide path occuring in 7 of the 14 circles).

Table 2 lists the calculated Bayes inference values (averaged over participants and repeated observations) for the five different stimuli (events) and two conditions (TWR, RTO) for case (a) non-answers not considered, and (b) non-answers taken as wrong answers.

Fig. 4 Bayes inference on probability ($\pm$stddev) of world state S1 (event/maneuver occuring) conditional on (false) decision d2 $=$ event not occuring, based on perceived evidence (case (b))
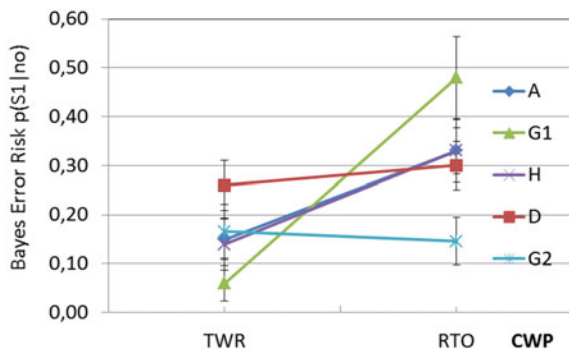
Fig. 5 Bayes inference on probability (±stddev) of world state S2 (event/maneuver not occuring) conditional on (false) decision d1 = event occuring, based on perceived evidence (case (b))
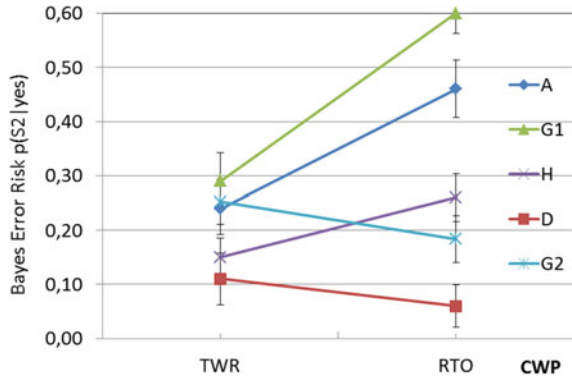
Table 2 Bayes inference for TWR and RTO-CWP from response data, for cases (a), (b). Std.dev. estimates from Binomial distribution

Event with S_1 or S_2	CWP	(a) Non-answers excluded		(b) Non-answers included	
		$p(S_1|n)$	$p(S_2|y)$	$p(S_1|n)$	$p(S_2|y)$
A: bank angle var	TWR	0.06	0.10	0.15 (0.04)	0.24 (0.05)
	RTO	0.06	0.13	0.33 (0.05)	0.46 (0.05)
G1: lights off	TWR	0.06	0.26	0.06 (0.04)	0.29 (0.05)
	RTO	0.13	0.50	0.48 (0.09)	0.60 (0.04)
H: gear-down	TWR	0.03	0.04	0.14 (0.05)	0.15 (0.04)
	RTO	0.04	0.04	0.33 (0.06)	0.26 (0.04)
D: Altitude var	TWR	0.22	0.03	0.26 (0.05)	0.11 (0.05)
	RTO	0.26	0.03	0.30 (0.05)	0.06 (0.04)
G2: above glide path	TWR	0.14	0.24	0.17 (0.06)	0.25 (0.04)
	RTO	0.10	0.18	0.15 (0.05)	0.18 (0.04)

As expected from Table 1, significant differences are observed for the Bayes inference analysis of RTO versus TWR performance with analysis case (b) (i.e. with non-answers included, right two columns). The calculated risk for the actual world state occurring to be in contradiction to the perceived (hypothetical) situation is very low for non-answers excluded (analysis case (a)) for both TWR and RTO conditions, and no significant TWR-RTO difference is observed, with the exception of stimulus G1 (lights off). The error risk increases significantly with non-answers included (case (b)), which in fact is not surprising. Not expected was the result that in the RTO-CWP the risk in most cases at least doubles as compared to TWR-CWP. The altitude variation (event D) and deviation from glide path (G2) in contrast exhibit no significant difference which can be explained by the fact that in both cases the majority of decisions were made based on MODE-S secondary radar display

information which includes altitude information in the labels with typically 25 foot interval and 4 s update rate.

4.3 Discriminability d′ of Aircraft Maneuvers

The results of the Bayes inference analysis is supported by a more sophisticated evaluation of data from Table 1 using signal detection theory (SDT). In contrast e.g., to percentage correct (p_c) evaluation of subjects decisions on dual choice tasks it separates the decision maker's discriminability d' from the subjective decision bias c (= decision criterion or individual tendency to more conservative, i.e. avoiding FA at the cost of decreasing H, or more liberal decisions) (MacMillan & Creelman, 2005).

Within the theoretical framework of SDT the two alternative stimuli S_1, S_2 for each event define independent statistical variables. Each set of decisions of a single subject for the 14 aerodrome circles with one of the events A, D, G_1, G_2, H represents a sample of the randomly presented S_1- and S_2-alternatives. For calculation of (parametric) discriminability d' the subjective responses are assumed to be drawn from independent equal variance Gaussian ($\mu_{1,2}$, σ) densities modelling the familiarity with situations S_1 and S_2 (MacMillan & Creelman, 2005). Any discriminability difference between TWR and RTO may be quantified by corresponding coefficients $d' = \mu_1 - \mu_2 = z(H) - z(FA)$, and subjective decision bias (criterion) c = 0.5(z(H) + z(FA)), with z() = z-score as calculated from the inverse cumulative densities.

Figure 6 depicts for analysis of case (b) the average (H, FA) data of the three visual discrimination tasks at positions A, G1, H in the receiver operating characteristic (ROC) space together with two sets of pairwise ROC curves (one pair for TWR and RTO conditions each). One set (solid lines) is parametrized by discriminability d', the other (dashed) by the subjective decision bias c. E.g. d' = 3 means that the Gaussian densities mean values of perceived situations S_1, S_2 differ by 3 normalized stddev (σ = 1). Under the above mentioned conditions each (d′, c)-ROC curve-pair is unambiguously determined by the single average (H, FA)-point. The d′ and c values are calculated via standard procedures (inverse cumulative densities from the (H, FA) data). Dotted lines indicate estimates of standard deviations s(d′) as described in MacMillan and Creelman (2005), based on the binomial variation of measured proportions from sample to sample.

The following Table 3 summarizes the discriminability d′ and criteria c (decision bias) corresponding to Fig. 6, and like Table 2 includes tasks D (altitude variation) and G2 and it includes both data analysis cases: optimistic (a) and pessimistic (b).

Again both data analysis cases are listed: optimistic (a) and pessimistic (b). In agreement with the Bayes inference the case (a) analysis (non-answers are not considered in the data analysis) shows no significant difference between TWR and RTO-CWP conditions, with the exception of task G1: for the lights-off stimulus even with non-answers not considered RTO exhibits a significant decrease of discriminability. This was already reported in MacMillan and Creelman (2005) for the percentage correct analysis.

Fig. 6 Measured data points
in ROC space of average hit
and false alarm rates
(Pessimistic analysis case
(b)) of visual-only
events/tasks **a** gear down, **b**
bank angle variation, **c**
lights-off for TWR (cross)
and RTO
(circle)—conditions,
together with the
isosensitivity and isobias
curves parametrized by
discriminability d' (solid
lines) and criteria c (dashed)
respectively. Dotted lines are
standard dev. based on
procedures described in
MacMillan and Creelman
(2005). Redrawn from
Fürstenau et al. (2013), with
permission
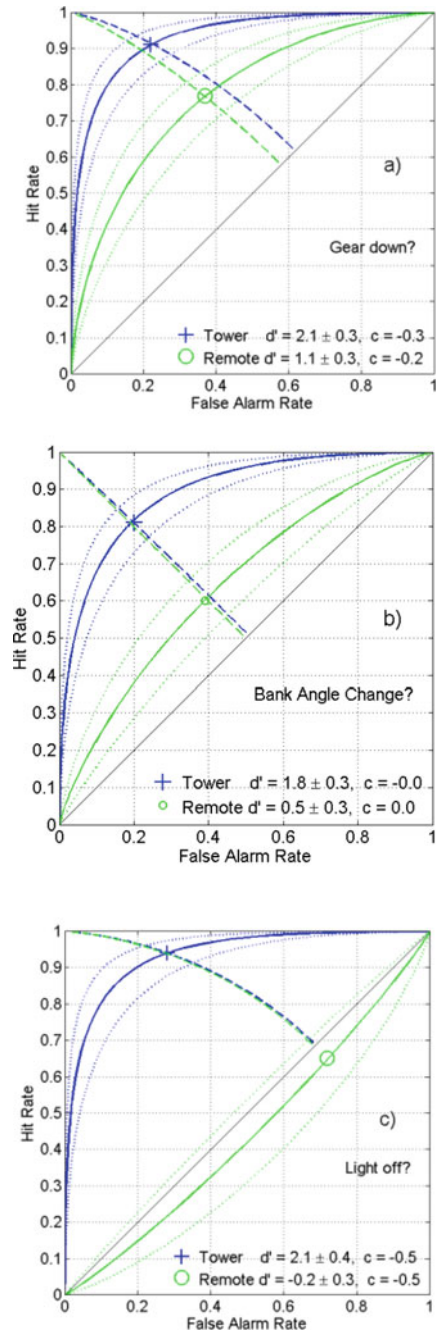
Table 3 Discriminability d′ (±stddev) and criteria c for both (a) optimistic and (b) pessimistic analysis as obtained from z-scores based on response matrices (hit and false alarm rates)

Event	CWP	(a) Non-answers excluded		(b) Non-answers included	
		d′ (±stddev)	c	d′ (±stddev)	c
A	TWR	2.81 (0.39)	−0.01	1.75 (0.30)	−0.02
	RTO	2.72 (0.45)	−0.11	0.54 (0.26)	0.00
G1	TWR	2.24 (0.40)	−0.45	2.14 (0.40)	−0.49
	RTO	1.05 (0.43)	−0.86	−0.20 (0.30)	−0.48
H	TWR	3.63 (0.53)	−0.30	2.12 (0.32)	−0.30
	RTO	3.47 (0.55)	−0.30	1.07 (0.27)	−0.20
D	TWR	2.69 (0.50)	0.49	1.90 (0.35)	0.22
	RTO	2.48 (0.48)	0.62	2.07 (0.40)	0.52
G2	TWR	1.74 (0.31)	−0.40	1.61 (0.30)	−0.37
	RTO	2.15 (0.33)	−0.31	1.94 (0.31)	−0.21

Also for case (b) analysis the Bayes inference results are confirmed: again a significant decrease of visual discriminability is observed if non-answers are attributed to erroneous decision (M, FA), for task G1 (landing lights off) even zero detectability. As expected tasks D and G2 requiring decisions on altitude (change) again exhibit no significant difference TWR versus RTO-CWP. Decision bias in most cases does not exhibit significant differences between TWR and RTO-CWP.
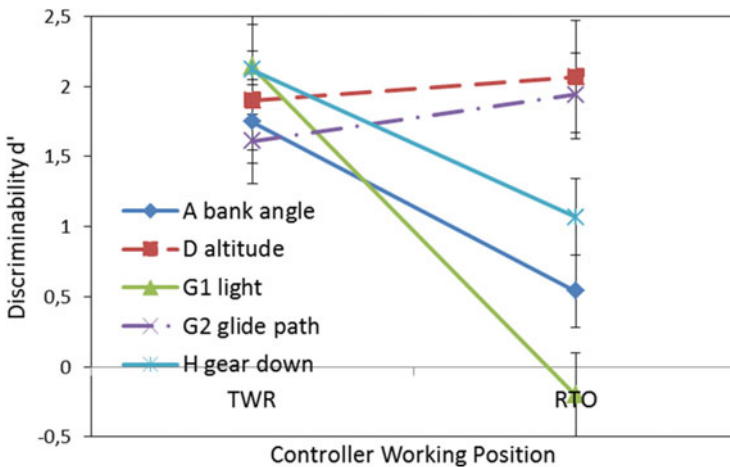
Fig. 7 Discriminability d′ (units = normalized stddev σ/μ, μ = mean) according to SDT derived from hit and false alarm rates in Table 1, for case (b): non-answers: = false answers. D and G2 (dash-dotted lines): decisions about altitude (variations). A, G1, H = visual-only information (solid lines). Error bars = standard dev. based on binomial distribution (MacMillan & Creelman, 2005)

Figure 7 depicts the discriminabilities d′ with standard deviations derived from Binomial distributions (MacMillan & Creelman, 2005) for analysis case (b): non-answers included. The figure summarizes and highlights the significant performance deficit of the RTO-CWP with respect to the visual-only information tasks (A, G1, H, solid lines). The increased RTO-CWP probability p(Si|dj), i ≠ j for drawing erroneous conclusions based on subjectively perceived evidence in the case of Bayes analysis (Figs. 4 and 5) is reproduced here by the decreased visual parametric discriminability d′ (based on the Gaussian (μ_i, σ) assumption for criterion distribution. Again also the difference between visual-only and visual and Radar information source (D, G2) is confirmed for the RTO case, however not for the TWR-CWP, indicating a usability deficit of the former.

The d′ calculation presupposes equal variance Gaussian densities for the subjective responses or familiarities with the two stimulus alternatives which was not possible to verify with our limited data set. We can obtain additional confidence in our results by means of the nonparametric discriminability index A with bias/criterion parameter b which is independent of the mentioned precondition.

4.4 Nonparametric Discriminability A

In this section we will confirm the parametric discriminability (d′) analysis with an additional one using the non-parametric discriminability index **A** (not to be confused with aircraft maneuver A; generally assumed independent of the Gaussian (μ, σ) assumption for familiarity). Discriminability **A** is defined as the average of the areas under the maximum and the minimum proper ROC-isosensitivity curve (constant d′, (Fürstenau et al., 2013; MacMillan & Creelman, 2005)) defined by a single (H, FA)-data point and varies between 0.5 (d′ = 0) and 1 (lim d′ → ∞). For the calculation of **A** and b we use corrected algorithms (functions of H, FA) derived in Zhang and Mueller (2005). Figure 8 shows one example (stimulus A: bank angle variation) of (**A**, b)-parametrized isopleths determined by the two TWR and RTO-CWP datapoints. Figure 9 depicts the **A**-values of the five tasks at A, D, G1, G2, H, for the two conditions TWR-CWP, RTO-CWP (again pessimistic analysis case (b): non-answers included as false answers).

The example (**A**, b)-isopleths in Fig. 8 for maneuver **A** show zero decision bias (b = 1), however a significant discriminability decrease for RTO-CWP (minimum A = 0.5 = positive diagonal), as expected.

In agreement with the d′-discriminabilities in the previous section the non-parametric indices **A** in Fig. 9 exhibit no significant differences between TWR and RTO-CWP conditions for events D, G2 (event sub-set with altitude stimulus; altitude information additionally provided by radar via Mode-S transponder), whereas the A-decrease for the visual-only subset {A, G1, H} is again evident. Moreover even a reduction of the number of erroneous decisions by attributing a 50% chance to non-answers to be correct instead of assuming 100% wrong answers) leaves the RTO-performance decrease for visual-only tasks significant. The drop to chance
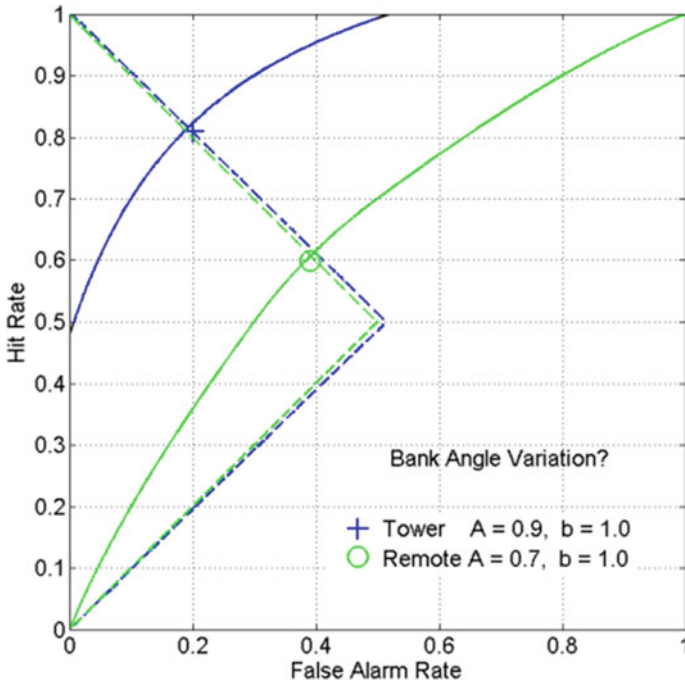
Fig. 8 Maneuver A as example isosensitivity curves for TWR and RTO case (b) analysis (straight lines, A-isopleths) and decision bias (dashed, b-isopleths). Reprinted from Fürstenau et al. (2014), with permission

level of RTO-CWP discriminability for case G1 is again confirmed and attributed to the RTO-resolution and contrast deficit which prohibits recognition of A/C even with lights on for short response times Tr: when participants at RTO-CWP after task initialization had waited some 10 s or so without recognizing landing lights they often simply guessed lights to be off or gave no answer, contributing to FA-errors.

4.5 Error Prediction Using the Information Processing/Time Pressure Hypothesis

In order to determine appropriate solutions for rising the RTO-CWP performance to at least the level of the TWR-CWP we have to find explanations for the measured discriminability deficits. The RTO-CWP performance for decision making using videopanorama and PTZ replacement of the tower far-view should be at least as good as the TWR-CWP so that users can be certain that replacement of the out-of-windows view has a potential of even improving their work condition. Referring to Figs. 4, 5, 7, and 9 this means that the decision error and discriminability differences
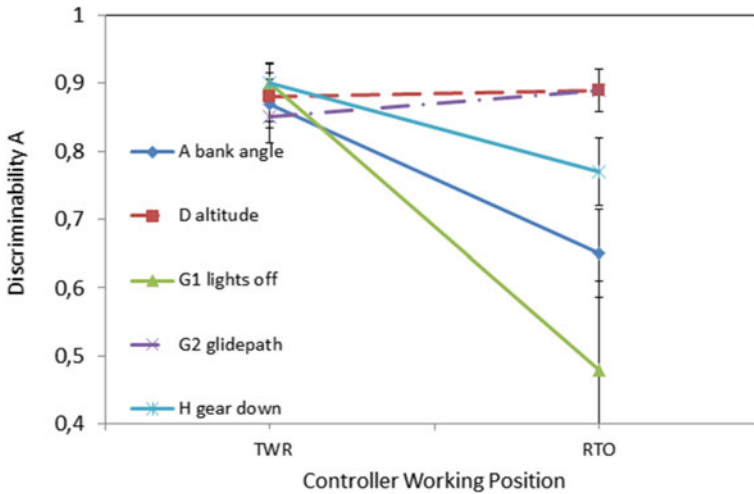
Fig. 9 A as calculated according to Zhang and Mueller (2005) from hit and false alarm rates in Table 1, case b. D and G2 (dash-dotted lines): decisions about altitude (variations). A, G1, H = visual-only information (straight lines). Error bars = standard dev. based on binomial distribution (MacMillan & Creelman, 2005). Reprinted from Fürstenau et al. (2014), with permission

between TWR and RTO-CWP of the visual discrimination tasks have to vanish in the final improved RTO-CWP design. In order to approach the required improvements in a scientifically founded way we tried to narrow down the origin of this deficit via an information processing hypothesis.

A (algorithmically) simple theoretical model with some potential for explaining the observed performance differences quantified in terms of decision-error probability, is based on the perceptual control/information processing theory (PCT/IP) of Hendy et al. (1997). Because our experiment was not initially designed for testing this theory we can only expect a first impression on the relevance of the corresponding assumptions. The core idea is to formalize the information processed as part of the total information required for a correct answer (Br measured in bits) as function of time pressure TP. TP is the ratio of required time Tr (to acquire Br) and the available time Ta: $TP = Tr/Ta$. Assuming constant cognitive processing rate (channel capacity C: $Tr = Br/C$) the rate of information processing demanded RID is related to TP via $TP = RID/C$, with $RID = Br/Ta$. Hendy et al. (1997) derived simple algorithms for modeling dependent variables like operator workload (OWL), success ratio, and number of errors as function of TP. For the latter they suggested an exponential dependency for the increase of decision errors with TP, where TP increases linearly with the number N of objects to be analysed (in our case N = 1): $TP = t_0(1 + b_1 N)/Ta$, and t_0 = minimal decision time for N = 0. For error probabilities we modify Hendy's algorithm in order to use our maximum error probability p_{err} (lim TP → 0) = 0 = p_{min} (= zero error for vanishing time pressure) and p_{err} (TP ≫ 1) = 0.5 = p_{max} (= just guessing, no information available) as boundary conditions. Keeping

the original assumption that errors start to grow exponentially with TP but then level off at p_{max} we arrive at a logistic function with threshold and sensitivity parameters as one possible model:

$$p_{err} = 0.5\left(1 + exp\left\{-\left(\frac{TP - \mu}{\beta}\right)\right\}\right)^{-1} \tag{2}$$

μ $(0 \leq \mu \leq 1)$ models the threshold where the observer starts shedding most information due to increasing workload (stress due to TP increase). It fulfills the conditions that $\lim(TP \gg \mu)$ $p_{err} \to 0.5$. $\lim(TP \to 0)$ $p_{err} \to 0$. The latter condition is fulfilled as long as $\mu/\beta \gg 1$, i.e. steep slope (= error sensitivity $dp_{err}/dTP = 1/8\beta$ at $TP = \mu$ and/or large threshold). Figure 10 shows for the three visual discrimination tasks the results of nonlinear fitting of the respective two data points $p_{err}(TP)$ at $TP(Tr(TWR))$, $TP(Tr(RTO))$ with the two boundary conditions ($p_{err}(\lim TP \to 0) = 0$, $p_{err}(\lim TP$
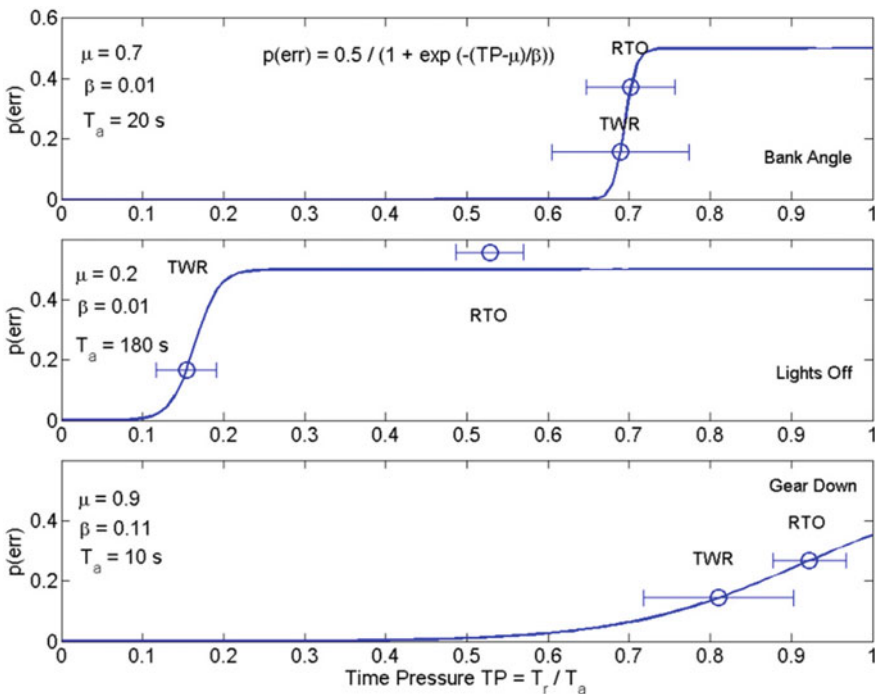
Fig. 10 Decision error probabilities for TWR and RTO-CWP versus time pressure TP ($\pm$stderr of mean, n = n(error) + n(correct) = 80 …100) for tasks where visual/PTZ-information was used for decision making. Standard errors of p(error) are smaller than the circles of data points. Logistic error model (Eq. 2) derived from IP/TP-theory (Hendy et al., 1997) for fitting $p_{err}(TP)$. Reprinted from Fürstenau et al. (2014), with permission

$\to \infty) = 0.5$) using model—Eq. (2) for the three visual-only tasks. For characterising the experimental results in terms of (μ, β) we have to use the total number of erroneous decisions for the full set $(n(S1) + n(S2))$ of trials per subject instead of the conditional probabilities, misses and false alarm rates $M = 1 - H$, FA: $p_{err} = (n1\,M + n2\,FA)/(n1 + n2)$ as used for the SDT analysis.

The results indicate the principal applicability of the logistic error model because all three cases yield reasonable threshold ($\mu \leq TP = 1$) and error sensitivity parameters β. The RTO-performance deficit always seems to correlate with some kind of time pressure. According to IP-theory decision errors should increase significantly due to increasing stress when Tr approaches Ta and to shedding of information when Tr > Ta. This is reflected by our results only for event H (gear down) with the shortest Ta = 10 s. Variation of threshold μ with event(stimulus) can be explained by the fact that the three specific events provide quite different stimulus conditions for the decision making as described in Sects. 2 and 3. The fact that only for the gear-down task an approximately exponential increase of errors is observed at TP $\approx$ 1 according to Hendy et al. (1997) with $\mu \approx 1$ whereas a threshold behavior at lower μ is quantified by the IP/TP model for tasks A, G1. This indicates at least one more performance limiting factor besides time pressure, such as PTZ-camera contrast/resolution and operator training. For lights-off decision the RTO-HMI contrast deficit should play a major role: the average response appears completely at random. Nevertheless also in this case a long waiting time of the observer (until first A/C sighting) before beginning to gather visual evidence might lead to increasing stress due to uncertainty.

5   Conclusion

The present analysis of two-alternative decision making with safety related aircraft maneuvers explains the observed discrepancy in the percentage correct analysis of the corresponding observation data (p_c, neglecting non-answers), as compared to the subjective success criteria (Friedrich & Möhlenbrink, 2013). The perceived safety in Friedrich and Möhlenbrink (2013) was rated as insufficient by participants which agrees with the objective data analysis presented in this Chapter.

The detailed analysis, based on detection theory (SDT) and Bayes inference confirms the vanishing of the difference between TWR- and RTO-CWP (suggesting sufficient RTO performance) when neglecting non-decisions during simultaneous decision making at TWR- and RTO-CWP. If however, non-decisions are taken into account and interpreted as false responses (misses(S_1) or false alarms(S_2)) we arrive at significant error increase under RTO as compared to TWR conditions. Correspondingly reduced discriminability indices A(non-parametric) and d'(parametric, Gaussian assumption for familiarity with stimulus)) are obtained and confirmed by Bayes inference, the latter quantifying the probability of a world state in contradiction to the evidence based decision.

The results indicate a usability deficit of the RTO-HMI (videopanorama and PTZ) in its present version due to time pressure as one possible reason. Data analysis

with a modified version of the Hendy et al. information processing / time pressure theory (IP/TP) (Hendy et al., 1997) indicates additional origins of performance decrease due to threshold behavior of decision errors significantly below the TP = 1 value. It is expected that increased automation (e.g. automatic PTZ-object tracking and augmented vision, e.g. data fusion with approach radar) will increase usability, and in combination with improved operator training could solve the performance problem. This is supported by the analysis of the remote tower metrics (RTM) as discussed in the previous Chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation". There a difference was found in usage of information sources between TWR- and RTO-CWP. As expected from the visual resolution deficit of the RTO-videopanorama a major switching tendency was found from panorama to PTZ-camera as information source for decision making (Chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation", Table 5). In any case further experiments are required for clarifying the role of time pressure and validating the effect of a higher level of automation and/or measures for improved usability. Experiments are preferably realized as human-in-the loop simulations with appropriate design for quantifying time pressure variation, with a task design that avoids non-answers.

Because of the significant effort required for the HITL-experiments and field tests, the initial results of the IP/TP-model suggest as intermediate step model based computer simulations for preparing corresponding HITL- and field experiments. For this purpose the commercial tool IPME (Integrated Performance Modeling Environment (Fowles-Winkler, 2003)) appears useful which integrates the PCT/IP-based approach together with a resource based theory so that by means of simulations it would allow for further clarification of the influence of different performance shaping functions.

References

Ellis, S. R., & Liston, D. B. (2010). *Visual features involving motion seen from airport control towers.* In *Proceedings of the 11th IFAC/-IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems 9/31–10/3, 2010.* Valenciennes, France

Fowles-Winkler, A. M. (2003). Modelling with the integrated performance modelling environment (IPME). In *Proceedings of the 15th European Simulation Symposium*. SCS European Council, ISBN 3–936150–29-X.

Friedrich, M., & Möhlenbrink, C. (2013). Which data provide the best insight? A field trial for validating a remote tower operation concept. In *Proceedings of the 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM 2013)*.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Papenfuss, A., & Kaltenhäuser, S. (2009). Steps towards the virtual tower: Remote airport traffic control center (RAiCe). In *Proceedings of the EIWAC 2009, ENRI International Workshop on ATM & CNS, Tokyo, 5.–6.3.2009* (pp. 67–76).

Fürstenau, N., Mittendorf, M., & Friedrich, M. (2013). *Discriminability of flight maneuvers and risk of false decisions derived from dual choice decision errors in a videopanorama-based remote tower work position.* In D. Harris (Ed.), *EPCE/HCII 2013, Part II, Lecture Notes on Artificial Intelligence (LNAI)* (Vol. 8020, pp. 105–114).

Fürstenau, N., Mittendorf, M., & Friedrich, M. (2014). Model-based analysis of two-alternative decision errors in a videopanorama-based remote tower work position. In D. Harris (Ed.), *EPCE/HCII 2013, Lecture Notes on Artificial Intelligence (LNAI)* (Vol. 8532, pp. 143–154).

Hannon, D., Lee, J., Geyer, T. M., Sheridan, T., Francis, M., Woods, S., & Malonson, M. (2008). Feasibility evaluation of a staffed virtual tower. *The Journal of Air Traffic Control*, 27–39.

Hendy, K. C., Jianquiao, L., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors, 39*(1), 30–47.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory*. Psychology Press.

Van Schaik, F. J., Lindqvist, G., & Roessingh, H. J. M. (2010). *Assessment of visual cues by tower controllers.* In *Proceedings of the 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems 8/31–9/3, 2010*. Valenciennes, France.

Schmidt, M., Rudolph, M, Werther, B., & Fürstenau, N. (2007). Development of an augmented vision videopanorama human-machine interface for remote airport tower operation. In *Proceedings of the HCII2007 Beijing, Springer Lecture Notes on Computer Science* (Vol. 4558, pp. 1119–1128).

Watson, A. B., Ramirez, C. V., & Salud, E. (2009). Predicting visibility of aircraft. *PLoS One, 4*(5), e5594. Published online 2009 May 20. https://doi.org/10.1371/journal.pone.0005594

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrica, 70*(1), 203–212.

Human-in-the-Loop Simulation for RTO Workload and Design

Multiple Remote Tower Simulation Environment

S. Schier-Morgenthal

Abstract The research on remote tower operation faces multiple challenges. Separate traffic flows are getting dependent, the design of the tower controller working places needs to be revised and new sensor technique must be tested. In accordance to this great bandwidth of design and adaption works a development and validation platform is of substantial need. Among the validation tools recommended by the European operational concept validation methodology, simulations play a crucial role. Within the present chapter a comprehensive simulation approach is discussed. This approach connects fast-time and human-in-the-loop simulation to determine traffic effects and the consequences for the air traffic controllers. Moreover, a transfer into the field using passive shadow mode tests is provided, replacing simulation components step by step with operational data.

Keywords Remote tower · Simulation · Field trials

1 Motivation

The introduction of Remote Tower Operations (RTO) into the air traffic systems is a major change in several aspects. Traffic streams which used to be independent are now getting dependent due to the fact that the responsibility is combined in the same control center (cf. Hagl et al., 2019). The working procedures and task load within remote tower differ from conventional tower operation and provide the potential of workload balancing (cf. Josefsson et al., 2019). Last but not least new sensor techniques are necessary to allow remote operations (cf. Papenfuss et al., 2020). These aspects need to be examined carefully and comprehensively to develop operational concepts for a safe and sustainable remotely controlled air traffic.

The European Operational Concept Validation Methodology (EUROCONTROL, 2010) provides a basic guide to evaluate and develop new systems and procedures for air traffic management. A major challenge to conduct research according to E-OCVM

S. Schier-Morgenthal (✉)
German Aerospace Center (DLR), Institute of Flight Guidance, Braunschweig, Germany
e-mail: sebastian.schier@dlr.de

is the design of an appropriate validation platform. This platform must be able to examine RTO concepts and techniques in a comprehensive manor. For instance, a new working place for the remote tower controllers can only be successfully developed, if the dependencies of the different airports are considered. Comparably, new sensor developments need to take into account the abilities to gather data as well as the potential to integrate this data into the new working positions. This comprehensive validation is only possible if a seamless process over the different validation techniques is provided.

This chapter summarizes developments in simulation techniques which were introduced between 2007 and 2020 to address the challenge of a comprehensive RTO validation environment. As a first step the methods proposed by E-OCVM are evaluated towards their potential to support RTO research. Out of the provided methods, Fast-Time Simulation (FTS), Human-In-The-Loop simulation (HITL simulation) and shadow mode trials will be concluded as the most relevant once. Following this analysis, an approach that connects FTS and HITL simulation in principle is derived. This approach is extended by the additional requirements HITL simulations have to fulfill in RTO research, leading to a new design of HITL tower simulations. Last but not least, the transfer of HITL simulation to shadow mode trials is addressed by showing a step-by-step transfer into operations. As the results of this new approach towards RTO research, the design of the validation platform and several case studies are presented. The chapter is concluded by a discussion of the applied methods and a look into the future of RTO validation activities.

2   Method: Design of a Comprehensive RTO Validation Platform

2.1   E-OCVM Guideline

Starting the research on RTO, the main objective was to determine whether tower and ground controllers are able to work in an arbitrary distance from their airport, without the conventional direct view out-of-windows, i.e. without the physical control tower (cf. Schmidt et al., 2006). Primarily, this is a matter of feasibility because in the beginning it was not clear if RTO would be possible at all. The limiting factor within this feasibility subject is the human operator. If some of the information provided in a conventional tower is not available (e.g. acoustic emissions from the airfield) or available only with reduced quality (e.g. video image vs. real out-of-windows view), there is a need to evaluate if the error probability of the operator's decision-making increases under RTO conditions.

Selecting the appropriate methods for the remote tower validation the European standard of air traffic system validation E-OCVM (cf. EUROCONTROL, 2010) needs to be considered. E-OCVM defines a process model from the initial idea to the final operational system. Within this model, the stage V2 (cf. Fig. 1) addresses
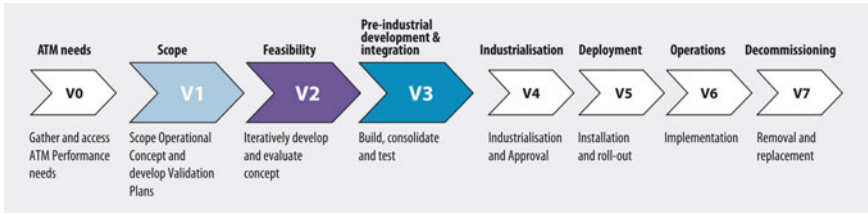
Fig. 1 E-OCVM lifecycle model (*Source* EUROCONTROL, 2010, p. 16)

feasibility questions. EUROCONTROL suggests to take a case-based approach (cf. EUROCONTROL, 2010, p. 50ff.) on this V2 stage including the human performance case. For this human performance case the tools gaming, fast-time simulations, real-time simulations, shadow mode and live trials are suggested (cf. EUROCONTROL, 2010, p. 49). The following analysis provides an overview of the methods:

- **Gaming** focusses on interaction between different parties (cf. Freese et al., 2020). Therefore, an abstract model of the working environment is provided to the participants. This model is designed to examine the principle interactions of the involved parties. In case of remote tower these parties could be air traffic control, airport authority or pilots.
- **Fast-Time Simulations (FTS)** are able to run numerous scenarios in a short period (cf. EUROCONTROL, 2010, p. 49). As such, FTS can determine statistic effects of large traffic samples. For RTO research, FTS is able to quantify the dependencies of the different traffic streams. The design and configuration of FTS is challenged by human performance issues. As human operators can only work in real time, their behavior needs to be modeled compatible to fast time.
- **Real-Time Simulation (RTS)** is in fact an imprecise term. E-OCVM defines: "Real-time simulation techniques are important in providing human-in-the-loop experience of a proposed concept" (cf. EUROCONTROL, 2010, p. 49). In this understanding, HITL simulation is the adequate term as this is a hyponym of RTS. HITL simulations integrate the operator into a reproduction of his or her working environment (cf. Rothrock & Narayanan, 2011). Working processes are executed in real-time and provide human performance data which can be aggregated to a feasibility assessment.
- **(Passive) Shadow mode trials** use the operational environment by inserting new systems and procedures (cf. EUROCONTROL, 2010, p. 49). In general, two operators work within shadow mode trials. One operator conducts the conventional process. Her or his decisions are transferred to the operational system. The second operator uses the new system or procedure. His decisions are not transferred into operations, but recorded and analyzed. In case of RTO research, one controller would monitor and manage the traffic from the conventional tower, while a second one uses a remote tower working position without issuing commands to the pilots.
- **Live trials** use the operational environment together with the new system or procedure (cf. EUROCONTROL, 2010, p. 49). In contrast to shadow mode, all

decisions made are transferred into operations. Fallback procedures to step back to the conventional system are most times in place. Using this method for RTO, one tower controller would manage the traffic from a remote tower working position, while a second tower controller is on standby in the conventional tower.

Concluding the methods defined by E-OCVM, two methods can be excluded for validation purposes. Gaming has only little to no relevance for RTO research as there is a high interest to maintain all interactions with pilots and other facilities (e.g. airport authorities) as they are (cf. Vereinigung Cockpit, p. 26; Krüger, 2019). Live trials are dedicated to V3 stages of E-OCVM and therefore will be conducted after the feasibility assessment is completed (V2 stage is finished).

FTS, HITL simulation and shadow mode trials all maintain some deficiencies which do not allow a comprehensive evaluation of RTO concepts. FTS is able to determine all traffic dependencies, but is limited by the accuracy of the used controller model. Making wrong assumptions on the controllers' abilities to manage traffic under RTO conditions will lead to wrong traffic dependencies. HITL simulations do not need a controller model as they put the operator into the loop, but are limited due to the real-time constraint. A comprehensive analysis of all possible traffic situations would take multiple years of simulation which is not executable, especially with the limited availability of the air traffic controllers. Shadow mode trials have the same deficiency as HITL simulations and moreover they own the disadvantage that the environment is not under full control. Rare und unfavorable events like distinct weather situations or system failures cannot be inserted as requested by the validation objectives. Moreover, shadow mode trials take a higher effort due to consideration of security issues, operational requirements on the utilized hardware and stronger restrictions on testing procedures in comparison to simulations. Following these conclusions, all three methods need to be combined to a comprehensive approach. This is done via the following steps:

1. FTS-HITL-Coupling: Transfer of results between FTS and HITL simulation.
2. HITL-Adaption: Adaption of the HITL simulation to the needs of RTO research.
3. HITL-Field-Transfer: Stepwise transfer of simulated concepts to shadow mode trials.

The approach towards these steps are presented within the following sections.

2.2   FTS-HITL-Coupling

The E-OCVM suggest FTS as the simulation tool to determine traffic effects:

> They [FTS] are best used to test the sensitivity of a proposed concept to different assumptions and scenarios. (*Source* EUROCONTROL, 2010, p. 49)

As discussed above, FTS design includes the challenge to model the human operator. Sophisticated simulation environments for human performance models are

available (e.g. Fowles-Winkler, 2003). Based on Wickens' multiple resources model (cf. Wickens & Hollands, 2000) and the human information processing/time pressure theory (cf. Hendy et al., 1997), initial tests were performed (cf. Mahmoudzadeh Vaziri & Fürstenau, 2014). However, the usage of this kind of performance modeling requires a valid input. Specific RTO human factors data needs to be gathered and specified in detail by a method which takes the operator into account. As new RTO concepts cannot be examined in live operations without compromising safety, HITL simulation is the only method which can deliver the necessary data. In consequence an optimization cycle between FTS, selecting relevant scenarios and HITL simulations, gathering the human performance data needs to be established. This process is designed as shown in Fig. 2 (cf. Schier et al., 2011).

Initially, scenario requirements are defined and modelled in an FTS using a basic controller model (Fig. 2, step 1). In a second step, traffic scenarios with a relevant influence of the controller are selected by analyzing the FTS data. After an adaption to the RTS needs (e.g. additional flightplan data) HITL-simulations is executed. Evaluating and comparing the HITL simulation data with the FTS data, optimizations of the FTS controller model can be derived. With this optimization, the FTS can be rerun and provides improved data on possible traffic dependencies in a large set of traffic situations. This cycle can be executed multiple times and thereby develops on one hand a detailed controller model and on the other hand a comprehensive set of traffic dependencies for RTO research.
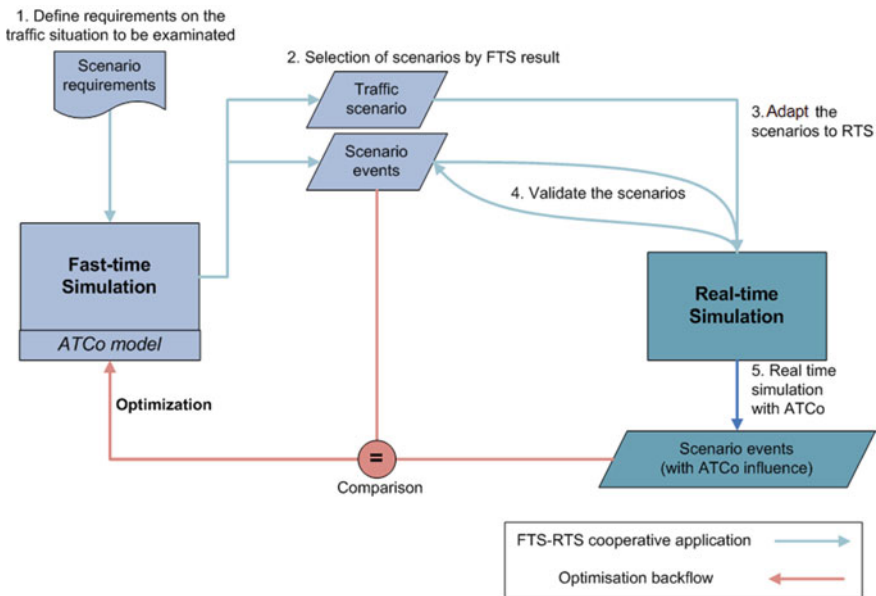

Fig. 2 Scenario definition process (cf. Schier et al., 2011)

2.3   HITL-Adoption

HITL simulation in general consist of five parts (cf. Rothrock & Narayanan, 2011, Table 1.1, p. 3). The simulation objects are the central data elements representing real-world items. The simulation dynamic modifies the objects dependent on time. The user interfaces allow the participants of the HITL simulation to modify the objects. Last but not least the configuration defines the initial state of the simulations objects and the resulting data is a collection of user interaction data and simulation object data. Figure 3 shows the common HITL simulation components.

RTO research requires a HITL tower simulation to assess feasibility. The German Aerospace Center (DLR) operates tower simulation facilities since the late 1990s to evaluate human factors aspects (cf. Kaltenhäuser, 2003). These facilities focus on aircraft and vehicles as their simulation objects. The simulation dynamic is based on the EUROCONTROL Total-energy model (cf. Nuic et al., 2010) and the user interfaces are mock-ups of the tower controller systems (e.g. flightstrips, weather information system, airside situational display). In comparison with other HITL simulations, tower simulations provide two special functionalities: The projection system and the pseudo pilots. The projection system emulates the out-of-the-window view. The pseudo pilots are simulation staff who interact with the air traffic controller comparable to real pilots. Therefore, they use a simulated radio system and a user interface to control multiple aircraft. Figure 4 shows the tower simulation components.
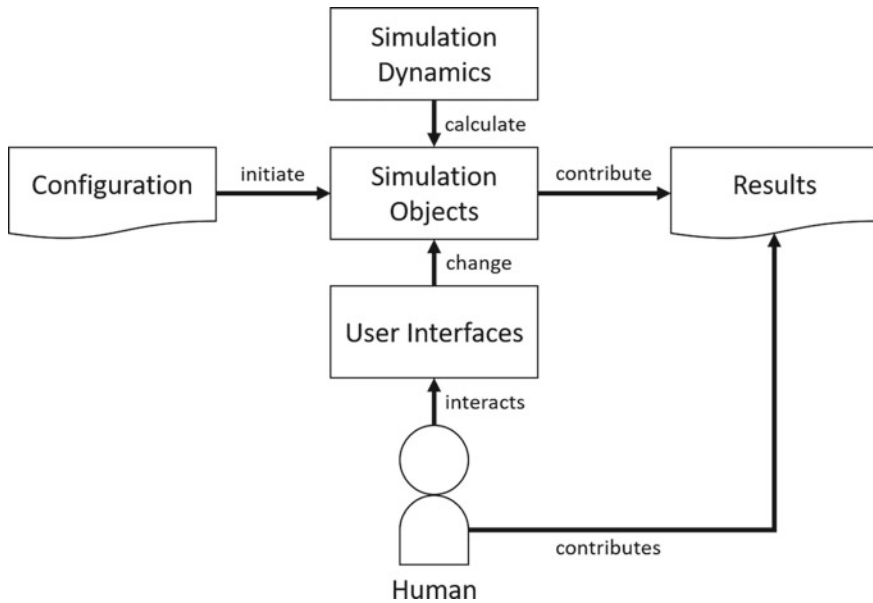

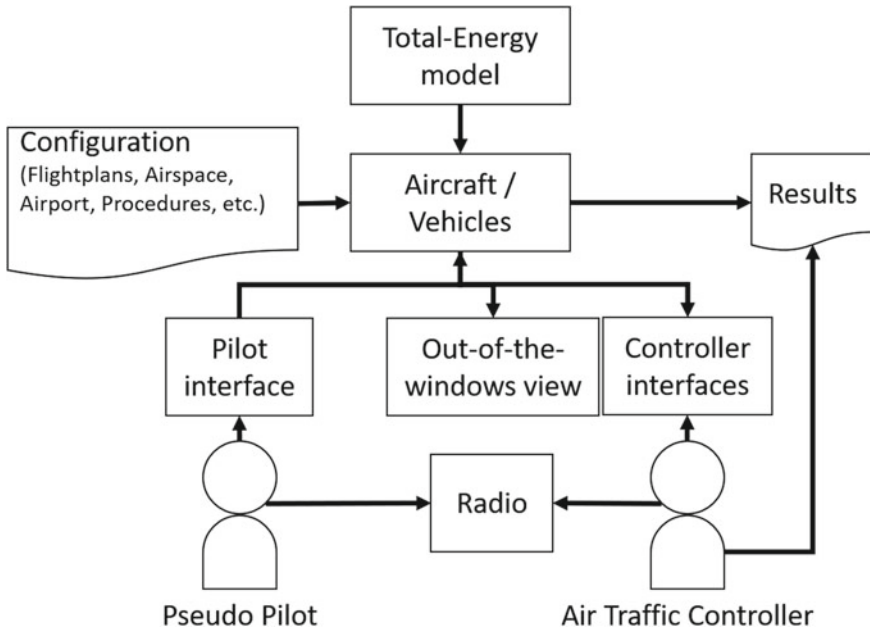**Fig. 3**   General components of a HITL simulation

Fig. 4 HITL tower simulation

Tower simulations are designed to validate new systems and procedures for conventional tower operations. RTO includes several aspects how conventional tower operations are changes: Workshare and responsibility can be flexibly organized; controller interfaces follow other design principles and the out-of-the-window view is provided by sensor systems. The validation platform must be able to represent those differences. To specify requirements for an RTO HITL simulation platform, a case study of an early DLR RTO project is performed. RAiCe (Remote Airport Traffic Control Center, cf. Fürstenau, 2011) was conducted between 2008 and 2012. It included two HITL simulation campaigns which used existing simulation equipment by adapting it to RTO needs.

The first RAiCe simulation experiment (RAiCe1, see chapter "Assessing Operational Validity of Remote Tower Control in High-Fidelity Simulation") examined the differences between work under remote tower conditions and work under conventional tower conditions. Twelve air traffic controllers were invited for this purpose. The simulation setup was based on the DLR Apron- and Tower Simulator (ATS) which offered a 200° projection system and the standard components for a HITL tower simulation as described above. For RAiCe1 the ATS was extended with a remote tower console. This console was designed within the project Remote Airport Tower Operation Research (RAPTOR, cf. Schmidt et al., 2007). It consisted of a wooden desk with integrated screens (e.g. for flight strips, camera control, radar, etc.) and a back projection system for the simulated remote out-of-windows view. Initially, this back-projection system was designed to be used with real cameras at

Fig. 5 RAiCe1 setup: remote tower console (left) and 200° projection system (right)

an airport for live testing purposes. For RAiCe1, the console was equipped with a software interface to display the simulated data (virtual out-of-the-window view as well as flightstrip and radar data). The simulation setup is shown in Fig. 5.

The data analysis of the successfully conducted simulations revealed some of the major challenges for remote tower control. Moreover, multiple findings regarding RTO HITL simulations were derived:

1. **360° projection**: Due to the limited panorama viewing angle of 200° effects of traffic pattern observation in the back of the controller could not be analyzed. In consequence a 360° projection system should be available for further simulations.
2. **Console integration**: The console was an inflexible wooden design which was adapted for a single screen size and specific systems. In consequence integrating tower tools like flightstrips and approach radar was a challenging and time-consuming task.
3. **Remote Tower video panorama**: The out-of-the-window view was provided by a back-projection system consisting of five UHD-projectors. This system was not adaptable to any changes of the working position or any requests for other projection size and resolution.

As a follow up of RAiCe1, a second simulation experiment (RAiCe2) was performed (cf. Moehlenbrink et al., 2010 and chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-in-the-Loop Simulations"). RAiCe2 focused on the effects of flexible airport assignment and work share among RTO air traffic controllers. Specifically, the simulation experiment addressed the question of how two controllers would organize their work under simultaneous control of two airports. The simulation setup was based on the remote tower console used in RAiCe1 (cf. Fig. 6). The back-projection system was replaced by two rows of five displays. Each row displayed a virtual out-of-the-window view of one airport. Additionally, approach radar, weather information and paper-based flight strips were integrated. The simulations were driven by two different applications due
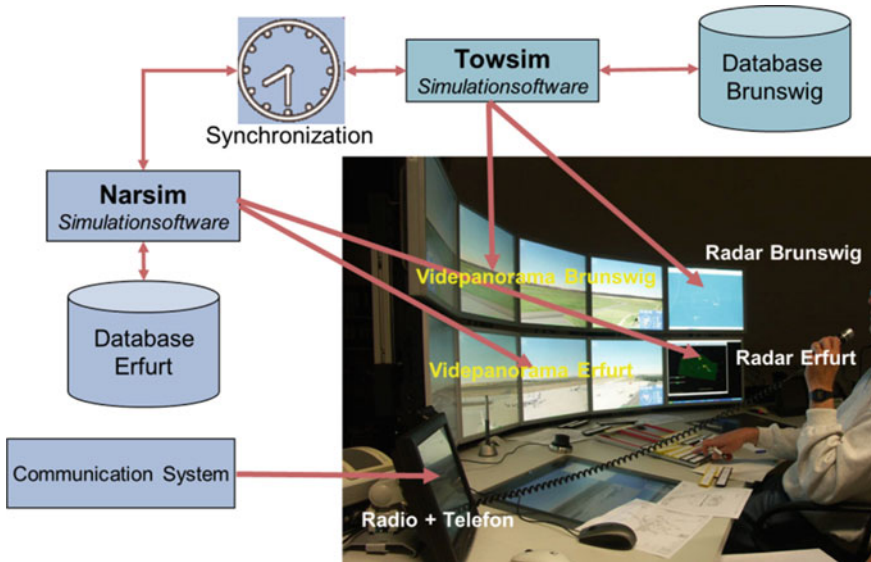
Fig. 6 RAiCe2 setup: remote tower console for Braunschweig and Erfurt

to a change of the simulation dynamics. The first airport (Braunschweig—EDVE) was available in an earlier software version (UFA's Towsim (cf. Sood et al., 2015). The second airport (Erfurt—EDDE) was modelled with the new NARSIM engine (cf. Have, 1993) which was dedicated to become the new simulation dynamic.

The simulation experiments were successfully conducted with twelve air traffic controllers working in six teams. Two major conclusions were drawn from this experiment:

1. **Console integration**: The wooden design impeded the integration process of the numerous tower tools and utilities. Screen sizes could not be adapted and only very limited space for the flight strips was available. Due to the fixed shape of the console, extensions were not possible.
2. **Synchronization**: Synchronization of the simulations and their recording appeared as a major issue. For analyzing parallel events and special situations, voice and aircraft data had to fit together with an accuracy of a second.

The findings of RAiCe1 and RAiCe2 were combined into a requirement list for the design a of a new tower simulation facility that supports the validation of RTO concepts. The requirement list is summarized as follows:

- **Flexible outside view**: The simulated outside view must be adaptable in size and setup to the RTO concept under validation. Projection angle, resolution and position in relation to the controller desk can change depending on the concept. A projection angle of 360° in maximum must be available.

- **Flexible controller desk**: The desk of the controller must be capable of containing different screens and tools in changing order. Thereby, the integration and flexible adaption to the RTO concept under validation is possible.
- **Synchronized data recording**: All recording systems must save their data synchronized with the simulation clock and writing timestamps with a minimum resolution of seconds to allow post-study analysis of parallel events.

2.4 HITL-Field-Transfer

Shadow mode trials utilize the operational environment and insert new systems or procedures (cf. EUROCONTROL, 2010, p. 49). The operational environment of conventional tower research is the tower building at the airport which is stationery. As such, shadow mode trials for conventional towers must be conducted inside or in close location to the tower building to provide comparable out-of-the-window view and an access to the operational data (e.g. flightplan, weather information, etc.). These requisites limit shadow mode trials in most cases regarding the available space and trial times. Moreover, an access to resources outside of the airport (e.g. simulation capabilities) is difficult or impossible due to security reasons.

In contrast to conventional towers, RTO is not stationary. The controller position can be located at any building with the necessary data access. Space and time limitations induced by the tower building are not relevant to RTO research. Resources apart from the airport and the tower building can be utilized. This optimization potential of the RTO shadow mode trials can be exploit via a pattern from computer science. Within computer science, integration of new functionalities into existing systems are a major research topic. Among others, so called stubs are patterns to be used within integration (cf. Meszaros, 2017, p. 529). Depending on the type of stub (e.g. responder, procedural stub, entity chain snipping), certain functions or data sources of the target system are emulated by the stub. Figure 7 shows an example of stub usage. The shown module one can be tested, using the stub. As soon as module two of the target system is available, the stub es replaced.

Applying this pattern to RTO shadow mode trials, allows validation activities even if some operational data sources are missing. For instance, a new RTO flightstrip system can be tested without setting up a camera system to generate the out-of-the-window view. Instead of the camera system, a virtual view or a recorded video can be used as a stub. The controller participating in the trial is provided with a complete

Fig. 7 Stub usage for testing purposes
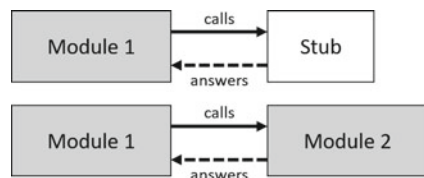
working position, but the effort to connect operational data sources is limited to the necessary sources.

Following the comprehensive approach of the RTO validation platform, the effort to build the stub is additionally reduced. The EOCVM approach suggests to conduct HITL simulations in advance of shadow mode trials. As such, the simulation functionality can be used as stubs in the trials. Figure 8 visualizes this approach applied on the RTO flightstrip system example.

Regarding the example of the RTO flightstrip system, HITL simulation with virtual out-of-the-window view and all controller interfaces (e.g. weather display, airside situational display) would be conducted in advance of the shadow mode trials. For the trials, the flightstrip system under test is disconnected from the simulation and connected to operational flightplan data. The operational flightstrip data is fed into the simulation to generate the outside view. This seamless transfer from HITL simulation to shadow mode trials needs to be considered in the design of the RTO validation platform. As such, two additional requirements are defined:

- **Simulation stub**: The validation platform must provide simulated data and interfaces which are not available from the operational environment in shadow mode trials.
- **Operational interface**: The simulation engine of the validation platform must be able to calculate simulation objects based on operational data.
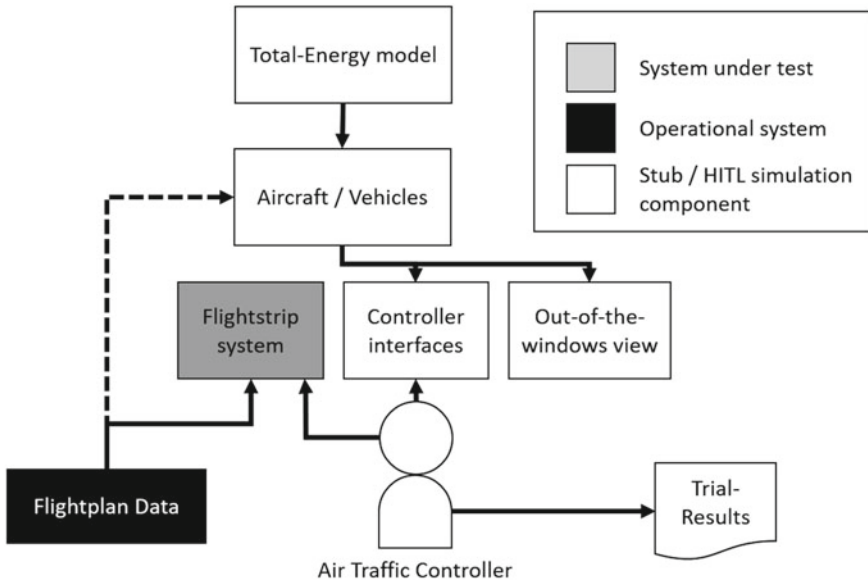

Fig. 8 Stub example of the RTO flightstrip system

2.5 Specification

Concluding the method part, a specification of a comprehensive validation platform for RTO research is derived. For this specification, each process and prerequisite, defined in the former method sections, is translated into a requirement with a distinct identifier for the RTO validation platform.

Following the E-OCVM, the validation platform consists of the core components FTS, HITL simulation and shadow mode trial environment (cf. Sect. 2.1). FTS and HITL simulation must be components in the validation platform (cf. Table 1, RE1 and RE2). Both components are connected via a process in two directions (cf. Sect. 2.2). The transfer from FTS to HITL simulation requires a controller model inside the FTS (cf. Table 1, RE3), an adaptable FTS analysis (cf. Table 1, RE3) and a traffic scenario translation for the HILT simulation (cf. Table 1, RE4). The transfer from HITL simulation to FTS requires a HITL simulation analysis (cf. Table 1, RE5) and the controller model.

The HITL simulation itself needs some adaptions to encounter all aspects of RTO research (cf. Sect. 2.3). As such a 360° out-of-the-window view is necessary (cf. Table 1, RE6). The view as well as the controllers' desks need to own the flexibility for adaptions (cf. Table 1, RE7 and RE8). Last but not least, data recording needs to be synchronized (cf. Table 1, RE9).

Table 1 Requirements of the RTO validation platform

ID	Name	Description
RE1	FTS simulation	The platform must contain a fast time simulation
RE2	HITL simulation	The platform must contain a HITL simulation
RE3	Adaptable controller model	The FTS must contain an adaptable controller model
RE3	FTS analysis	The platform must provide an adaptable FTS analysis
RE4	Scenario translation	The platform must be able to generate a HITL scenario based on the FTS data
RE5	HITL analysis	The HITL simulation must contain an adaptable controller model
RE6	360° projection	The HITL simulation must provide a projection system with a display angle of 360°
RE7	Flexible projection	The HITL simulation must provide projection system which is adaptable in angle, resolution and position
RE8	Flexible controller desk	The HITL simulation must provide a controller desks which can contain different screens and tools in changing order
RE9	Synchronized recording	The HITL simulation must record all data synchronized to one clock with a resolution of one second
RE10	Operational interface	The HITL simulation must be able to accept operational data as input
RE11	Simulation stub	The HITL simulation must be able to generate data based on operational data

The transfer from HITL simulation to shadow mode trials is enabled via a stepwise approach using the simulation components as stubs (cf. Sect. 2.4). This requires the HITL simulation components to accept operational input (cf. Table 1, RE10). Moreover, components that are not fed with operational data must be able to complete the given data based on their simulation functionality (cf. Table 1, RE11). The requirements are summarized in Table 1.

3 Implementation of the RTO Validation Platform

The development of the remote tower simulation platform was initiated in 2009. As a first implementation step, an FTS and a HITL simulation software had to be selected as being the core components. To satisfy the requirements RE1 (FTS) and RE3 (controller model), AirTOp by Airtopsoft SA (cf. Airtopsoft, 2021) was chosen (cf. Walther, 2010). AirTOp is a broadly used FTS tool, providing airport and airspace modeling and simulation. As a HITL simulation (requirement RE2) NARSIM by Netherlands Aerospace Centre was selected (cf. Have, 1993). NARSIM is specially designed as a research simulation engine. As such, any parameter can be recorded and new functionalities can be integrated by the users.

The data analysis and scenario transfer from FTS to HITL simulation was initially based on a self-developed DLR application. The Extensible Workflow Management For Simulations (EWMS, cf. Scharnweber & Schier, 2009) and its successor EWMS2 (cf. Schier & Morlang, 2017) provide standard calculation steps for aeronautical data as well as imports and exports to chosen simulation tools. The content of EWMS and EWMS2 was transferred to KNIME (cf. Berthold et al., 2009) in 2018 which is a commercial of the shelf software. The data transfer from NARSIM to the AirTOp controller model was integrated into EWMS/EWMS2/KNIME data analysis procedures as well.

Finally, the operational data sources were connected. The following connections have been established until 2020:

- Flightplan data: EUROCONTROL data services
- Radio: Hardware radio receivers and audio streams
- Aircraft position: OpenSky ADSB service (cf. Schäfer et al., 2014)
- Video data: Resource for video stream transfer.

Figure 9 shows the resulting setup.

The hardware setup of the DLR Apron- and Tower Simulator was completely reconstructed based on the specification. Within the specification and design process two general trends were identified. One trend demands for 360° projection system (requirement RE6) and a most realistic environment. The second trend aims at a most flexible environment where out-of-the-window view and controller desks can easily be adopted (requirements RE7 and RE8). Several findings within the RAiCe2 experiment supported these trends (cf. Schier et al., 2013). Within reconstruction, it became clear that both trends could not be satisfied by one single facility. Consequently, a
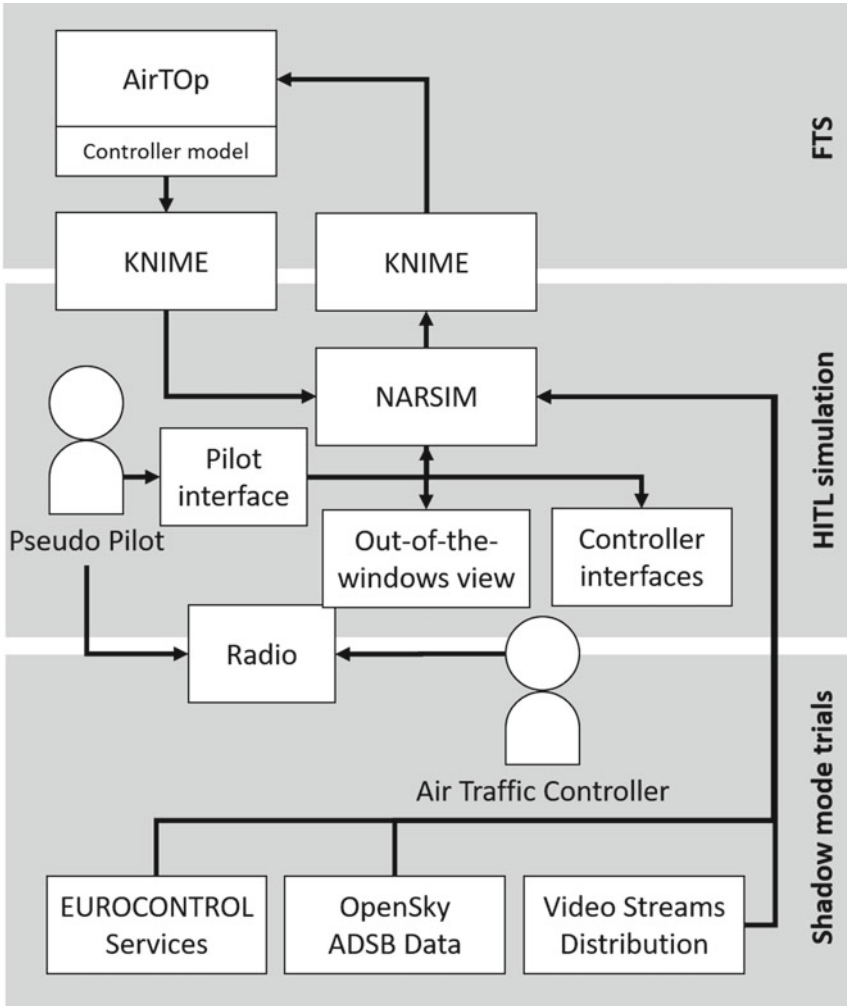
Fig. 9 The RTO validation platform

new 360°-system was realized (ATS360) offering maximum realism with regard to the simulated environment and scenarios (cf. Fig. 11). The ATS360 includes a 360° high resolution projection system and mockups of operational systems such as the DFS electronic flight strip system (Tower Flight Data Processing System—TFDPS) and the weather and information system (information data handling system—IDVS). Thereby requirement RE6 was satisfied (Fig. 10).

Additionally, a flexible facility called TowerLab was designed (cf. Fig. 11). To provide an adaptable outside view system (requirement RE7), allowing for modifications with regard to size, resolution and alignment a loose connection of monitors was constructed. This construction can be changed within minutes if other setups are

Fig. 10 ATS360: the close-to-reality apron and tower simulator of DLR
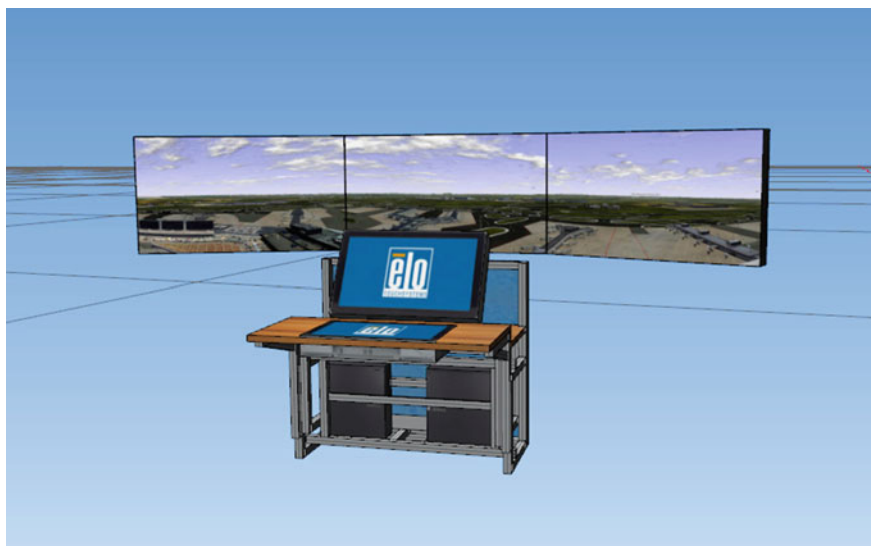
Fig. 11 TowerLab: design drawing of MoToKo with outsideview system

needed. Moreover, a generic console, called MoToKo (modular tower console) was designed. This console encounters all experiences made in RTO research so fare. Fixed display cut-outs, predetermining the screen size were avoided (requirement RE8). The console is extensible and adaptable by a standard construction kit and allows for easy movement to other positions.

TowerLab and ATS360 were used within numerous simulation experiments (some of them described in the present Part III of the book), from research on advanced surface management systems (cf. Carstengerdes et al., 2013) to analysis of tower controller visual sequences (cf. Manske & Schier, 2015). In the following sections the application of the comprehensive remote tower validation platform is shown, providing multiple case studies.

4   Results of the MRT Simulation Platform Application

4.1   Case Study: Application of FTS-HITL-Coupling

The first application of the FTS-HITL-Coupling took place in the project RAiCe2 (cf. Moehlenbrink et al., 2010). Possible options of sharing task load in remote tower setup with one and two controllers controlling simultaneously two airports were examined. Therefore, a scenario with numerous parallel events that challenge the controllers were required. As such, challenging scenarios should be identified by FTS and then transferred into HITL simulation. The scenario retrieval process was started by defining a rough controller model (cf. Walther, 2010). The model assumed that tower controllers' performance is limited to the fact that one command per time can be given. As such controlling two airports remotely includes the possibility that upon parallel events on both airports, one event is delayed. This happens for instance if aircrafts want to takeoff on both airports at the same time (cf. Fig. 12). The controller is only able to give one takeoff clearance at a time instant. Following this constraint, one takeoff needs to be delayed.

Situations like these could be determined within numerous fast time simulations runs using the rough controller model that implements only the radio communication behavior. Depending on the number of parallel events, the scenario for the HITL simulation was chosen and transferred via EWMS.

One metric to quantifying the quality of scenarios within HITL is the accuracy by which traffic events are modelled. In the environment of tower simulations, traffic events are defined as clearly identifiable actions of an aircraft (e.g. landing, takeoff, etc.). The accuracy refers to their occurrence in time. As such, the FTS-HITL-Coupling quality can be evaluated by the time difference between modeled event occurrence in the FTS and in HITL simulations.

As a result, it was determined that 123 of the defined events took place within a time difference of 60 s which was defined as the acceptable threshold (cf. Fig. 13). With about three events per scenario (cf. Schier et al., 2011) this was accepted as
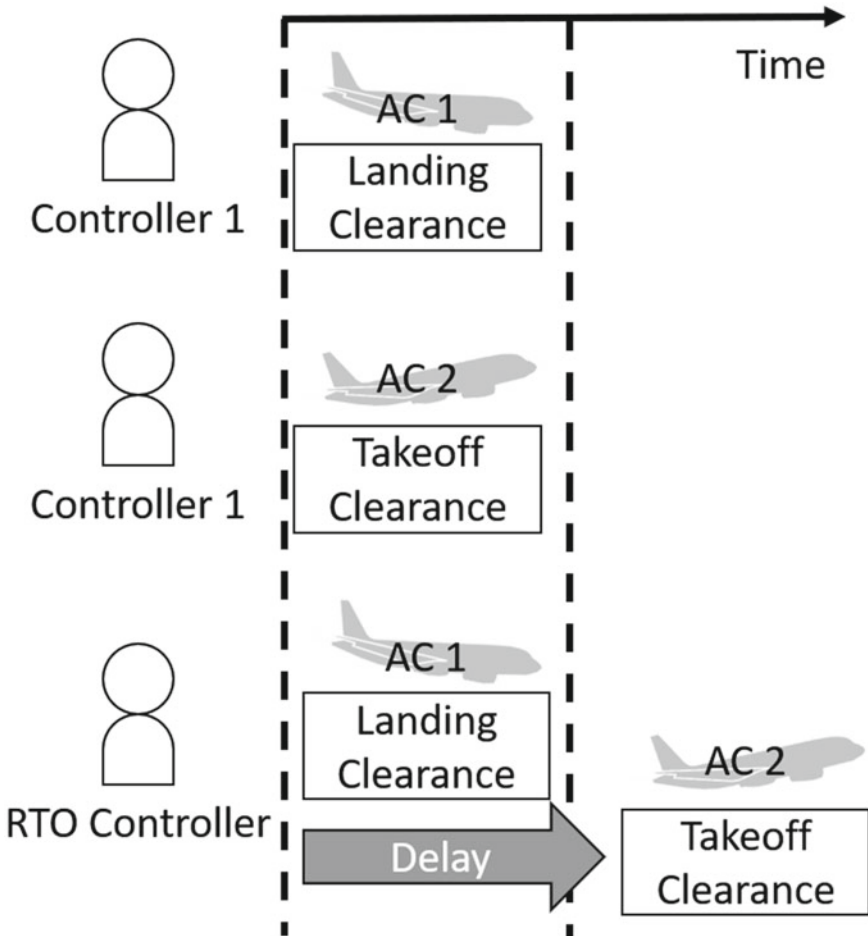
Fig. 12 Challenges of parallel events in remote tower operations

sufficient for the conducted 38 HITL trials. The HITL simulations were analyzed regarding the duration of different commands issued via radio. The average duration was transferred back to the FTS controller model. By this process the FTS quality could be improved for further scenario retrievals.

4.2   Case Study: Remote Tower Human Factors Study

The "Remote Tower Human Factors Study" (RTC-HFS) as a contracted research for the Deutsche Flugsicherung GmbH (DFS) was the initial use of the TowerLab (cf.

Histrogramm showing time difference of traffic situations and number of occurence
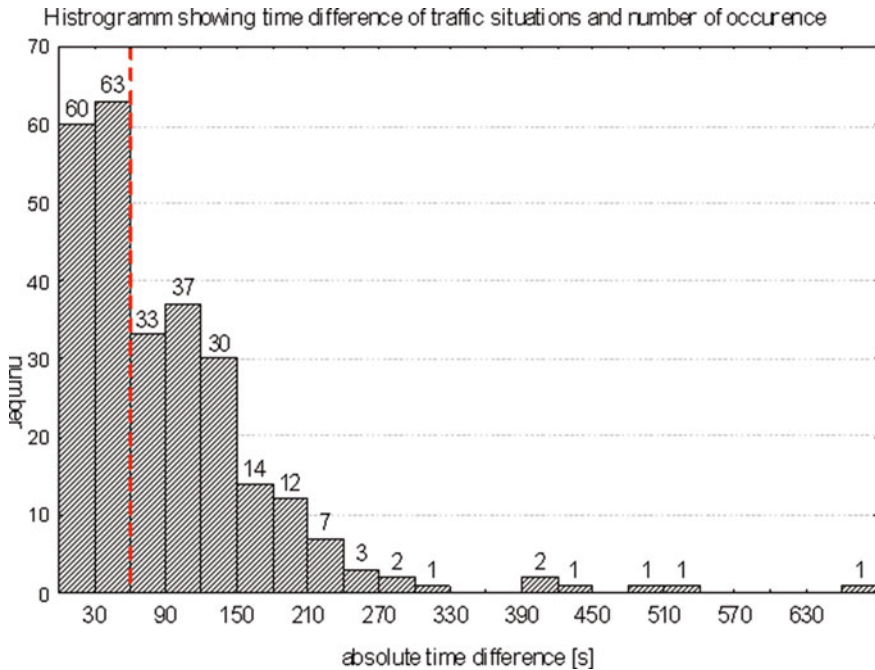
Fig. 13 Histogram of time differences of traffic situations and number of occurrences (cf. Schier et al., 2011)

Papenfuss et al., 2012). The objective was to analyze task load limits of controllers in remote tower controller working position (CWP).

To a large extent the simulation design was driven by requirements defined by DFS for remote tower CWP. As such a specific screen size and resolution had to be provided for the simulated out-of-windows view. Additionally, DFS looka-like controller tools were requested. The flexible TowerLab infrastructure including the display wall was adapted to these needs. A MoToKo console was configured according to the requirements of DFS CWP, including TFDPS mockup (the DFS flight strip system), DFS weather display and approach radar (cf. Fig. 14).

Six air traffic controllers took part within this successful first TowerLab simulation campaign (cf. Papenfuss et al., 2012). Besides primary data collection, feedback of the air traffic controller was requested after the simulation. The TowerLab design was accepted and additionally proved its advantages by the very short construction and adaption phase. The comments provided by the domain experts primarily focused on the software:

(1) **Synchronization**: Further improvement of data synchronization between simulation engine and all connected systems (e.g. flight strip and radio emulation system) appeared desirable.

Fig. 14 RTC-HFS: remote tower controller working position (CWP-) design

(2) **Helicopter model**: The situations displayed within this campaign included several helicopter missions. The used model was a very rough one adapting fixed wing models. For further campaigns it was suggested to implement an improved helicopter model.

The integration of these optimizations was initiated for the following simulation campaigns.

4.3 Case Study: Remote Tower Center (RTC) Study

As second part of the RTC human factors study an investigation in the remote tower center concept was performed (cf. Papenfuss et al., 2012). This campaign focused on the possibility to set up a center for controlling multiple airports remotely from a single location. Within this center it is possible for the controllers to work one shift at one airport and change to another airport on the next shift. It was the goal to investigate constraints and challenges of these airport changes.

The simulation design made use of the experiences gained in the first RTC-HFS campaign and multiplied the designed working positions for three airports (Dresden—EDDC, Erfurt—EDDE and Braunschweig—EDVE) as shown in Fig. 15. Additionally, a CWP for a ground coordinator was designed. The ground coordinator

Fig. 15 Remote tower center (RTC) design for controlling three airports from afar. Foreground: ground coordinator; back: three airport CWP's for executive controllers

is in charge of coordinating startup and air route clearances as well as initiating flight-plan for flights following visual flight rules.[1] In contrast to the executive controllers, the ground coordinator was not in need of an outside view, but advanced flightplan systems and communication devices had to be provided.

The simulations were successfully conducted with 12 air traffic controllers shifting through the different working positions. Collected comments of the controllers and the ATM experts showed significant improvement on the data synchronization requested within the remote tower human factors campaign. Again, the missing helicopter model was addressed.

4.4   Case Study: Multi Remote Tower Study

Another RTC simulation campaign was conducted within the European SESAR (Single European Sky ATM Research) framework. In a DFS—DLR cooperation a Multi Remote Tower (MRT) concept was addressed. Air traffic controllers were given a high amount of traffic at a single remote tower working position, observing one airport. The results of this condition were compared with a multi remote tower condition: The same amount of traffic was distributed between two airports and controlled by the operator in parallel (cf. Moehlenbrink & Papenfuss, 2014).

The simulation design combined the experiences made in RAiCe2 with multi remote tower working positions and the latest results of the remote tower center study (cf. Fig. 16). The single working position was slightly improved as compared to the center study by an advanced camera control and weather display. The multi remote tower CWP was built on the same basis, but encountered several additional challenges. Especially the requirement to not have displays of radar or flight strips within

---

[1] In Germany, flights following visual flight rules are not under duty to file a flightplan.

Fig. 16 Multi remote tower setup (left: single working position, right: multi remote working Position), front eye-tracking device and exercise lead working position

the line of sight to the outside view had to be solved by the construction process. The simulated out-of-window view in this case was based on the latest video panorama prototype version (using HD-technology, see chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation").

A final design for the multi remote tower console was retrieved and the simulation with 20 controllers was successfully finished. Synchronization process between simulation, voice system, flight strips and eye tracking system could be established.

4.5 Case Study: SESAR2020 Shadow Mode Trials

The TowerLab was chosen as a platform for shadow mode trials in SESAR2020. Among other objectives, SESAR2020 aims on evaluating new camera and tracking technique as well as controller interfaces for RTO. The systems under test are provided by industrial partners. Within the planned trials, air traffic controllers will work in shadow mode using the new camera and tracking technique. Afterwards, they will give their feedback on the new functionalities.

A special challenge of these trials is the dependency of all systems under test. The tracking algorithms need to have the video stream of the cameras. The cameras need to be connected to their control interface. And the control interface requires access
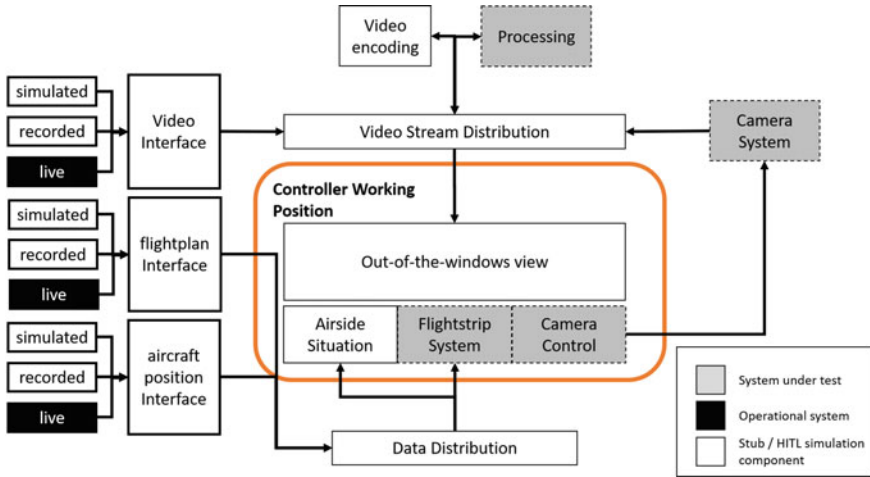
Fig. 17 TowerLab setup for SESAR2020 shadow mode trials

to the tracking functionality. All systems need to be tested thoroughly, but will not be in place at the same time. Moreover, the controllers will need additional data (e.g. airside situation, weather data, flightplans) which enables them to develop a mental picture of the situation.

Following the stub-method suggested for shadow mode trials (cf. Sect. 2.4) and the connections to operational data sources inside TowerLab, dissipates these dependencies. Figure 17 shows the resulting setup.

The out-of-the-window view is connected to a video stream distribution system. This system is initially fed by the TowerLab sources. This can be a simulated, virtual video stream for cases when all conditions (e.g. weather) should be under control. If video data with operational quality is required, recorded videos or the DLR camera system can be used. As soon as the camera system under test is in place and connected, this data will be fed into the video distribution system.

Additionally, a data distribution system is in place to provide flightplan and aircraft position data. As this data is not generated by the systems under test, TowerLab sources will be used. This can either be simulated flights (with flightplan and calculated position) or real data (recorded or live). Thereby, the flightstrip system and the camera control, which are part of the system under test, can be used with their full functionality. For instance, the tracking functionality can be activated by selecting a flight in the flightstrip system.

The SESAR2020 shadow mode trials are planned in November 2021. Integration work is ongoing in February 2021, enabling the stage where DLR camera system feds the video distribution, while the camera system under test is integrated.

5 Conclusion and Outlook

In this chapter, a summary of activities conducted to develop a comprehensive RTO validation platform is given. With this validation platform, various studies and developments with different research questions could be conducted. Additionally, with regards to the simulation methodology, a move was made from monolithically simulators representing one specific workplace to a flexible, coupled simulation environment which serves many different approaches and a seamless transit from one to another.

RTO is a complex research topic as it induces multiple changes to the air traffic system. Traffic which used to be independent is dependent under RTO conditions. Controllers have their responsibilities shared across different airports and are able to balance their workload. Last but not least new CWP and sensor techniques have to be designed and evaluated. These aspects of research generate new requirements to the validation platforms. Following E-OCVM, FTS, HITL simulation and shadow mode trials are the primary methods to conduct these validations. For a comprehensive research, FTS-HITL-Coupling is necessary which allows a scenario selection via FTS and detailed human factors analysis via HITL simulation following an optimization cycle for both tools. Additionally, HITL simulations need to fulfill requirements for 360°, flexible projection system as well as adaptable out-of-the-window view and CWPs. Last but not least, a HITL-Field-Transfer is necessary that leads seamlessly from HITL simulation to shadow mode trials. This transfer is achieved via the stub-pattern from computer science integration and testing methods.

The method of a comprehensive RTO validation platform which allows evaluation from FTS to shadow mode trials is implemented in DLR's simulation platform. AirTOp, NARSIM and operational interfaces are connected to an innovative and powerful platform which has proven their capabilities in various projects of RTO research. Smaller deficiencies encountered in these projects have been fixed already. For instance, the requested helicopter model was developed (cf. Janssen, 2014) and an eye tracking system has been integrated (cf. Manske & Schier, 2015), forcing synchronization of data recordings to a resolution of milliseconds.

As basic RTO concepts have been defined and feasibility is under assessment, research is getting more into detail and thereby requires a higher level of realism. While first studies addressed very basic principles of work organization of two controllers, research is now focusing on large remote tower centers with up to 15 airports or MRT working positions with three or more airports. Beside the general task to monitor and guide the aircrafts, additional controller tasks (e.g. tactical planning, telephone calls, coordination work) and special situations (e.g. construction work, emergencies, adverse weather) are under evaluation. This higher level of realism and details has a new demand for the validation platform: Events and interactions which are not initially affected by RTO should now be available for validation purposes. Specifically, additional interaction partners such as approach control, remoter tower center supervisor, airport staff and pilots (e.g. in case of emergency) should be available. These additional interaction partners have not been modelled either in FTS

or in HITL simulation. In HITL simulation they can be integrated by additional staff, but this leads in some cases to unwanted training and group effects (cf. Schier et al., 2017). As a consequence, virtual agents show a high potential for RTO validation purposes as they can take certain roles and follow distinct rules of behavior. Thereby an unlimited set of interaction partners can be integrated without losing the comparability in repeated simulation runs. This concept has been proven for airport management simulations (cf. Schier et al., 2018) where a certain number of interaction partners is required. For the future this concept should be transferred to the RTO validation platform to allow research on the collaborative aspects of RTO.

References

Airtopsoft SA. (2021). *Airside aircraft movements*. https://airtopsoft.com/airside-aircraft-movements/, February.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., & Wiswedel, B. (2009). *KNIME-the Konstanz information miner: version 2.0 and beyond*. AcM SIGKDD explorations Newsletter 11.1, ISSN 1931-0145.

Carstengerdes, N., Schaper, M., Schier, S., Metz, I., Hasselberg, A., Gerdes, I. (2013). Controller support for time-based surface management—First results from a feasibility workshop. In *Proceedings of the 3rd SESAR Innovation Days (2013)*. EUROCONTROL. The Third SESAR Innovation Days, 26.-28. Nov. 2013, Stockholm, Sweden. ISBN 978-2-87497-074-0.

Fowles-Winkler, A. M. (2003). Modelling with the integrated performance modelling environment (IPME). In *Proceedings of 15th European Simulation Symposium*. SCS European Council, ISBN 3-936150-29-X.

Freese, M., Lukosch, H., Wegener, J., & König, A. (2020). Serious games as research instruments— Do's and don'ts from a cross-case-analysis in transportation. In *European Journal of Transport and Infrastructure Research*, Delft University of Technology, ISSN 1567-7141.

Fürstenau, N. (2011). BWMMS session 3a, b: Steps towards the Remote Tower Center. VDI Verlag. In *9th Berlin Human-Machine Systems Workshop*, 5.Okt.—7.Okt. 2011, Berlin. ISBN 978-3-18-303322-5. ISSN 1439-958X.

Hagl, M., Friedrich, M., Jakobi, J., Schier-Morgenthal, S., & Stockdale, C. (2019). Impact of simultaneous movements on the perception of safety, workload and task difficulty in a multiple remote tower environment. In *2019 IEEE Aerospace Conference*, ISSN 00002004, March.

ten Have, J. M. (1993). The development of the NLR ATC research simulator (Narsim): "Design philosophy and potential for ATM research". *Simulation Practice and Theory, 1*(1), 31–39. ISSN 0928-4869, 15 July 1993. https://doi.org/10.1016/0928-4869(93)90009-F

Hendy, K. C., Jianquiao, L., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors, 39*(1), 30–47.

EUROCONTROL. (2010). *European Operational Concept Validation Methodology (E-OCVM)*. European Organization for the Safety of Air Navigation, Version 3.0 (Vol. I), February.

Janssen, E. (2014). Analyse möglicher Realisierungen einer Drehflüglerdarstellung im Totale-Energie-Modell für Luftverkehrssimulationen. Masterarbeit, DLR-IB 112-2014/7.

Josefsson, B., Polishchuk, T., Polishchuk, V., & C. Schmidt, C. (2019). Step toward remote tower center deployment: Optimizing staff schedules. *Journal of Air Transportation, 27*(3). ISSN 2380-9450.

Kaltenhäuser, S. (2003, July). Tower and airport simulation: Flexibility as a premise for successful research. *Simulation Modelling Practice and Theory, 11*(3–4). ISSN 1569-190X.

Krüger, H.-J. (2019). der flugleiter. Gewerkschaft der Flugsicherungen e.V., Ausgabe 2/2019, ISSN 0015-4563.

Mahmoudzadeh Vaziri, G., & Fürstenau, N. (2014). *Integrated performance modeling environment: User guide*. DLR IB-112-2014/5.

Manske, P., & Schier, S. (2015). Visual scanning in an air traffic control tower—a simulation study. In *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015)*, Las Vegas, USA, July.

Meszaros, G. (2017). xUnit test patterns: refactoring test code (1st Ed.). Addison Wesley. ISBN 978-0131495050, May.

Moehlenbrink, C., & Papenfuss, A. (2014). *Eye-data metrics to characterize tower controllers' visual attention in a Multipe Remote Tower Exercise*. ICRAT 2014, Istanbul, Turkey, May.

Moehlenbrink, C., Friedrich, M., Papenfuss, A., & Jipp, M. (2010). Monitoring behaviour of tower controllers. In *Human centred automation*, Seiten 269–283. Shaker Publishing. HFES 2010, Berlin. ISBN 978-90-423-0406-2.

Nuic, A., Poles, D., & Mouillet, V. (2010). BADA: An advanced aircraft performance model for present and future ATM systems. *International Journal of Adaptive Control and Signal Processing, 24*(10). ISSN 1099–1115.

Papenfuss, A., Friedrich, M., Schier, S., & Jakobi, J. (2012). Human Factors Studie zu Remote Tower Control: Ergebnis- und Empfehlungsbericht. Projectreport, DLR-IB 112-2012/34.

Papenfuss, A., Reuschling, F., Jakobi, J., Rambau, T., Michaelsen, E., & Scherer-Negenborn, N. (2020). Designing a fusion of visible and infra-red camera streams for remote tower operations. In *2020 IEEE Aerospace Conference*, ISSN 00002004, March.

Rothrock, L., & Narayanan, S. (2011). *Human-in-the-loop simulations*. Springer, ISBN 978-1-4471-6017-5.

Schäfer, M., Strohmeier, M., Lenders, V., Martinovic, I., & Wilhelm, M. (2014). Bringing up OpenSky: A large-scale ADS-B sensor network for research. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. ISBN 9781479931484, April.

Scharnweber, A., & Schier, S. (2009). EWMS—Extensible workflow management for simulations. In *CEAS European Air & Space Conference*, Manchester, United Kingdom, October.

Schier, S., & Morlang, F. (2017). *Asking the experts—A flexible approach towards simulation based Training and validation data analysis*. Workshop Modellierung und Simulation in der Luftfahrt, Braunschweig, Germany, May.

Schier, S., Walther, J., Papenfuss, A., Möhlenbrink, C., & Lorenz, L. (2011). An approach to support controller workplace design in a multi airport-environment using fast and real time simulations. 60. Deutscher Luft- und Raumfahrt Kongress 2011, 27.-29. Sept. 2011, Bremen.

Schier, S., Rambau, T., Timmermann, F., Metz, I., Stelkens-Kobsch, T. (2013). Designing the tower control research environment of the future. Deutscher Luft- und Raumfahrtkongress 2013, 10.–12. Sept. 2013, Stuttgart, Germany.

Schier, S., Rosenau, L., & Freese, M. (2017). Trainieren mit CLAUDI – Das Potential kognitiver Agenten für Simulationen und Planspiele 6. Interdisziplinärer Workshop Kognitive Systeme, München, Deutschland, März.

Schier, S., Duensing, A., & Britze, S. (2018). Sparring partners for human-in-the-loop simulations: The potential of virtual agents in air traffic simulations. *CEAS Aeronautical Journal*, Springer.

Schmidt, M., Rudolph, M., Werther, B., & Fürstenau, N. (2006). Remote airport tower operation with augmented vision video panorama HMI. In *Proceedings 2nd International Conference on Research in Air Traffic Management (ICRAT 2006)*, DS Public, ISBN 86-7395-210-7, Belgrade, June.

Schmidt, M., Rudolph, M., Werther, B., Fürstenau, N. (2007). Development of an augmented vision videopanorama human-machine interface for remote airport tower operation. In *Human Interface and the Management of Information*, Part II, HCII 2007, LNCS 4558, Seiten 1119-1128. Springer. HCI 2007, 22.-27.7.2007, Beijing, China. ISBN 978-3-540-73353-9.

Sood, R., Wolff, D., Neubert, A., Seitz, A. (2015). *UFA ATTower—An air traffic Procut*. UFA Inc., information brochure. http://www.ufainc.com/products/attowertowsim-tower-simulation/, January.

Vereinigung Cockpit. Flugsicherheitskonzept SafeSKY 2020—Eine Perspektive aus dem Flight Deck. Aus dem Internet. https://www.vcockpit.de/fileadmin/dokumente/themen/sichererhimmel/VC-Flugsicherheitskonzept_2020. Aktuell am: 23.02.2020.

Walther, J. (2010). Implementierung eines Prozessmodells zur simulativen Bewertung der Towerlotsen Arbeitsbelastung bei zeitgleicher Ausübung von Kontrollfunktionen an zwei Flugplätzen. DLR-Internal Report [DLR-IB 112-2010/31].

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance.* Prentice-Hall.

Assessing Operational Validity of Remote Tower Control in High-Fidelity Simulation

Anne Papenfuss and Christoph Moehlenbrink

Abstract In this chapter results from simulation studies are presented which were conducted to assess the operational validity of the remote tower concept at a very early maturity level. The goal was to gain empirical evidence to lead further developmental activities, and to learn about critical design issues and human factors of the remote tower control concept. A high-fidelity simulation study with a sample of twelve tower controllers was conducted to assess operational validity of an experimental workplace for remote tower control (RTC). This set-up was compared to a simulation workplace representing a conventional tower. The core of the experimental RTC workplace is a panoramic display, presenting high resolution video data of the remotely controlled airport. Besides the feasibility of the concept, the study addressed the relevance of the view outside the tower window for air traffic controllers decision making and the benefit of information augmentation. Two functionalities were tested, being highlighting aircraft based on automatic movement detection as well as overlay of aircraft call signs. Eye tracking, questionnaire, and interview data were gathered. Results indicate that the concept is valid for control of smaller airports with little air traffic. The augmentation of call signs onto the video panorama reduced head-down times for the radar display.

Keywords Simulation · Validation · Eyetracking · Questionnaire

1 Introduction

1.1 Motivation

Remote control of smaller airports with little traffic is a concept for future air traffic control. It has the potential to reduce the costs of providing air traffic control to these airports. However, the concept implicates significant changes regarding the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
N. Fürstenau (ed.), *Virtual and Remote Control Tower*, Research Topics in Aerospace,
https://doi.org/10.1007/978-3-030-93650-1_12

controllers[1] work environment, as the controller shall move from the tower to another location, up to hundreds of kilometres away. One key challenge is the substitution of the view out of the tower window (far view). Yet, controllers can be assisted by replacing the far view with a panoramic video transmission which further allows for novel automated support (Fürstenau et al., 2008; van Schaik et al., 2010).

One objective within this context of developing a replacement of the far view relates to "what" information in general, and specifically of the far view, an air traffic controller uses, in order to guarantee safe operations on the (remotely controlled) airport. Until now there is no conclusive evidence which information tower controllers specifically use from the view outside to control air traffic, and which information is not used. Nevertheless, for smaller airfields with little sensor technology the view outside certainly is the most comprehensive source of visual information. From the human factors perspective it is of interest to understand, in how far the substitution of the far view causes changes on the tower controllers working methods.

For this reason, operational validity of this novel concept means that the controller sees everything he needs to. The function of and the information derived by the far view, therefore, is a central topic in this study. In addition, another focus is set on the examination of the utility of controller assistance at the remote working position through information superimposition and a zoom camera with a semi-automatic tracking function. A high-fidelity simulation set up was chosen, to represent the essential changes for the work environment and to assess its impact on air traffic control operations (see also chapter "Multiple Remote Tower Simulation Environment (S. Schier)").

1.2 Related Work

From 2002 to 2004 the concept study "Virtual Tower" was conducted at German Aerospace Center (DLR) that initialised research concerned with remote control of small airports (Fürstenau et al., 2004a). Within the project RApTOr (Remote Airport Tower Operation research), first steps of the idea were realized, and an experimental remote tower operation controller working position (RTO-CWP) was developed (Fürstenau et al., 2008; Schmidt et al., 2007). Based on a task analysis it was concluded that the far view is a crucial information source. It is not clearly determined into detail what parts of the view outside controllers use for their decisions and what information stays unused. Therefore, the core of the RTO-CWP is a reconstruction of the far view by means of a live stream of high resolution videos and an additional zoom camera. For demonstration and evaluation of the technical concept, an experimental system was set up at the research airport of Braunschweig-Wolfsburg (EDVE). For a more detailed technical description of the camera system

---

[1] In this chapter the term controller refers to both male and female operators. For ease of reading the male personal pronoun will be used.
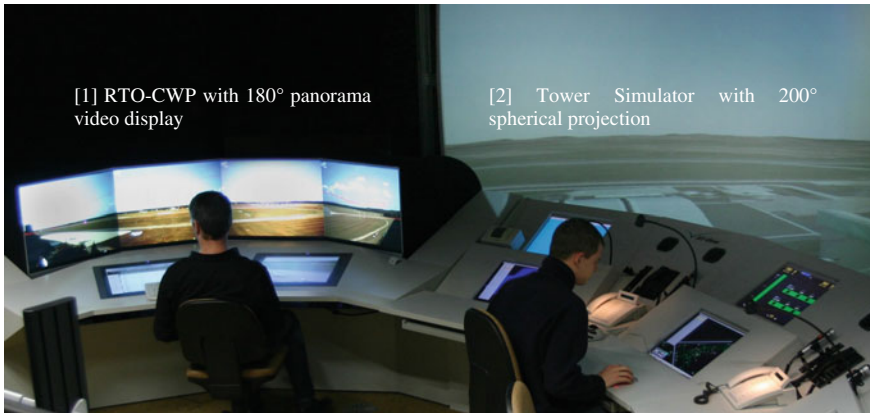
Fig. 1 The experimental workplace for remote tower operations (RTO-CWP) [1] integrated within the infrastructure of the conventional tower simulator [2] (IEEE copyright)

and the configuration of the RTO-CWP see the references mentioned above and chapters "Remote Tower Experimental System with Augmented Vision Videopanorama", and "Remote Tower Prototype System and Automation Perspectives" of the present book. The experimental RTO-CWP is depicted in Fig. 1.

In a preliminary simulation study, this experimental work place was tested with two controllers from the research airport Braunschweig-Wolfsburg; the airport also chosen for the simulation study. The study was concerned with questions regarding the feasibility of this novel concept in general and specifically with the usability of the workplace and its novel assistance functionalities (Möhlenbrink et al., 2010). Results of this study were promising with regards to the controllers' acceptance of this approach to remotely controlling air traffic, as well as operational feasibility of the workplace. The gaze-behaviour of the controllers was recorded during the simulations in order to describe the information search process of tower controllers in an objective way. There was a clear influence of the working position on the gaze behaviour. Due to the small sample size questions regarding the influence of the information augmentation could not be answered.

1.3 Aerodrome Control Work Environment

The tower controller is responsible for safety of operations within the aerodrome. The most fundamental tasks are the control and surveillance of traffic on the runway, taxiways and park areas as well as the surveillance and coordination of the whole aerodrome. Usually at regional airports, an executive controller (EX) and a coordinator (CO) work together in a team to control the aerodrome. The executive controller is in contact with the pilots via radio, while the coordinator is more concerned with

coordinating the arriving and departing traffic with other sectors and assisting the tower controller with the documentation on the flight strips. Both controllers share the responsibility for the safety of the operations. Nevertheless, the executive is the team leader and makes the final decision about the sequence of aircraft. In comparison to other air traffic control working positions, tower controllers do permanently have to react upon changes in the environment (Dittman et al., 2000). Therefore, they mentally have to integrate a variety of different information and cues, which enables them to take proactive actions (Tavanti & Bourgeois, 2006). The direct view on the airport is the most specific visual information source for aerodrome control, in contrast to the en-route controller. Additional visual information sources are flight strips, radar, and a weather display. Additionally, the controller is using radio for communication with the pilots, ground radio, and telephone. There are further information sources, especially assistance systems, that are not available at all airports (Papenfuss & Möhlenbrink, 2009).

1.4 Characteristics of Regional Airports and Consequences for Air Traffic Control

In general, regional airports do not possess a wide range of sensor technology due to high costs of these systems. Approach radar information is provided that covers the aerodrome traffic but not traffic on the ground. The traffic at a regional airport typically is a mix between flights operating under conditions of Visual Flight Rules (VFR) and Instrumental Flight Rules (IFR), with a high percentage of VFR traffic. Even though these rules refer to different meteorological conditions they also characterize features of the pilots. IFR traffic is mainly operated by commercial airline pilots. In contrast, flights operating under VFR conditions are mainly flown by private pilots and pilot trainees. Whilst IFR traffic is relatively predictable as a flight plan is scheduled hours in advance VFR traffic does not require a prior flight plan. Therefore, its occurrence is rather unpredictable for the air traffic controller.

These boundary conditions implicate several consequences for air traffic control at regional airports. The VFR traffic cannot be anticipated like IFR traffic and demands flexible reaction of the controllers. The fact that pilots of VFR traffic are usually less experienced has influences on the air traffic control service. It is likely that there are more deviations from the standardized controller-pilot radio communication, and there is less confidence that the pilots follow the commands from the tower correctly. Thus, the control of the mixed traffic demands more attention and checking from the controller than only controlling IFR traffic and is one contributor to mental workload (Vogt et al., 2006).

1.5 Controller Assistance via Information Super-Imposition and Automatic Zoom Camera Tracking

The replacement of the outside view by a live video offers new possibilities for controller assistance that could compensate for the missing sensor equipment of regional airports. The live video can be used for automated information extraction and information augmentation. Motion detection algorithms can detect moving objects like aircraft (AC), vehicles, or bird swarms. This information can be used for assistance like highlighting these objects on the panorama, to use this object data for the control of the zoom camera (automatic tracking), or to use the data for further data procession e.g. fusing radar position and video image position of AC for conformance checking. Furthermore, information gained through data fusion of other sensors can directly be superimposed onto the outside view of the tower controller. It is assumed, that in the future also at regional airports the means are provided to get position and identifier data (callsigns) from all aircraft. Data of a multi-lateration system or ADS-B sensors can directly be superimposed at the respective position on the video panorama (Fürstenau et al., 2008). Thus, information presented to the controller in the conventional workplace separately on his radar display and the flight strips, will be integrated into the outside view of the RTO-CWP.

1.6 Research Questions

In order to evaluate the concept of providing remote tower services via a controller working position with a video panorama (in the following called "RTO-CWP"), three different research questions were addressed in the study. (1) In how far are the RTO-CWP and the solution for replacing the far view (video panorama) suitable for a team of air traffic controllers in order to control air traffic at a regional airport. Where are the limitations of the concept? How is the workplace rated by controllers regarding acceptance and usability?

As pointed out above, the far view is a crucial part of aerodrome control. Therefore, besides the new work environment a second research questions deals with (2) the relevance of the far view and what information the controller uses from this information source. As the concept of a video panorama allows for novel assistance functions, the study investigates in a third research question (3) in how far the superimposition of different information provides a benefit for the controllers.

2 Method

2.1 Subjects

Twelve tower controllers from six German airports participated in the high-fidelity simulation study. The sample was rather heterogeneous regarding age (Range: 24–59 years, mean $M = 39$, standard deviation $SD = 12$) and job experience as tower controller (Range: 2–37 years, $M = 17$, $SD = 11$). All controllers were used to work with simulated traffic on high fidelity simulators because it is part of their education and training as a tower controller.

2.2 Experimental Design

A 2×3-factorial experimental design was tested with the within subject factors weather (good visibility, low visibility) and work place variant (conventional tower simulation baseline TO-BL, RTO-CWP baseline RC-BL), RTO-CWP and information augmentation RC-MD, RC-CA). After a series of three simulation runs for each team of two controllers, roles (executive, coordinator) and weather condition changed, resulting in three runs per team constellation. Weather condition and work place variant were varied in between the groups in order to control learning effects regarding the use of the new workplace.

2.3 Simulation Setting

The RTO-CWP was integrated into the High-Fidelity-simulation environment at DLR in Braunschweig and equipped with all necessary displays and HMI in order to represent a quasi-operational workplace. The simulated airport was Braunschweig-Wolfsburg. For the tower simulator, a 4-projector front projection (each 768 * 1024 pixel) on a 200° spherical surface was used, imitating the view out of the window without reduction of visual angles and proportions (life-size). The RTO-CWP was equipped with a 4-projector rear projection system (30″ diagonal, 1050 * 1400 pixel resolution) that realized the 180°-panorama of the corresponding four simulated cameras within a reduced operator viewing angle of ca. 125°. Due to the smaller distance of the projector to the rear projection and the better quality of the devices, the video on the RTO-CWP was brighter, had a higher contrast, saturation, and acuity in comparison to the tower simulator. The simulation setup of both, the tower and RTO simulation, comprises an operator's position for executive and coordinator, as well as working positions for two pseudo pilots and a supervisor.

2.4 Experimental Task

The controller team had to manage air traffic in 30 min simulation runs at the airport Braunschweig-Wolfsburg according to published legal guidelines from the DFS. They divided their tasks in executive and coordinator. The traffic scenarios were designed with the same demand (12 AC in 30 min) and overall traffic mix of 60% VFR traffic. In order to reduce learning effects, the callsigns of the AC differed between the simulation runs. This traffic rate of 24 AC/hour had to be chosen significantly higher than usually expected for the small low-traffic airports in order to understand to create challenging situations for the controllers to understand, in how far the remote tower concept influences the working methods of the controllers and if there are limitations to the concept.

2.5 Experimental Conditions

As a methodological approach for validation of the novel work environment, working in a common 200°-tower simulator is compared with working at the RTO-CWP. In the human-in-the-loop simulations four different conditions of the workplace were realized.

(1) *Tower Simulator Baseline (TS-BL)*

In this condition, controller teams worked on the conventional high-fidelity airport tower simulator with life-size 200° projection of the outside view. This condition was instructed as working in a real tower at the airport Braunschweig-Wolfsburg.

(2) *RTO-CWP Baseline (RC-BL)*

Controller teams were working at the new experimental workplace RTO-CWP with video panorama and zoom camera. This experimental condition served as a baseline to compare working in the conventional tower simulator with working at the RTO-CWP. Compared to the TS-BL this working environment was introduced as being 200 km away of the airport.

(3) *RTO-CWP and Movement Detection (RC-MD).*

Compared to the RC-BL condition, image processing algorithms are realized for the detection of moving objects. Moving objects were superimposed with a coloured frame. It was further possible for the tower controller to direct the zoom camera on a moving object or start the automatic tracking of that object by manually clicking on a detected moving object.

(4) *RTO-CWP and Callsigns (RC-CA)*

In this condition, the callsign of each AC was superimposed next to the respective AC. The position of the AC was gathered from the simulator in order to place the callsigns onto the video panorama. Moreover, semi-automatic tracking function was available, like in the RC-MD condition.

Weather (good visibility vs. low visibility)

There were two weather conditions, both in the range of conditions to operate VFR flights, which differed in the range of sight. In the good visibility condition the viewing distance was unlimited, in the low visibility condition the viewing distance was diminished to 3.5 km (about 2.17 miles).

2.6   Controllers Working Positions

The two controller positions (executive and coordinator) were equipped with a generic approach radar application, a weather display, paper-based flight strips, radio, and the far view, displaying the visual scenery as seen from the tower position. Within the tower simulator, two identical radar applications were placed in front of each of the tower controller position. It was necessary, because of the large distance of the coordinators operating position to the radar display placed in front of the executive.

At the RTO-CWP, there is an additional control display in front of the executive for controlling the zoom camera and its functions, but just one radar right in front of the coordinator's operating position. Nevertheless, the spatial dimensions of the RTO-CWP allow the executive to sufficiently check the radar display. The configurations of the working positions are depicted in Figs. 2 and 3.

Fig. 2 Design of working positions in the tower simulator encompasses [1] two radar displays, [2] weather display, [3] flight strips; [4] far view (200° life-size projection). The executive is positioned to the left, the coordinator to the right of the flight strips

Fig. 3 Design of the RTO-CWP with [1] radar, [2] weather display, [3] flight strips, [4] far view (video panorama), [5] control display for zoom camera. The executive is positioned to the left, the coordinator to the right of the flight strips

At the two pseudo pilot working positions, located in a separate room, two trained pseudo pilots controlled each AC via command pad and mouse clicks, according to the radio advices given by the tower controller. They translated heading, speed and altitude commands and clearances advised by the tower controller into inputs for the simulation software.

3 Dependent Variables and Data Analysis

3.1 Feasibility, Acceptance and Usability of the Workplace

In this study, questionnaires and interviews were used to assess usability, feasibility and acceptance of the experimental workplace and the assistance tools in order to gain insights about operational validity. The study took part on an experimental workplace; therefore, it was expected that handling issues with the human–machine-interfaces might occur. After each simulation run, both controllers had to fill out a set of questionnaires, addressing the relevant aspects of human–machine interaction with regards to feasibility and acceptance. Questions from the EUROCONTROL SHAPE questionnaires (Eurocontrol, 2008) were used. Ergonomic concepts addressed with the questionnaires were mental workload and the impact of automation (SHAPE-AIM), trust in automation (SHAPE-SATI) and situation awareness (SHAPE-SASHA). Only items concerning the experimental workplace and the new components like information augmentation, zoom camera, touch pen, tracking function were selected. Each item was rated on Likert scales ranging from 5 to 7 point scales, depending on the original questionnaire. Further questions from the system usability scale (SUS) were used to access subjective rating concerning the usability of the controller working positions on a five-point Likert scale (Brooke, 1996).

Sum scores were calculated for each scale, negative items were inverted beforehand according to the manuals (Eurocontrol, 2008).

Furthermore, after each simulation run feedback of the controllers concerning the usability of the RTO working environment and the different information augmentation variants was gathered in semi-standardized interviews. The traffic situations of the simulation runs were used as triggers for the controllers, to comment on operational requirements for remotely controlling an airport. The feedback of the controllers was clustered in two main topics: usability of the workplace and feasibility of the concept. Comments of the controllers are reported in an aggregated manner.

3.2   Assessing the Relevance of the Far View

As outlined, the far view is the most distinct information source of the tower controllers' work place. Even though its relevance in general is rated high, it stays unclear what specific information is used from the far view. For validation of the new work environment that aims at replacing the outside view, it is crucial to understand what information controllers need from the outside view in order to ensure that this information is available at the new workplace, as well. The radar display of the en-route controller, as the main visual information source, presents a synthetic, integrated graphical representation of several radar sources. It contains information about the three-dimensional position, velocity and heading of AC in the airspace. In comparison, the far view is more complex and less explicit regarding the information, the controller perceives by "looking outside". Information claimed to be derived by tower controllers, ranges from the position and velocity of AC to the visual range at the airport and animals on the runway (Papenfuss & Möhlenbrink, 2009). Although the far view is not clearly understood, it is argued, that it is as a necessary element in the safety chain for tower air traffic control (Möhlenbrink et al., 2010). As a general rule, in order not to miss any unexpected event, tower controller are expected to look outside as often as possible. Therefore, the so called head-down-times (i.e., when the controller is not looking outside, but on other information sources) should be minimized, in order to minimize the risk of not detecting unpredictable events, as well (Hilburn, 2004). The head-down and head-up times of tower controllers have been investigated by means of using objective metrics, like controllers gaze position, in order to derive percental dwell times (e.g. Pinska, 2007). Different publications report controllers using between 20 up to 54% of time for the far view (Lange, 2014; Oehme & Schulz-Rückert, 2010; Pinska, 2006) and 51%–80% for head-down times respectively. Even though these studies provide indicators for the frequency of use of far view, this metric is not explaining, what information controllers do actually perceive when they look outside in order to collect information for their decision making.

The relevance of information or data is an essential feature, to assess and evaluate human–machine interaction. For instance the psychological concept of situation

awareness that is known to have an impact on human–machine performance (Durso et al., 1999), refers to the "perception of the relevant information in the environment" (Endsley, 1995, pp. 34). In order to better understand the relevance of the far view and the acquired information in this study a multi-method approach is used. Durso et al. (2008) suggest that relevance of information can be quantified and computed on three dimensions—number of tasks which require the information, task frequency, and the criticality or importance of the tasks that use the information.

In order to achieve an applicable approach, in this simulation study relevance of an information source is defined as frequency (percentage) of use and uniqueness and criticality of information. Accordingly, eye tracking is used to gain objective metrics that describe the distribution of visual attention. Because of the spatial limitation of the head tracking unit, only one working position, RTO-CWP or tower simulator, could be used for eye gaze measurement. It was decided to measure eye gaze behaviour at the RTO-CWP. As this metric is derived in a simulation setting, furthermore a ranking of information sources regarding the perceived importance for the daily work of the tower controller is used. In a questionnaire, controllers were asked to rank the information sources they use in their daily work according to their importance. The list consisted of 20 different information sources. The rank of the information sources is calculated as the mean ($N = 12$), where the first rank is weighted with the score 20, the last rank with the score one. To identify the severity of situations, where information of the far view is needed, a questionnaire was used, as well as free interviews. In the final questionnaire controllers were asked to remember situations, they had perceived as critical. Out of those situations they were asked to choose one they remembered particularly well and to describe this situation and what they did to solve this situation. Afterwards, they were asked to select all those information sources out of a list from which they derived relevant cues for the specific situation that triggered their actions. By means of interviews, controllers where further asked to specify what information they use from the far view that is not available through other information sources.

3.3 Benefit of the Assistance Tools and Analysis of the Eye-Tracking Data

The concept of the assistance tools investigated in this study aims at supporting the tower controller by providing relevant information within the main visual information source, the far view. Initial approaches used a transparent display to superimpose relevant information for tower control (Fürstenau et al., 2004b; Peterson & Pinska, 2006); see also chapter "Introduction: Basics, History, and Overview" (Introduction). In one approach wind information was superimposed over the runway (Schmidt et al., 2006). Controllers have to submit wind information to the pilot when they give a landing clearance. Additionally, they visually have to scan the runway to make sure that it is not occupied. Normally, the wind information is provided on an additional

display. It is assumed, that by superimposition of relevant data head-down times of the controllers can be minimized (Peterson & Pinska, 2006).

To assess the benefit of the assistance tools used in this study a multi-method approach was used. The introduced augmentation assistance tools (movement detection, callsigns) were evaluated by means of a dichotomous scale ("Yes", "No"), asking for usability, benefit and reliability. Additionally, during the simulation runs at the RTO-CWP, eye data was recorded as an objective measure of the information acquisition process and the head-down times of the controllers.

The head-mounted eye-tracking system in combination with an optical head-tracker via infrared cameras allows a parallel tracking of both controllers in the complex simulation environment (Möhlenbrink et al., 2010). Dwell times for the different information sources were determined as an index for the visual attention distribution. The analyses of this data focused on the differences of this index between the roles controller and coordinator, as well as the differences between RTO-CWP baseline and information augmentation conditions. For these analyses, a 3d-model of the RTO-CWP was generated, which includes the position of the screens of the visual information sources (compare Fig. 3). Based on that 3d-model, the eye gaze is transferred during the measurement into position data on the screens. After eye data recording, the Eye-Tracking Analyser (EyeTA) was used to calculate fixations and respective dwell times for defined areas of interest (AOI). In this study, the AOIs were identical to the information sources at the RTO-CWP (compare Fig. 3), thus being far view, flight strips, radar display, weather display and zoom camera. The EyeTA tool was developed at the Institute of flight guidance for semi-automatic analysis of large eye-tracking data sets. For calculation of fixations, the dispersion threshold algorithm suggested by Salvucci and Goldberg (2000) was implemented with minimum fixation duration of 100 ms. Fixations are used as indicators for conscious information acquisition; so in combination with the AOIs they represent the information controllers gathered during the simulation run, in order to make decisions.

4   Results

4.1   Feasibility, Acceptance and Usability of the Concept

The scores for all questionnaires in the four conditions are shown in Fig. 4. The four graphs show scores for mental workload (AIM, Fig. 4a), trust in automation (SATI, Fig. 4b), situational awareness (SASHA, Fig. 4c) and system usability (SUS, Fig. 4d). The dotted lines refer to the minimum and maximum values of the scales, as well as the mean. First, the absolute values for the scores are described. For the AIM scale, a lower score is better as it refers to less negative impact of the novel automated system on mental workload. The scores for all four experimental conditions are relatively low, indicating no major negative influence on mental workload. The SASHA scale is
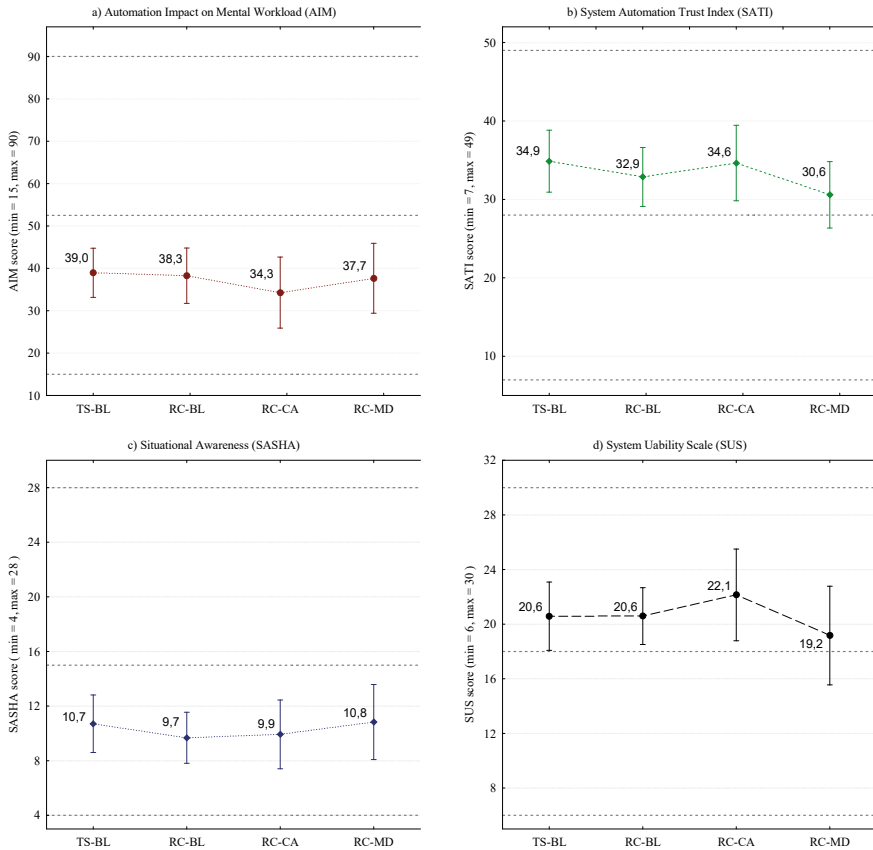
Fig. 4 Mean values for the operational feasibility questionnaires; vertical bars indicate 0.95 confidence intervals. **a** Impact of automation on mental workload (AIM, min = 15, max = 90). **b** Trust in automation index (SATI, min = 7, max = 49). **c** Situational awareness (SASHA, min = 44, max = 28). **d** System usability scale (SUS, min = 6, max = 30). X-axes indicate experimental conditions: TS-BL = conventional tower simulator, RC-BL = RTO-CWP Baseline, RC-CA = RTO-CWP Callsign Augmentation, RC-MD = RTO-CWP Movement Detection Augmentation. Dotted grey lines indicate minimum, maximum and medium scores for each of the scales

interpreted in the same way; a lower score means less negative influence on operators' situation awareness. The scores for all four experimental conditions are also relatively low, indicating no major negative influence on situation awareness. The scales of SATI and SUS indicate a higher trust in automation and usability, corresponding to higher score values. For all four experimental conditions, the scores are located for SUS as well as SATI slightly above the middle of the scale, indicating a slightly larger than medium trust in the automation and a slightly larger than medium usability of the work places. The AIM score for the callsign condition also is the lowest for all four conditions; the highest was measured for the conventional tower simulator ($M_{\text{TS-BL}}$ = 39.0, $SD_{\text{TS-BL}}$ = 13.7). The trust in automation index tended to be highest for the

TS-BL condition compared to the lowest value for the RC-MD condition ($M_{\text{TS-BL}}$ = 34.9, $SD_{\text{TS-BL}}$ = 9.3, $M_{\text{RC-MD}}$ = 30.2, $SD_{\text{RC-MD}}$ = 6.8, 7 items, 7-point-Likert scale). Regarding the situational awareness, the RC-BL had the best (lowest) score compared to the RC-MD condition ($M_{\text{RC-BL}}$ = 9.7, $SD_{\text{RC-BL}}$ = 4.2, $M_{\text{RC-MD}}$ = 10.8, $SD_{\text{RC-MD}}$ = 4.6, 4 items, 7-point-Likert scale). There is a non-significant effect for the usability scale that the condition RC-CA is rated more usable than the condition RC-MD ($M_{\text{RC-CA}}$ = 22.3, $SD_{\text{RC-CA}}$ = 5.5, $M_{\text{RC-MD}}$ = 18.4, $SD_{\text{RC-MD}}$ = 5.9, 6 items, 5-point-Likert scale).

As the high standard deviations show, inter-individual differences were quite high. None of the four questionnaires used to assess feasibility and usability of the workplace did show in a paired t-Test a significant difference between the real tower simulator baseline (TS-BL) and the RTO-CWP baseline (RC-BL). Regarding the two types of information augmentation, two subsamples were created for the "Callsign" and for the "Movement Detection" condition. T-Tests were conducted to analyse, whether the two subsamples rating the RTO-CWP had similar ratings in the RTO-CWP baseline condition RC-BL. The results showed that both subsamples did not differ significantly, but for the situation awareness questionnaire SASHA. Here, the controllers taking part in the MD-condition tended to rate the negative impact on situational awareness less than the controllers taking part in the CA-condition. (MBE = 7.37, SD = 3.02, MCA = 11.0, SD = 4.31, t(20) = −2.09, p = 0.05). In a next step, a Repeated Measures ANOVA of the influence of the information augmentation on the SHAPE questionnaire scores was conducted. Bonferroni post-hoc test were used to check for significant differences between the conditions. No significant differences could be found between the RTO-CWP baseline RC-BL and the motion detection condition RC-MD. The comparison between the RC-BL and the callsign condition (RC-CA) revealed a significant decreased mental workload (AIM scale) when augmenting callsigns into the far view ($M_{\text{RC-BL}}$ = 41.8, $M_{\text{RC-CA}}$ = 34.4, $F(1,15) = 5.43$, $p < 0.05$).

One general usability issue of the RTO-CWP is that it transforms the tower workplace into a sole computer work place. This can cause new ergonomic issues. A feared feeling of "loss of reality" on this computer work place was not rated as critical, because the en-route controller also only uses a computer work place. Compared to the set-up of the real tower (TS-BL), controllers reported that they could not use the head position as an indicator for the position of an AC on the airport or the traffic pattern, whilst working on the remote tower workplace. The change of those dimensions in the experimental remote tower work place (RC-BL) was reported to lead to some initial problems with the judgement of distances. All controllers described differences in working at the RTO-CWP compared to a real tower regarding the possibility to detect far away AC due to the resolution of the video cameras (see chapter "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation"). This effect was especially apparent for VFR traffic in the traffic pattern. Furthermore, controllers reported that the area of the approach that was visible on the 180° video panorama was too small. They would start to search the AC earlier in a real tower with real far view.

The workplace and its features, like the zoom camera, were rather intuitively to use. In general, the controllers wanted personalized settings for the assistance tools; for instance, switching information augmentation on and off, selecting single AC that should only be augmented, or defining a set of fix positions for the zoom camera, which can be accessed via shortcuts.

In general, the concept of replacing the view outside the tower through a high resolution video was commented as sufficient in normal operations for small airports with a simple layout and a medium traffic amount. Most controllers rated this concept as suitable for low traffic situations (start and end of day of operation) or contingency purposes. High accuracy and reliability of the video pictures is mandatory. Additional sensor solutions for night and bad visibility operations were mentioned as necessary and helpful. Some controllers raised the concern of liability, when something causes an incident or accident that could have been detected with a real far view but not with the video panorama. There were comments that some relevant information is only available by being present at the airport, for example, precise information about weather and the runway surface status. Furthermore, controllers mentioned that especially for dealing with VFR traffic precise knowledge of the vicinity of the airport is necessary, e.g. landmarks used for navigation or the location of hospitals. There were concerns that this knowledge and experience gets lost whilst working on a remote working position.

The controllers were confronted with an experimental workplace, which displayed only 180° of the aerodrome. Controllers had divergent opinions about the missing backside view. Some preferred to see it permanently; some preferred to have the possibility to see these areas via a moveable camera. Others mentioned the possibility to adapt procedures to ensure safe operations in this area, e.g. not to use the backside traffic pattern, or to allow only one AC in the backside traffic pattern at a time. In general, new procedures or changed working styles were both reported as possibilities to overcome those limitations introduced by the RTO working position. For example, controllers stated that they would use more extensively position reports of the pilots to control air traffic.

4.2   Relevance of Far View

Eye tracking data of the simulation runs are used as an objective metric, to describe the attention distribution of controllers during their control task. Eye tracking data is available only for the RTO-CWP, so all data refer to the attention distribution whilst working remote. Eye-tracking studies regarding the attention profile in a conventional tower at a small to medium sized airport have been conducted by Lange (2014), Oehme and Schulz-Rückert (2010), and Pinska (2006). Altogether, 43 eye data files containing 30 min simulation runs could be gathered, 22 from the executive and 21 from the coordinator's operating position. 72% of the recorded eye gaze position data could be matched to one of the predefined AOIs ($SD = 13\%$).
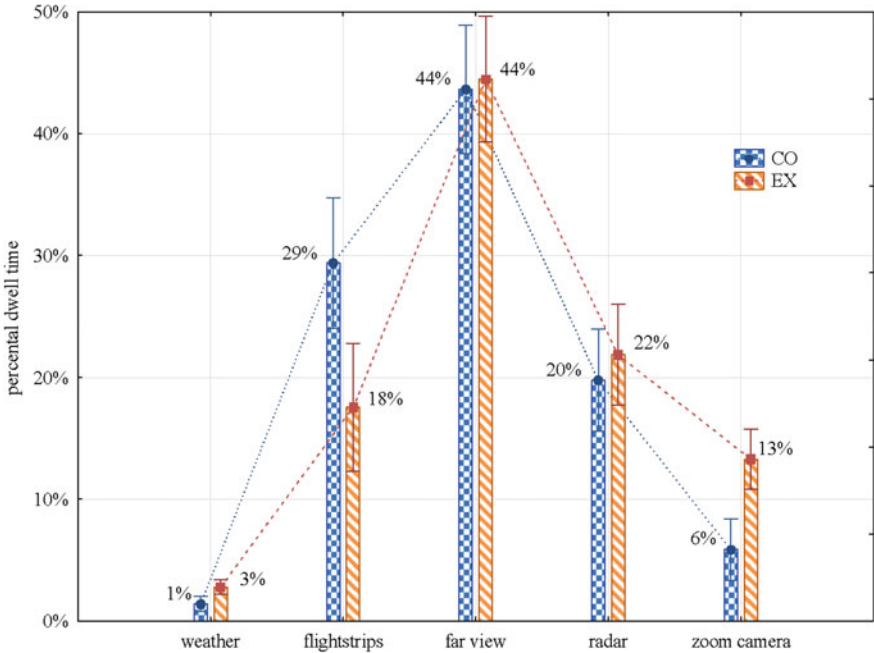
Fig. 5 Mean percental dwell times M of the visual attention on the different visual information sources of the RTO-CWP, separated for the two roles executive (role EX, $n = 22$) and coordinator (role CO, $n = 21$). Vertical bars indicate 0.95 confidence intervals

Both roles show a comparable profile for using information sources with the highest rank for the far view ($M_{\text{FarView}} = 44\%$, $SD_{\text{FarView}} = 12\%$), compare Fig. 5.

For the executive position, the ranking is (1) far view—(2) radar—(3) flight strips—(4) zoom camera and (5) weather, for the coordinator this ranking is (1) far view—(2) flight strips—(3) radar—(4) zoom camera and (5) weather. In order to understand if the role (executive, coordinator) influences the distribution of the visual attention, a repeated measurement ANOVA was conducted, with the factor role (executive, coordinator) and visual information sources as repeated factor. The results indicate a small effect for the interaction of role and display ($F(4,164) = 5.18$, $p < 0.001$, $\eta^2 = 0.11$). The post-hoc Bonferroni-test show that there is a significant difference for the attention distribution on the flight strips ($M_{\text{EX}} = 18\%$, $SD_{\text{EX}} = 9\%$; $M_{\text{CO}} = 29\%$, $SD_{\text{CO}} = 15\%$; $p = 0.01$) (cf. Fig. 5). The coordinator spends about 10% more visual attention on the flight strips compared to the executive, which can be explained with the share of tasks between the operators.

The result of the ranking of the information sources for the top ten information sources for the aerodrome working position is depicted in Fig. 6. This ranking differs significantly from a corresponding one obtained through a work analysis for the conventional tower workplace of a large international airport (see (Papenfuss & Möhlenbrink, 2009) and chapter "Remote Tower Prototype System and Automation
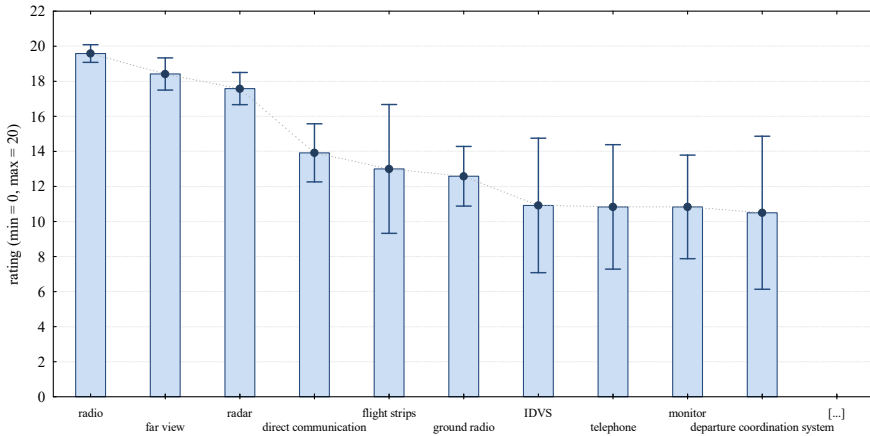
Fig. 6 Top ten of the most relevant information sources for the aerodrome working position, data from $N = 12$ controllers. Bars resemble the mean rating (minimum $= 0$, maximum $= 20$); whiskers indicate 95% confidence intervals

Perspectives", Sect. 2). Direct communication means the face-to-face interaction with team members in the tower. The information source IDVS resembles an integrated data processing system mainly used for weather data and operational data (e.g. runway in use. The information source monitor stands for additional cameras for surveillance of the airfield that are presented on additional monitors in the tower cab. According to this ranking, the far view is seen as the seconded most important information source ($M = 18.4$, $SD = 1.4$) followed by the radar ($M = 17.6$, $SD = 1.4$), only radio communication is ranked more important ($M = 19.6$, $SD = 0.8$). Standard deviation for the ranking of these information sources is rather small, compared to the large standard deviations for the lower ranked information sources, e.g. flight strips, IDVS and telephone. This means that all controllers had a rather similar ranking for the first three information sources, but strongly differed in their individual ranking for the following information sources.

Table 1 Use of information sources in critical situations

Information source	Response ($N = 12$)	Frequency
Far view	10	83%
Radar	7	58%
Radio	7	58%
Telephone	1	8%
Ground radar	1	8%
Other displays	0	(0%)
Monitors of video cameras	0	(0%)
Alerts	0	(0%)

The results of the analysis of the critical situations are shown in Table 1. In ten out of the 12 described situations, the far view provided the relevant information used for decision making. Both radar and radio communication provided relevant information in 7 cases and telephone and ground radar were named each in one case. The information sources other displays, monitors of video cameras as well as alerts were not selected.

For the ten situations, where the far view provided relevant information, controllers where further asked to specify what information they got from the far view. These answers ($n = 10$) were analysed and grouped into four functional categories of information: anticipation of abnormality ($n = 2$), discovery of abnormality ($n = 2$), composition and verification of the traffic picture ($n = 5$) and unspecified/all ($n = 1$), compare Table 2.

Within the free interviews, controllers were asked, in which operational situations the far view provides information that is not provided by any other information source. Accordingly, one of the most critical information is, whether the runway is occupied or not. Furthermore, the far view is used to timely verify whether pilots executed the commands given by the controller. The controllers need this verification to estimate if AC will be separated properly or if further control steps are needed to ensure safe operations. One situation identified by controllers is the timing of the base turn (turn base clearance) of a VFR AC. Controllers mentioned that it can be hard to timely decide only with the help of the radar whether the pilot already followed the command (to turn), because the update rate of the radar is relatively low. A change in the attitude cannot be detected and a change of the position will be visible with recognizable delay. In the far view, the change of the aircrafts attitude can be verified more precisely, sometimes by reflections on the aircrafts surface that are clearly visible to the controllers.

Another situation that was mentioned by controllers is the cross checking of, especially, VFR-pilots action on the ground. It might be of interest to check whether the pilot took the correct taxiway, because confusions can happen if a pilot is new

Table 2 Overview of functional categories and samples of information

Id	Category	N	Description
1	Unspecified/all	1	All information
2	Anticipation of abnormality	2	AC performs "swing" manoeuvre, rapid change of attitude
			Collision risk (Turn-off + next Landing)
3	Discovery of abnormality	2	Observation of military jets and small AC
			Observation of motion track of AC
4	Composition and verification of traffic picture	5	Strength and location of smoke
			Identification truck/vehicle on ground radar
			Verification of information
			Precise and timely information of position, speed and separation of AC

to an airport. Controllers rate the experience of pilots through cues derived by radio communication. If a pilot sounds unsure or has problems to follow the phraseology, controllers will check the pilot in the far view more closely, e.g. if the pilot takes the correct taxi way.

4.3 Benefit of Assistance Tools

The controller rated the assistance tools by means of a specific questionnaire. Questions are structured in three categories—regarding the tracking functionality of the zoom camera (question 1–3), the usability of information augmentation of either "callsigns" or "automatic movement detection" (question 4–6), as well as the potential of these functionalities to provide operational benefit (question 7 and 8). Results are presented for both roles executive and coordinator separately. Question 3 was only answered by the executive role, because the coordinator did not have had the means to directly interact with the zoom camera. Additionally, the two augmentation types "Callsigns" and "Movement Detection" are contrasted for category 2. The results are shown in Tables 3 and 4. The first number indicates the times the question was answered with "Yes"; the number in braces indicates accordingly the times answered with "No".

With regards to the tracking functionality (questions 1–3), controllers ratings are relatively clear that this functionality is a sensible feature. There are issues regarding the handling and reliability of the feature, where controller's ratings provided no clear results or tendencies. In the free interviews some usability issues where further specified. The tracking functionality of the zoom camera was regarded as useful for traffic in the traffic pattern, to detect the AC type, to determine whether AC has left the runway, or in critical situations in general, where usually one extra controller follows the critical objects with his glasses. In the experimental set-up chosen in this study, the tracking was not reliable enough. Furthermore, the tracking was sometimes too slow to follow the AC. The usage of the zoom camera was rated as "intuitive", although too sensitive especially in high zoom levels. The amount of training for precise manual control was rated "high". Controllers described desirable advanced control options. One is, to select a specific AC in the approach radar display and

Table 3 Answers regarding tracking functionality

			EX Yes(No) $N = 12$	CO Yes(No) $N = 12$
Tracking	1	Is the tracking functionality a sensible feature?	11 (1)	11 (1)
	2	Did the tracking functionality work correctly?	4 (8)	6 (6)
	3	Was it possible to use tracking functionality without handling problems?	7 (5)	– /–

Table 4 Answers regarding information augmentation

			Callsigns CA		Movement detection MD	
			EX $n=6$	CO $n=6$	EX $n=6$	CO $n=6$
Usability	4	Was the additional information a reliable assistance?	4 (2)	5 (1)	1 (5)	1 (5)
	5	Did you use the additional information?	5 (1)	4 (2)	4 (2)	3 (3)
	6	Was the additional information correct?	4 (2)	5 (1)	2 (4)	3 (3)
Operational benefit	7	Did the system provide new possibilities for your work?	3 (3)	2 (4)	0 (6)	2 (4)
	8	Did the system changed something fundamental in your work routines?	2 (4)	2 (4)	1 (5)	0 (6)

the coordinates of the AC position will be used to automatically position the zoom camera.

The augmentation type "callsigns" was rated as reliable, the controllers used the information that was provided, and the information was rated as correct. There was no clear opinion regarding the operational benefit of the assistance tool or whether it changed work routines fundamentally.

In comparison to the augmentation type "callsigns", "movement detection" was perceived, as not reliable. There was no clear opinion regarding the usage of the information as well as their correctness. There was a clear answer from the executive that this feature did not provide an operational benefit. In comparison, two coordinators saw an operational benefit. No fundamental change in work routines was apparent.

The comments of the free interview regarding the augmentation assistance tools give a more thorough picture: concerns were raised that through augmentation other information is masked. In general, the augmentation of callsigns was rated as more helpful than the highlighting of moving objects (movement detection). The augmentation of callsigns was rated as helpful for parking AC, as well as AC on the final. Both augmentations were rated as helpful to detect far away objects. Some issues with the lower resolution of the video panorama, like detecting far away AC, could be compensated through augmentation. For example, through the motion detection AC in the traffic circuit are highlighted and thus quickly visible to the controllers.

On the other hand the danger of over-reliance was reported, especially for the augmentation of the callsigns. The reliability and accuracy of the data was mentioned as a main factor, whether an augmentation solution is accepted or not. If a controller perceives the information augmentation as not reliable it would be turned off in order to avoid wrong decisions based on faulty information.

Even if controllers did not indicate a fundamental change in work routine, the effect of the augmentation tools on head-down times was controlled by means of the

eye tracking data. The hypothesis was that information augmentation increases the head-up times, thus time spent on the far view should be increased. The comparison of the dwell times on the AOIs for the different information augmentation conditions shows only a significant difference for the use of the radar-display between the baseline condition and the two information augmentation conditions. No systematic difference could be found for the dwell times on the far view. The results are depicted in Fig. 7, with (a) the difference for callsigns and (b) movement detection conditions.

Due to the experimental design, each participant worked either in the movement detection (MD) or callsign (CA) condition in the specific role, but all participated in the baseline condition. Therefore, it is impossible to calculate a repeated measure ANOVA (BL, MD, CA). However, paired t-tests were calculated for the controller subsample participating in the BL- and MD-condition and for the subsample participating in the BL- and CA-condition. First, it was tested whether the dwell times of the two groups differ significantly in the BL-condition. The results of the t-Test (M_{BL-CA} = 21.4%, $M_{BL-MD} = 20.7\%$, $t(20) = 0.21, p > 0.05$) show that there is no systematic difference in the baseline condition, thus the two subsamples are comparable. While in the movement detection condition significantly more attention is distributed to the radar display ($M_{RC-BL} = 21\%$, $SD_{RC-BL} = 10\%$, $M_{RC-MD} = 25\%$; $SD_{RC-MD} = 14\%$, $t(8) = -2.31, p < 0.05$) compared to the baseline, in the callsign condition RC-CA highly significant less attention is distributed to the radar display (M_{RC-BL} = 21%, $SD_{RC-BL} = 5\%$, $M_{RC-CA} = 16\%$; $SD_{RC-CA} = 5\%$, $t(12) = 4.73, p < 0.001$) as compared to the baseline condition.
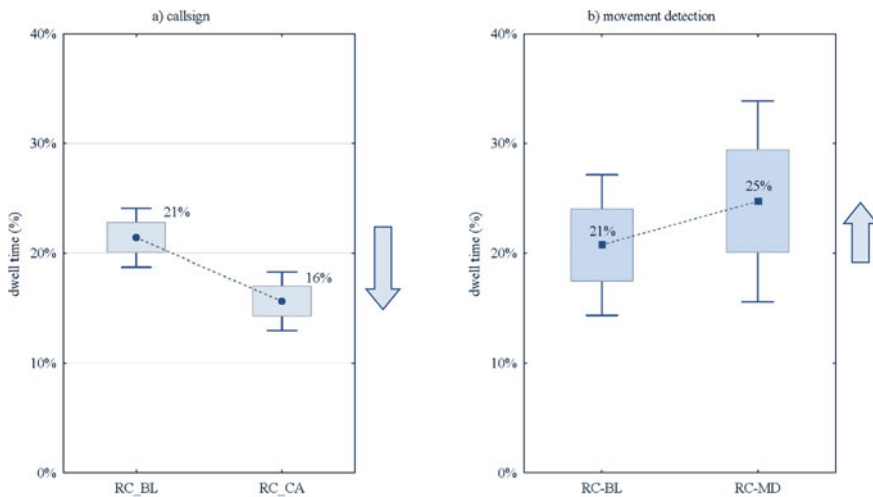
Fig. 7 **a** The graph shows the mean percental dwell time on the AOI RADAR for baseline (RC-BL) and callsigns (RC-CA) condition. With callsign augmentation, less visual attention is spent on the RADAR display. **b** The graph shows the mean percental dwell time on the AOI RADAR for baseline (RC-BL) and movement detection (RC-MD) condition. With movement detection, more visual attention is spent on the RADAR display. Boxes indicate standard error, bars 1.96 * standard error

5 Discussion

5.1 Feasibility, Acceptance and Usability of the Concept

The results of the questionnaires, used to assess feasibility and usability of the workplace, are promising with regard to the direct comparison of the tower simulator and the new workplace, the RTO-Console. No significant differences were found between those two conditions; regarding trust in automation, mental workload, situation awareness, and usability. One aim for the design of the RTO-CWP was to build it as similar as possible to the conventional tower in order to increase the acceptance for the concept. So, besides enabling remote control, working methods should not be affected by the concept. Therefore, the result indicates the introduced changes not to have a negative influence, compared to a conventional tower simulator.

There are objective differences between these workplaces, like the reduced panorama viewing angle (ca. 125°) that in turn reduces the required head rotation for the 180°—reconstructed far view. The results of the interviews indicate that some of these changes introduced by the RTO-Console could be compensated through training. Nevertheless, there are issues with the new workplace that need further consideration. One aspect is the resolution of the video panorama. Especially for smaller regional airports with a high percentage of VFR traffic the visibility of small AC in the traffic circuit is necessary.

In general, the feedback concerning the usability of the workplace was positive. The RTO-Console, the zoom camera and the tracking functionality are rated as "intuitive to use" and suitable for smaller airports with moderate traffic. The use of high resolution video stream was seen as a "good" approach. Controllers used the video data on the panorama instead of the radar for immediate verification of instructions given to the pilots, like start-up or a new heading, as the update rate of the radar is too low. As regional airports have a higher percentage of VFR traffic and, therefore, a higher chance of unexpected events, this timely update via the video is beneficial.

There are limitations of the results obtained, because in the simulation controllers only had to control air traffic no ground based traffic was included, e.g. fuelling trucks, follow-me cars or vehicles for construction work. The coordination with and monitoring of this traffic is a substantial part of the controllers task at a small airport. Furthermore, weather was kept stable over each simulation run. The evaluation of weather situations was indicated by the controllers as potentially difficult, when using the RTO-Console.

Because in this experimental workplace the reconstructed far view was limited to 180°, procedures to allow safe traffic control for AC flying in the areas without visual information could be applied. This certainly has effects on the capacity and must be seen as a trade-off between technical equipment invested in the remote control of the airport and applicable procedures. In general the concept for remote control as discussed in the present work applies for smaller airfields that do not have a major

issue with capacity. Thus, a reduction of capacity in remote control operations could be a valid approach.

Regarding the two information augmentation solutions, a positive effect of these assistance functions was expected on the scales of the SHAPE-questionnaires; nevertheless no significant differences to the RTO-Console Baseline could be found. The tower simulator was rated best on the scale for trust in automation when looking at the trends. As there was no automation in this condition, the automation in the other conditions tended to be a source for mistrust in the system. On the other hand, the assistance by augmenting callsigns tended to be less mentally demanding compared to the standard work environment. Seemingly, assistance can diminish mental workload if it is highly reliable. Maybe no effects of the information augmentation could be found, due to the experimental status of the workplace, as well as of the assistance tools. In general, the reliability, the accuracy and the adjustment to individual preferences of the assistance tools were mentioned as momentous by the controllers. In the RTO-Console main assistance solutions are based on the video data. Especially, the detection of AC was rated as helpful when they are in a distance that they have the size of only 1–4 pixels in the video images. This requires the reliable detection of extremely small objects on the basis of video data. If more advanced technical camera and data transmission solutions are available, a higher resolution of the video could resolve these problems. If not, an adjustment of the control procedures is essential.

5.2   Relevance of the Far View and the Visual Information

The eye tracking data of the present work show that the tower controller directs most visual attention on the video panorama, which replaces the view out of the window. Overall the percentage was 44% of the time; these results are consistent with other results in the literature (Lange, 2014; Oehme & Schulz-Rückert, 2010; Pinska, 2006, 2007) as these studies indicate a valid range of 20–54% of visual attention at the far view. So, whilst working at the simulated remote tower environment, air traffic controllers had comparable working methods regarding their distribution of visual attention as in a conventional tower. In the preliminary study, executive and coordinator had different attention distribution profiles. The coordinator spent less time on the far view than the executive (Möhlenbrink et al., 2010). These results could not be replicated; both roles had remarkably similar profiles. However, compared to the executive the coordinator spent more visual attention on the flight strips, because he had to document clearances and flight progress on the flight strips. Furthermore, the coordinator had no means to interact with the zoom camera independently, so less attention was spent on this display. The differences in the results of the two studies might be due to the fact that in the main study reported in this paper, controllers were confronted with an unfamiliar airport, an unknown team member and a partly new work organization. It can be assumed that controller teams develop distinguishable profiles of visual attention for their different task sets over time. But the similar visual attention profiles can also be interpreted in terms of redundancy. The coordinator

visually follows the traffic in the same manner, as the executive does. This behaviour might be an enabler for effective cross-checking and cross-monitoring behaviour, necessary for building up team situation awareness and for adapting continuously within the team (e.g. Dickinson & McIntyre, 1997; Paris et al., 2000).

Furthermore, controllers ranked the different information sources they use for tower control. In this ranking, the far view was rated the most influential visual information source, followed by radar. But the overall most important information source is radio communication. For remote tower control solutions for radio communication seem rather straightforward to achieve. Nevertheless, radio communication becomes a research topic if two or more airports are controlled from a remote tower center (Wittbrodt & Thüring, 2010; see also chapter "Planning Remote Multi-airport Control… " on remote multi airport control).

The analysis of critical situations showed that in those situations remembered by the controllers relevant information was derived most often by the far view, mainly in order to verify information perceived by other information sources. Furthermore, the far view provided information that helped to build up an understanding of the actual traffic situation. This topic is also discussed in detail in chapters "Visual Features Used by Airport Tower Controllers: Some Implications for the Design of Remote or Virtual Towers" and "Detection and Recognition for Remote Tower Operations" of this volume. In most cases reported by the controllers, the far view was not the only source for relevant information. Triggers for information search were achieved e.g. via radio communication. The far view was used to timely verify this information. It has to be discussed in how far the video panorama reconstruction of the far view can provide the same quality of information in the sense, that it is not potentially distorted by the sensor technology or data transmission.

In order to gain further understanding about the use of information provided by the far view, one approach would be to apply masks with smaller and a higher number of AOI's to the eye tracking data. In that case, it can be distinguished which functional part of the airport, e.g. runway, final, apron, the controller looked at. Another approach is the application of dynamic AOI (Gross & Friedrich, 2010). This approach seems promising for analyzing eye data gathered in a simulation in combination with traffic and further process data, to understand better the information acquisition process of operators during complex cognitive tasks.

5.3   Benefit of Assistance Tools

One assistance tool investigated in this study was the tracking functionality of the zoom camera. The zoom camera could be attached to the highlighted objects (movement detection or callsign), realizing an automatic tracking. Controllers approved this feature as sensible. There are issues with the reliability of this tracking as well as with the handling. As the study investigated an experimental workplace with advanced functionalities, besides usability problems the operational value of the feature was shown.

Regarding the two types of information augmentation (callsign, movement detection) the callsigns were rated as more reliable and rather correct, compared to the movement detection. Furthermore, controllers claimed the callsigns to be more helpful than the movement detection. In the callsign condition, the controllers were provided with valuable information (the identifier of the AC) directly superimposed on the far view. With this additional information, identification of AC in the approach or on the apron is easier. It is a drawback of the augmentation of callsigns that controllers rely too much on this information, rather than verifying the information derived by radio communication, flight strips and radar by means of the real time information of the far view. This concern was also raised by the controllers in the study. It might have led to rather conservative ratings with regard to the question if the assistance tools provide new possibilities for the work. Besides these subjective ratings, the results of the eye tracking data show that, through the augmentation of callsigns in the far view, visual attention was significantly drawn from the radar display. Yet the augmentation of motion detection results showed the opposite effect. Potentially, in this condition more attention was drawn to the radar, as AC in the traffic pattern or approach were detected earlier than without augmentation. The radar was then used to identify these AC. Nevertheless, in the callsign condition no direct increase of head-up times (higher dwell times on the far view) could be observed. It might be that controllers distributed their "additional" attention quite individually, so no general pattern could be found.

In order to achieve an increase in head-down times as one goal of information augmentation, the callsign of an AC has to be regarded as more relevant than the information provided by motion detection alone. This goal of decreasing head down times has to be seen in relation to the benefit provided by the regular cross-checking between the different information sources and quality of information they provide. Especially flight strip data, that resembles planned data or the expected state of the traffic, should not be mixed with the data resembling the actual traffic situation, like radar and the far view so that controllers effectively can monitor and react upon critical differences between expected and actual state of traffic.

6 Conclusion

Overall, within the present investigation the concept for Remote Control of small low traffic airports showed no significant differences compared to working on a conventional tower simulator as indicated by subjective ratings and usage of information sources. The findings show that the described work environment does not change fundamentally working procedures of tower controllers, supporting the perspective of a medium-term application. In fact, a remote tower solution comparable to the concept under investigation, went operational in Sweden in 2015 (LFV, 2015).

The design of the RTO-CWP as realized in this high-fidelity simulation enables a controller team to successfully handle the traffic of a regional airport. Reconstructing the out-of-windows view of the tower through a high resolution video panorama

proved to fulfil the information needs of the controllers in most cases. There are issues like achievable video resolution and contrast within a reasonable cost frame, which determine the ATC-performance under certain conditions and task requirements. It might lead to changed procedures and maybe capacity reduction in remote control operations. This has to be evaluated in the context of the actual traffic demand at small airports.

With regard to low visibility conditions (night, fog), enhancement of the camera technology towards the infra-red spectrum could even improve the visual information, compared to the conventional tower view. In general, new assistance tools like information overlay in the video panorama and automatic tracking of the zoom camera were rated as promising, given that high reliability can be provided. For operational use of this work place, future work has to deal with questions of redundancy and safety of the system. Nevertheless, to understand which information the tower controller needs at which time is mandatory prerequisite to design such a safe system. The results of this study demonstrate this need, as not every information augmentation introduced in this study, showed the intended effect on reducing head-down times. Obviously, the field of tower air traffic control has received and will receive more attention in the next years. High-fidelity studies are a suitable method to understand the impact of novel concepts on important human factors at a very early stage. The results do not only provide valuable input into the design and further development. High fidelity studies also proved to be a good method to actively involve the operators into the concept development at an early stage.

References

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. Taylor and Francis.

Dickinson, T. L., & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and application* (pp. 19–44). Lawrence Erlbaum.

Dittmann, A., Kallus, K. W., & van Damme, D. (2000). *Integrated task and job analysis of air traffic controllers—Phase 3: Baseline reference of air traffic controller tasks and cognitive processes in the ECAC area* (H. F. a. M. Unit, trans.). Eurocontrol.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1999). Situation awareness as a predictor of performance in en route air traffic controllers (O. o. A. Medicine, trans.). *Aerospace Medicine Technical Reports*. Federal Aviation Administration.

Durso, F. T., Sethumadhavan, A., & Crutchfield, J. (2008). Linking task analysis to information relevance. *Human Factors, 50*(5), 755–762. https://doi.org/10.1518/001872008X312369

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64.

Eurocontrol. (2008). The SHAPE questionnaires. Retrieved April 26, 2010, from http://www.eurocontrol.int/humanfactors/public/standard_page/SHAPE_Questionnaires.html

Fürstenau, N., Rudolph, M., Schmidt, M., & Werther, B. (2004a). *Virtual tower Wettbewerb der Visionen 2001–2004* (pp. 16–21). Deutsches Zentrum für Luft- und Raumfahrt.

Fürstenau, N., Rudolph, M., Schmidt, M., Lorenz, B., & Albrecht, T. (2004b). On the use of transparent rear projection screens to reduce head-down time in the air traffic control tower. In D. A. Vincenzi, M. Mouloua, & P. A. Hancock (Eds.), *Human performance, situation awareness and automation technology: Current research and trends* (pp. 195–200). Lawrence Erlbaum.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., & Halle, W. (2008). Augmented vision videopanorama system for remote airport tower operation. In I. Grant (Ed.) *Proceedings of the 26th International Congress of the Aeronautical Sciences (ICAS).* Optimage Ltd.

Gross, H., & Friedrich, M. (2010). *A toolchain for gaze analysis using dynamic areas of interest.* Poster presented at the HFES European Chapter.

Hilburn, B. (2004). Head-down times in aerodrome operations: A scope study. *Technical report*, Center for Human Performance Research.

Lange, M. (2014). *Information search for decision-making in tower air traffic control: A field study.* Master thesis, Technical University Chemnitz.

LFV. (2015). *Sweden first in the world with remotely operated air traffic management*. Retrieved September 22, 2015, from http://www.lfv.se/en/News/News-2015/Sweden-first-in-the-world-with-remotely-operated-air-traffic-management/

Möhlenbrink, C., Friedrich, M., Papenfuss, A., Rudolph, M., Schmidt, M., Morlang, F., & Fürstenau, N. (2010). High-fidelity human-in-the-loop simulations as one step towards remote control of regional airports: A preliminary study. In *ICRAT 1.-4.07.2010,* Budapest.

Oehme, A., & Schulz-Rückert, D. (2010). *Distant air traffic control for regional airports.* Paper presented at the 11th IFAC, IFIP, IFORS, IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France.

Papenfuss, A., & Möhlenbrink, C. (2009). Kognitive Arbeitsanalyse Lotsenarbeitsplatz. *Institutsbericht No. 112–2009/20*, Deutsches Zentrum für Luft- und Raumfahrt.

Paris, C. R., Salas, E., & Cannon Bowers, J. A. (2000). Teamwork in multi-person systems: A review and analysis. *Ergonomics, 43*(8), 1052–1075.

Peterson, S., & Pinska, E. (2006). *Human performance with simulated collimation in transparent projection screens.* Paper presented at the 2nd International Conference on Research in Air Transportation (ICRAT), Belgrade.

Pinska, E. (2006). *An investigation of the head-up time at tower and ground control positions.* EUROCONTROL Experimental Centre.

Pinska, E. (2007). Warsaw tower observations. In *EUROCONTROL EEC Note* No.02/07. EUROCONTROL Experimental Centre.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Application Symposium.* ACM Press

Schmidt, M., Rudolph, M., Werther, B., & Fürstenau, N. (2006). Remote airport tower operation with augmented vision video panorama HMI. In *Proceedings of the 2nd International Conference on Resarch in Air Transportation (ICRAT 2006)*, Belgrade (pp. 221–229).

Schmidt, M., Rudolph, M., Werther, B., & Fürstenau, N. (2007). Development of an augmented vision videopanorama human-machine interface for remote airport tower operation. In *Proceedings of the Human Computer Interaction International (HCII).* Lecture notes computer science (Vol. 4558, pp. 1119–1128). Springer.

Tavanti, M., & Bourgois, M. (2006). A field study on visual aspects of tower operations. In *EUROCONTROL Innovative Research Workshop 2006* (pp. 105–110). EUROCONTROL Experimental Centre.

van Schaik, F. J., Roessingh, J. J. M., Bengtson, J., Lindqvist, G., & Fält, K. (2010). *Advanced remote tower project validation results.* Paper presented at the 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France.

Vogt, J., Hagemann, T., & Kastner, M. (2006). The impact of workload on heart rate and blood pressure in en-route and tower air traffic control. *Journal of Psychophysiology, 20*(4), 297–314.

Wittbrodt, N., & Thüring, M. (2010). *Communication in future remote airport traffic control centres.* Paper presented at the Human Centered Automation—HFES European Chapter.

Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations

Norbert Fürstenau and Anne Papenfuss

Abstract We report on the analysis of subjective mental workload (WL) and objective task load (TL) measurements of a Multiple Remote Tower Operation (MRTO) simulation experiment with 12 air traffic control officers (ATCos). The experiment was performed as part of a project for the development of a remote tower center (RTC) for the centralized control of several airports (APs) from afar (Fürstenau, Virtual and remote control tower. Springer, Switzerland, 2016). Specifically, we were interested in the question if being responsible for two or more traffic systems at the same time, causes workload independently from actual traffic load. Subjective WL was measured by means of the one-dimensional quasi real time Instantaneous Self Assessment method (five level ISA scale) whereas objective TL data were obtained online by monitoring ATCo's communication with pilots (radio calls frequency RC and duration RD), both dependent on the environmental traffic load n. In addition to variance analysis (ANOVA) for quantifying linear correlations (WL/TL~n) a new cognitive resource limitation model for nonlinear (logistic) regression-based parameter estimates was applied to the data (Fürstenau et al., Theor Issues Ergon Sci, 2020). ANOVA results supported initially stated hypotheses on significant increase of subjective and objective WL/TL measures with increasing traffic flow n, as well as a WL increase under transition from one controller per airport (baseline) to two-airport control by a single ATCo (Lange et al., Analyse des Zusammenhangs zwischen dem Workload von Towerlotsen und objektiven Arbeitsparametern, 2011). Furthermore, a hypothesized mediator effect of communication TL was determined, mediating the dependency of ISA-WL on traffic load n. The extension of the of the (linear) ANOVA by the (nonlinear) logistic model-based analysis of ISA(n) and RC(n) data allowed for the quantification of theoretically founded WL/TL sensitivity (ν/ρ) and bias parameters, the latter characterizing the difference between work conditions. The validity of the regression-based parameter estimates was supported by the theoretical prediction of model parameters based on prior information (e.g. scale limits). Estimates of the nonlinear model parameters quantified the dissociation between the subjective WL and objective communication load measures. Derived from the assumption of

N. Fürstenau (✉) · A. Papenfuss
German Aerospace Center (DLR), Inst. of Flight Guidance, Lilienthalplatz 7, 38108 Braunschweig, Germany
e-mail: norbert.fuerstenau@dlr.de

cognitive resource limitation the logistic model provides a theoretical foundation for the discussion of the initially stated hypotheses regarding WL/TL characteristics. Specifically, a stimulus (RC)—response (ISA) power law analysis according to (Fürstenau and Radüntz, Power law model for subjective mental workload and validation through air-traffic control human-in-the-loop simulation, 2021) allowed via the Stevens exponent $\gamma(=\rho/\nu)$ to formalize and quantify the assumed mediator role of the objective communication TL between traffic flow and the subjective ISA-WL response.

Keywords Multiple Remote Tower · Mental workload · Task load · Air traffic control simulation · Work conditions · Cognitive resource limitation · Instantaneous self assessment · Psychophysics · Logistic model · Stevens law · Parameter estimation · Nonlinear regression

1   Introduction

A major goal of Multiple Remote Tower Operation (MRTO) is cost reduction by increasing the efficiency of (a team of) ATCOs as a resource. As a crucial human factor, the workload experienced in a MRTO work situation compared to a conventional tower environment, was determined to prove the feasibility of this concept. For this purpose a Human-in-the-Loop (HitL) experiment was performed in a specially designed RTC simulation environment for centralized control of two airports. Detailed description of the experiment and two-factor variance analysis (ANOVA) of a subjective workload measure and objective communication task load (TL) measures was published in (Lange et al., 2011).

In the present work we compared the statistical ANOVA results and linear correlations between dependent and independent variables with results of a new nonlinear model based data analysis to explore the relationship between features of the task and environment on subjectively experienced workload. The main goal of this analysis was to quantify the effect of dual tasks (controlling two airports at the same time) on experienced workload through the test of four hypotheses concerning the dependence of the WL and communication (TL) variables on the independent traffic load, and the interdependence of WL and TL. By means of model parameters for WL/TL-sensitivity and bias we show the relevance of considering effects of cognitive resource limitation for quantifying the workload differences under different work conditions.

A quasi real-time one-dimensional WL measure was used for monitoring the subjective workload (Instantaneous Self Assessment, ISA Kirwan et al. 1997) The basic hypothesis to be tested was the plausible prediction that a single controller being responsible for the two airports, on average subjectively experiences higher WL as compared to an ATCO being responsible for the same traffic load at one airport. Considering the simulator design of two panoramic video reconstructions one on top of the other with approach radar and weather displays to be visually

scanned more or less continuously, in addition to radio communication with aircraft (AC) approaching both airports, it seems plausible to predict a significant increase of subjectively experienced WL even with nearly zero traffic.

For this analysis it was of interest to understand, if traffic numbers and communication parameters are valid workload predictors for the MRT work environment. For controlling active MRTs, the authors expected that the ATCOs experience additional switching costs that influence the experienced workload. Within the study, three experimental work-place and task-share conditions were compared and the relative difference of the assessed TL and WL parameters were analysed and interpreted.

ANOCOVAs showed significant correlations through (linear) correlation analysis and F-tests between dependent variables Workload (ISA per 2 min simulation time interval), objective cumulative radio calls duration (RD/2 min), frequency of radio calls (RC/2 min) and independent predictor variables traffic flow n (aircraft/2 min) and work condition c. The results of WL/TL dependency on environmental load (n) were also discussed in terms of a functional ISA(RC(n), RD(n)) dependence, i.e. the workload generating traffic load being mediated through the task load represented by the communication metrics. ANOVA results showed that radio communication (RC, RD) load should be analysed as a mediator between traffic flow (environmental load) and experienced workload (Baron & Kenny, 1986). Moreover, linear ANOVA analysis supported the hypothesis that the communication load may be assumed to be sufficient for the explanation of experienced subjective workload.

The extended nonlinear data analysis was based on a recently suggested formal model (Fürstenau et al., 2020; Fürstenau & Radüntz, 2021) derived from the cognitive resource limitation theory (e.g. Wickens, 2002; Kahnemann et al., 1973; Kahnemann, 2011). Basic assumptions result in simple logistic (sigmoid) functions for the dependency of WL/TL on the environmental traffic flow n (aircraft/time interval), with asymptotic upper load level. Using prior information such as scale limits, domain experts knowledge, average duration of radio calls, the model allows for quantitative predictions of WL/TL sensitivity and offset parameters. These were compared with nonlinear regression based parameter estimates of the data means that provided evidence for the model assumptions. By combination of WL and TL measures through elimination of the common traffic load variable n a formal psychophysical power law stimulus (TL)–response (WL) relationship was obtained with a Stevens exponent of the order of 1, consistent with the typical stimulus—response characteristics. To our best knowledge the first psychophysics approach to workload was suggested by Gopher and Braune (1984) (Gopher et al. 1985) who provided initial experimental evidence by means a task load battery and Stevens power law WL analysis (see Sect. 2.2.5 and Appendix 3 at the end of the book). Recently, Bachelder and Godfroy-Cooper (2019) reported on power law data analysis of an aircraft control HitL simulation experiment that provided power law exponents in the range 0.2–0,4, i. e. the typical order of magnitude of Stevens exponents in (physical) stimulus—(subjective) response experiments. Logistic sensitivity and shift parameters of subjective and objective measures as well the Stevens exponent (see Sect. 4) were used to quantify the results as dependent on traffic and communication load

for the different work conditions and to discuss the hypotheses in comparison to the statistical ANOVA results.

Following this introduction we begin with a brief overview on WL and WL measures in Sect. 2. An extended overview is provided in a separate Appendix 3 of the book. Section 2 is followed by a description of the relevant details of the HITL simulation experiment in Sect. 3. In Sect. 4 we present a summary of the derived theoretical model with equations for parameter prediction and regression analysis of data mean values. A detailed derivation is provided in a separate Appendix D following the main Chapters of the book. Section 5 starts with a review of the initial ANOVA and linear correlation analysis of the measured data (Lange et al., 2011), followed by (generalized) linear (GLM) and nonlinear (logistic) model based regression analysis for quantifying WL/TL parameters. In Sect. 6 ANOVA results and theoretical WL/TL predictions are compared with nonlinear regression based estimates and we discuss the additional insight into the model based work- and task load characteristics within the context of the specific MRTO work conditions. In the Conclusion Sect. 7 besides summarizing the evidence for MRTO WL effects from the different approaches, we also highlight the additional support of the present experiment for the psychophysics power law (stimulus—response) relationship between objective and subjective TL/WL measures. Tables with preprocessed data used for the analysis are provided in an Appendix of this Chapter.

2 Mental Workload and Workload Measures

In this section we present a brief overview on workload, WL measures and models. Additional information can be found in Appendix 3 of the book.

2.1 Definition of Mental Workload

(Mental) workload, as a construct, was developed in the human factors community and gained importance with the rise of automation and is as such strongly related to automation and human–machine-system design. By definition, workload consists of three attributes, input load (or task load TL), operator effort and the performance, or work result (Johannsen et al., 1979). The term can be related to models of human cognition, where it is assumed that human performance is supported by a general pool of mental "effort" or undifferentiated resources (Kahnemann, 1973; Wickens, 2002). Mental workload is characterized by the demand imposed by tasks on human's limited mental resources, e.g. processing capacity and memory (Wickens & Hollands, 2000a, b).

Most important, workload is separated from taskload and separated from stress. Stress is a physiological reaction of the body to a demanding situation, so workload can cause stress, but other factors too, like noise or heat. Taskload is the input, or the

demand from the outside world of the operator, i.e. the task. Here, during the development of a widely used measurement tool the NASA-Task Load Index, it was considered that there is a subjective reaction towards taskload and that multiple dimensions shape the experienced overall subjective load (Hart & Staveland, 1988). Accordingly, the widely used workload scale NASA-TLX assess a subjects response to the six factors physical and mental demands, effort, frustration, performance and time pressure (Hart & Staveland, 1988; Moray, 1982). Whilst taskload can be measured objectively, workload is the subjective reaction of an individual towards this taskload. More precisely, "[…] workload is not merely a property of that task, but of the task, the human, and their interaction (Moray, 1988).

Nevertheless, it is assumed and of interest for human–machine-system designs, to identify a universal task load—workload relationship for a specific task. The challenge is, that any empirical measurement of workload will be an individual response to a given task input. Furthermore, an operator can actively change his or her experienced load by adopting an appropriate speed vs. accuracy operating criterion, if he or she is prepared to accept errors" (Moray, 1988, p. 130). Furthermore, the operator state influences experiences workload, like psychophysical characteristics, personality, experience, motivation, internal tolerated error level (performance) (Johannsen et al., 1979). Recently, the concept of the human performance envelope (Edwards, 2013)) was proposed as a model explaining the impact of human factors on human performance. It is an analogy to the engineering term of the flight envelope. It summarizes nine factors which are supposed to interact which each other and by this can explain observable human performance. In this model, workload is one of the factors, beside situational awareness, stress, team work, communication, fatigue, attention and trust. Recently, research is conducted to understand the interplay of these factors in order to explain, why human operators can perform well even when workload is high and why their performance can drop, even when the workload is not excessive (Friedrich et al., 2018). Summarizing these definitions of workload, it becomes clear that any linear model or prediction based on a single variable is a simplification, for the sake of ease of use during a highly realistic test. The question therefore is to find a model with sufficient predictive power and tolerable effort in terms of measurement and data preparation.

2.2 Operational Approaches and Models to Predict Mental Workload in ATC Based on Task Load

As outlined above, the modelling of a task load—workload function is of importance for the design of human–machine systems. In this section we present a brief overview of existing task load-workload models which are used in the ATC context.

In this domain, it is agreed that traffic load is associated with perceived and psychophysiological workload (Langan-Fox et al., 2009). A basic computational model of workload was proposed by Johannsen et al. (1979) with workload (==

effort) characterized as function of objective task input (load, or TL) and subjective criterions of the operator: Effort = f(Load, Operator-State, Internal Performance Criteria).

In parallel to workload models, indicators and methods to assess operator's mental workload were developed and tested. These indicators or measures can be categorized as either subjective ratings of workload, physiological measures and task load measures. The latter two categories are classified as objective indicators of workload, in contrast to the subjective reports. A lot of research has been conducted and still is conducted to identify robust and reliable objective workload indicators.

2.2.1 CAPAN Model (Sector Capacity Planning for Networks)

We are focusing on models based on task load indictors which are set in relation to subjective workload ratings. There is a broad basis of research and studies which could show the relationship and which revealed a set of objective task load indicators, relevant for the ATC task. For ATC, analyses were conducted to understand and model the impact of different properties of a given traffic situation on workload. This is of special importance, as in the current ATM concept for En-Route traffic, the workload of ATCOs is used to divide or aggregate neighbouring sectors in order to balance the taskload of ATCOs.

Air navigation Service providers (ANSPs) use models to predict mental workload based on expected time needed to fulfil control task per hour (Flynn et al., 2003; Russo, 2016). These estimations are used as a control mechanism to avoid overload situations and to reduce the predicted amount of workload up to a "medium" or controllable level. These models are implemented in real world application and useful, even though the theoretical and computational models still lack the needed reliability and cannot explain all variances in experienced workload.

Eurocontrol's network management, responsible for the configuration of the upper airspace sectors, applies the CAPAN methodology for sector capacity assessment, based on fast time simulation. Here, they are calculating a theoretical sector capacity based on predicted working times of ATCos based on a fast time simulation. Basically, these working times are interpreted as workload levels. In their model, five workload levels are differentiated. The idea is that in each hour the predicted working time should not exceed 42 min (T_R = required time). This resembles 70% of the time available (T_A). This threshold considers that ATCOs should have 30% of their time for tasks not bound to a discrete event (a flight). So, a predicted value of working time above 42 min is considered as an overload situation, which needs to be avoided. In the methodology, fast time simulations of the air traffic within the sector is used to predict ATCOs tasks and consequently summing up the time required to fulfil all basic tasks. In the CAPAN model, workload is the sum of the time durations needed for a basic set of ATCO tasks. So, models of workload are based on models of the ATC task. Table 1 summarizes the basic ATCo tasks and percental share which were observed in seven different ATC sectors.

Table 1 Overview on ATCO's basic tasks with typical percental shares in terms of time spent on tasks, and time required per hour (Russo, 2016)

ATCO task	Average percental share	SD	Average T_R (min)/h
Flight data management	7.9	1.2	4.7
Conflict search	23.4	2.6	14.1
Coordination	4.1	1.6	2.5
Radio communication (RC)	40.0	4.3	24.0
Radar	24.3	8.5	14.6

Additional information on the time pressure aspect of workload can be found in Appendix C of the book.

2.2.2 Radio Communication

The numbers in Table 1 refer to observations conducted in seven different ATC sectors. In all cases, radio communication had the highest overall share on T_R, on average 40% representing a summarized duration of 24 min per hour. Standard radio communication accounted in an example study for 35 to 47% of time needed for conducting the standard tasks and the task always had the highest percental share. This result is in line with a task analysis in the tower environment where radio telephony took 33% of the time (Papenfuss, 2013).

Time needed to communicate with aircraft, as well as number of transmissions on the radio channel correlate with subjective workload ratings (Moray, 1982), as determined by subjective measures, e.g. the Modified Cooper Harper Scale (Casali & Wierwille, 1983). Evidently, radio communication with pilots is conducted by ATCos for managing the traffic, e.g. by responding to clearance requests. ATCO's communication serves for solving of traffic events and situations, beside the other tasks they have to conduct. Nevertheless, time used for radio communication can be measured quite easily as radio transmission occupancy time.

2.2.3 Features of the Traffic Situation

Besides the time required (T_R) for radio telephony, features of the control tasks, the traffic situation, are considered for workload estimations. The term of complexity was introduced to bridge the gap between objective taskload metrics and the experienced workload (Athènes et al., 2002). Recently, machine learning approaches were used to model the relevant factors leading to sector closure and opening, taking into account 27 different complexity factors found in literature (Gianazza, 2017). From the results of a principal component analysis the authors concluded, that there are six main

influencing factors, being the volume of airspace, number of aircraft, incoming flow within the next 15 and 60 min, the average vertical speed of a/c within the sector and geometry of speed vector intersections.

2.2.4   Formal Relationship Between Traffic and Workload

In the CAPAN methodology (Russo, 2016) and the simplified approach (Flynn et al., 2003) regression analysis was used to determine the sector capacity, based on a set of fast time simulations. In both approaches, a quadratic regression is used to model the relationship between aircraft entering the sector (entry rate) and estimated workload. The a priori defined thresholds are used to determine the corresponding traffic numbers *n*.

Lee (2005) suggests that there is no linear relationship between controller workload and traffic count and compares linear, exponential and sigmoid-curved relationships between empirically observed workload ratings and traffic counts in an experimental simulator study. In his study, four controllers with very high familiarity with their sectors took part in a high-fidelity simulation study. Subjective workload was collected every 5 min with Air Traffic Workload Input Technique (ATWIT, (Stein, 1985)) on a scale ranging from 1 (low workload) to seven, representing high workload. In his study, the sigmoid-curve fitted the observed empirical workload ratings best, compared to a linear and exponential fit. The author concludes that the subjective workload is categorial, where sudden jumps from low to high workload can occur. The author also emphasizes, that from an operational point of view the transition from high to unmanageable workload is of importance.

The ATWIT method as a quasi online technique is comparable to the one-dimensional five level Instantaneous Self Assessment (ISA) method used in the present work (for details see Sect. 3.3 and Appendix C at the end of the book).

2.2.5   Psychophysics of Mental Workload

In Sect. 4 we describe a recently developed nonlinear resource limitation model for mental workload (Fürstenau et al., 2020; Fürstenau & Radüntz, 2021) that was used in the present work for the extension of the initial (linear) ANOVA based statistical data analysis. It takes up the psychophysics approach to workload of Gopher and Braune (1984), p. 521. In fact these authors argued that, "[…]if the human information processing system can be assumed to invest […] hypothetical processing facilities to enable the performance of tasks then subjective measures can be thought to represent the perceived magnitude of this investment, in much the same way that the perception of […]" a physical stimulus is changed with variation of its magnitude. Based on laboratory experiments with standardised cognitive tasks they proposed a scaling approach that can be traced back to the psychophysical measurement theory of Stevens (Stevens, 1957, 1975). Psychophysical research aims to describe the relationship between changes in the amplitude of a physical stimulus (S, e.g. brightness,

loudness) and the subjective perception P of these variations. In Stevens' power law the sensation magnitude is a power function of stimulus intensity S described by the constant b and Steven's exponent γ that is characteristic for the type of stimulus with a numerical value of the order of 1:

$$P = bS^{\gamma}$$

Recently (Bachelder & Godfroy-Cooper, 2019) reported on the application of the psychophysics power law to the analysis of a pilot workload estimation experiment. Theoretically predicted Stevens exponents of different tasks were in the range $0.24 \leq \gamma \leq 0.41$ and compared favourably with those obtained from regressions of the data using the power law: $0.21 \leq \gamma \leq 0.37$.

To our best knowledge, only very few experiments like (Bachelder & Godfroy-Cooper, 2019) use a HitL simulation with online measures for a test of the psychophysics hypothesis of mental workload. One advantage of HitL simulation is the possibility to measure workload parameters under realistic conditions for determining parametric interdependence, such as correlation with external traffic load (as simulation scenario parameter) and Radio Communication measures (RC, see above). In (Fürstenau & Radüntz, 2021) we have shown how after nonlinear transformation (P(WL), S(RC)) of WL and RC variables the measured quasi real time WL variable can be transferred into a Stevens type of subjective response P to physical stimulus S, with an exponent related to parameters of the cognitive resource limitation model (see Appendix D of this volume).

3   Experiment and Data Collection

This HitL simulation experiment was the first one to address the new remote control tower paradigm (remote tower operation RTO) of multiple airport traffic control from a special remote control tower center work environment (Remote Tower Center RTC). It was described in detail with initial analysis of WL—TL dependency in an internal DLR-report (Lange et al., 2011). Furthermore, workload data were analysed to understand the impact of workplace design on workload (Möhlenbrink et al. 2011; Möhlenbrink, 2011), as well to identify complexity factors of traffic situations in the tower environment, as addressed in a recent study (Josefson et al., 2018). For details on remote tower operation and the simulation environment see the respective chapters in (Fürstenau, 2016) and in the present volume. In this section we briefly review the relevant details of the experimental procedure and setup.

3.1 Sample and Procedure

The voluntary participant sample consisted of 12 tower controllers (all males) from the German air navigation service provider DFS, working at three different German airports. Their age was between 25 and 60 years (mean = 35, std = 11). They were invited as six pairs (ATCO1, ATCO2), each for two succeeding days and received a financial compensation. None of the controllers was familiar with the remote tower concept or had previous experience with controlling two airports simultaneously.

The experiment run over two days per couple of ATCOs. On the first day participants were introduced to the new concept of remote tower and MRT control. They could also familiarize themselves with the simulation setup and conduct two training runs each. On day two, the empirical data was gathered with altogether eight exercise runs each lasting for 25 min. Three different experimental working conditions were tested, which are described in detail in Sect. 3.3 To control for learning effects, each pair of ATCOs underwent the experiment in a different order of the three conditions.

3.2 Simulation Setup

The simulation setup is shown in Fig. 1. It consisted of a specially designed remote tower operations work environment for control of two (regional German) airports (AP1: Braunschweig, AP2: Erfurt). Both airports were represented by simulations of two 180° high resolution video panoramas consisting of rows of four displays each and arranged one above the other (see previous chapters) and a remotely controlled (simulated) pan-tilt zoom camera (PTZ) replacing the binoculars for each airport. The panoramic views simulated the high fidelity video reconstruction of the far view out of tower windows as experienced by RTO controllers in a remote tower center (RTC). Weather information text was presented as augmented vision overlay on the

Fig. 1 Experimental remote tower center work environment for two-airport control with two video panorama systems representing the two airports, including weather information at outmost left and right displays. At the right. **a** (left photo) On the table flight strips are available. Operator equipped with eye-head tracking device. **b** (right photo): approach radar is displayed at right edge of lower panorama

outmost displays of each AP. Apart from the replacement of a (simulated) video panorama with weather/wind information and PTZ, the RTC console contained the usual tower instrumentation, i.e. flight strips with different bays for AP1, 2 and two radar displays for the two airport control zones. Besides individual weather data also the specific ICAO code of each airport was superimposed on the panorama to ease up identification.

The experimental setup allowed to test and quantify several hypotheses (see Sect. 5.1), in particular that a single ATCO for control of two remote airports experiences higher workload than two ATCOs (baseline condition), each one responsible for one airport. Additionally, both controllers worked in the classical team setup of executive and coordinator. In that case, PL is responsible for traffic control and consequently communicates on a coupled frequency with pilots at AP1 and AP2.

3.3　Experimental Design and Workload Assessment

The initial study was set-up as a one-factorial (working condition) within-design For this data analysis traffic load n was included as a second quasi-independent variable. Three different work conditions were tested, meaning each pair of participants experienced all three working conditions. To control for learning effects, the sequence of working conditions was systematically varied.

- $c = 0$ (baseline condition): Both ATCOs independently control a single AP, ATCO1—AP1, ATCO2—AP2
- $c = 1$: two ATCOs share tasks for dual remote AP control, ATCO1 (responsible for communication with pilots) & ATCO2 (coordination, documentation and monitoring) for AP1 & 2
- $c = 2$: a single ATCO controls two APs simultaneously, new dual AP control condition: ATCO1 for AP1 & AP2

For the resource limitation model with logistic characteristics for analysis of the the workload and communication load data we combine measurements from conditions $c = 0$ and $c = 1$ into a new baseline $c = 0\&1$ (see Sects. 4, 5.2, 5.3).

As independent environmental load variable the number of aircraft n under the responsibility of the ATCOs was measured for each two-minute time interval of the simulation runs (traffic flow, $n = AC/2$ min). Therefore, radio communication data was analyzed and categorized and timestamps for each aircraft were derived. Environmental load n was determined by monitoring start time and AC call sign of initial call by the (pseudo)pilot for each AC and the time of leaving the frequency or final call of respective AC after touchdown and transfer to apron control. Counting the overlapping "active" time slots of all call signs for each 2 min interval yielded n. As further features of the time intervals, run and condition were added. For analysis, time intervals were sorted by equal n and conditions. It needs to be mentioned, that the original study was not designed to systematically vary traffic load but to create

challenging traffic situations. Data analysis revealed a range of one to 10 simultaneous aircraft within the scenarios.

Simultaneously, objective task load data were obtained through measurement of rate of radio calls (ATCO's radio calls RC $(2 \text{ min})^{-1}$) and cumulative transmission duration RD $(2 \text{ min})^{-1}$. RD values are given in % of maximum of $RD_{max} = 120$ s. Cumulative radio transmission duration RD/2 min was used as further load variable. The cumulative transmission time $<RD> = <RCD><RC>$, product of means $<>$ of ATCOs transmission time RCD per call, rate of radio calls RC per 2 min, as average across n_T = number of measured cases of traffic load $n > 0$ per 2 min interval.

Dependent variable: Controllers subjective WL was measured in fixed 2 min time intervals during the simulation run by means of the Instantaneous Self Assessment (ISA) method (Jordan & Brennen, 1992; Kirwan et al., 1997). Every two minutes, the experimenters handed in a print-out of the ISA scale where participants had to point to the value of the ISA scale. This procedure avoided that the participants had to call out their votes and that they did not influence each other in their voting.

ISA is a simple and quasi real time measure for HitL real-time simulations. It is based on a discrete scale of five levels, attributed to the subjectively experienced load during a couple of minutes immediately before the individual rating. It was developed at the Air Traffic Management Development Centre (ATMDC) of the National Air Traffic Services (NATS, UK) (Kirwan et al., 1997; Brennen, 1992; Jordan & Brennen, 1992; Jordan, 1992; Tattersall & Foord, 1996), with the goal of being quick and unobtrusive in order to avoid adding significant WL to the primary task. The following table lists the description of the different subjective ISA levels associated with the corresponding numbers as defined by (Kirwan et al., 1997). In the instructions, the ideal WL level is at 3 (Table 2).

Because the scale levels represent the subjective decision of participants on the experienced load during task execution, the level differences are probably not equidistant. In the theoretical model of Sect. 4 any deviation from linearity is included in the exponential nonlinearities of the model equations.

Table 2 ISA workload categories, after (Kirwan et al., 1997)

Level	WL heading	Spare mental capacity	Description
5	Excessive	None	Behind on task; loosing track of the full picture
4	High	Very Little	None essential tasks suffering. Could not work at this level very long
3	Comfortable Busy Pace	Some	All tasks well at hand. Busy but stimulating pace. Could keep going continuously at this level
2	Relaxed	Ample	More than enough time for all tasks. Active on ATC tasks less than 50% of the time available
1	Underutilized	Very much	Nothing to do. Rather boring

4 Nonlinear Work- and Task Load Model

In this section we provide and briefly describe the equations of a recently developed nonlinear (logistic) resource limitation model of mental workload (WL) which were used for characterizing the experimental ISA-workload and communication task load (TL) data by means of logistic scaling and shift parameters (for details of the model derivation see Appendix D of the book). One major goal of the theoretical model based data analysis is the formalization and improved quantitative testing (Sects. 5.2 and 6) of MRT related hypotheses beyond the statistical (linear) ANOVA in Sect. 5.1.

A motivation for the extension of the usually accepted ANOVA F-test (i.e. p-values as sufficient evidence) by means of theoretical model based approach comes from increasing critique on the use of p-value based conclusions as provided in Sect. 5.1 (e.g. Lambdin, 2012). Main advantage of a plausible theoretical model for WL-data analysis as compared to the linear statistical (ANOVA) approach is the possibility of numerical prediction of model parameters based on prior information (scale limits, traffic characteristics). The present logistic resource limitation models can be used for quantitative theoretical predictions of subjective WL and objective TL sensitivity and bias parameters that characterize the new MRT work environment. In Sect. 5.2 the nonlinear ISA(n) and RC(n) model functions are used for regression based parameter estimates using the experimental ISA and RC means across participants (clustered for equal traffic load), to be compared in the discussion Sect. 6 with the theoretical predictions and with ANOVA results in Sect. 5.1.

Basic assumption for the derivation of our model is cognitive resource limitation (e.g. Kahnemann, 1973; Wickens & Hollands, 2000a, b; Wickens, 2002). It assumes separate limited resources distributed between visual (peripheral, focal), auditory (perceptual), vocal, and cognitive (attention, memory, processing speed) channels which to a certain extent may be treated as independent. The logistic model per-se takes into account this limitation of resources in the form of exponential nonlinearities when approaching these limits, in our case phases of over- und under-utilization (see Figs. 2, 3 in Sect. 4.1, and Fig. 1 in Appendix D of the book).

In previous work (Fürstenau et al., 2020) it was shown for a comparable HitL simulation experiment (approach sector radar control scenarios) with a sample of 21 ATCOs that logistic regression analysis of measured ISA-means (across participants) vs. independent traffic load variable n provided sensitivity and bias parameter estimates with high significance, in reasonable agreement with the theoretical predictions. Although the nonlinear theoretical models can only represent the characteristics of the means across sufficiently large statistical samples it could be demonstrated that the logistic WL-sensitivity parameter is useful as well for clustering of individuals into groups of low and high WL.

In what follows we present a brief overview on the logistic resource limitation model describing the dependence of subjective workload ISA(n) and objective radio communication task load $RC(n)/(2\,min)^{-1}$ (number of calls per time interval between ATCO and pilots) on traffic load n (=aircraft per time interval), and the ISA(RC) power law including numerical predictions of model parameters (for mathematical

details see Appendix D at the end of the present volume). The validation of the psychophysics power law (Stevens law) for the stimulus (TL)—response (WL) relationship between suitably normalized and transformed ISA and RC variables was published recently (Fürstenau & Radüntz, 2021). Here we focus on those basic equations which are used for theoretical prediction of characteristic WL/TL parameters (see Table 3) and for regression based parameter estimates of the measured ISA(n) and RC(n) means in Sect. 5.2 (see tables of pre-processed data in the appendix of the present chapter).

For the model based parameter prediction and regression based parameter estimates we define the combined data of conditions $c = 0$ and $c = 1$ as baseline (see Sect. 3.3 for definitions) in order to improve the statistical quality of parameter difference with respect to single operator MRT control of two AP ($c = 2$). In contrast the initial ANOVA analyses of (Lange et al., 2011) as summarized in Sect. 5.1 was separated for the three conditions.

4.1 Logistic Mental Workload Model ISA(n)

For the present model the ISA measure (like the communication variable RC(n), see below) is treated as a continuous variable that was discretized for the purpose of subjective reporting. The nonlinear logistic WL model (sigmoid functions, see Figs. 2, 3) takes into account the ISA-WL scale limits with exponential asymptotic approach to the upper limit as given by:

$$I(n) = \frac{I_u}{1 + \exp\left\{-\frac{n-\mu}{\nu}\right\}} = \frac{5}{1 + k \exp\left\{-\frac{n}{\nu}\right\}} \tag{1}$$

Using prior knowledge on the ISA scale (ISA$_{max}$: $_= I_u = 5$, $I_{d} = 1$) and operational parameters (domain experts information on operational traffic flow range and capacity n_c), allows for theoretical prediction of model parameters (see Table 3). The scaling parameter ν quantifies the exponential convergence towards the upper and lower asymptotes. It also characterizes the maximum slope I' $= dI/dn = I_u/4\nu$ at inversion point $n = \mu$ with $I(\mu) = I_u/2$. It is easily verified that lim ISA($n \rightarrow \infty$) = I_u and lim ISA($n \rightarrow -\infty$) = 0 (with n < 0 as mathematical design aspect with no real meaning). Because ISA scales' lower limit is $I_d = 1$, we assume the intersection with $n = 0$ (zero traffic) at ISA(n = 0) = I_{min}: $_= I_d = 1$. This yields $k_{th} = I_u/I_d - 1 = 4$ and $\mu = \nu \ln(k)$ that relates shift parameter μ to scaling parameter ν which reduces Eq. (1) to a 1-parametric model for the baseline scenario with ATCOs controlling single APs:

$$I(n) = \frac{5}{1 + 4 \exp\left\{-\ln(4)\frac{n}{\mu}\right\}} \tag{2}$$

Table 3 Summary of model parameters, their meaning and numerical (theoretical) predictions (see Sects. 4.1–4.3)

| Model variables and parameters	Description	Num. prediction (experimental condition, independent variable factor 2: $c = 0$ & 1, 2)	
		Baseline: $c = 0$ & 1, two ATCos, 2 airports	MRT: $c = 2$, one ATCo, 2 airports
n [AC/2 min]	Traffic load (independent variable, factor 1)	0 …10	0…10
n_c [AC/2 min]	Nominal capacity limit (prior info)	7	6 (plausibility)
ISA workload	**Logistic model (I_u, I_d, μ, ν)**		
I	ISA WL measure, 5 levels	1…5	1…5
I_u	Upper scale limit (model, asymptote)	5	5
I_d	Lower scale limit (def./prior info)	1	2 (plausibility)
I_0	Lower asymptote (model)	0	0
μ [AC/2 min]	Shift (sigmoid inversion point)	5	def:: = ν ln(k) = 1.46
ν [AC/2 min]	Scaling ν = μ/ln(k)	3.61	:= ν(c = 0&1) = 3.6
k	exp(μ/ν) = I_u/I_d − 1	4	1.5
dI/dn = I′(n = μ)	Slope at inversion = $I_u/4ν$	0.347	0.347
p_I	Normalized ISA: = I/I_u	0.2…1	0.4…1
$dp_I/dn = p_I′$	Normalized slope (sensitivity)	0.069	0.069
P_I	Transformed: = $p_I/(1 − p_I)$ = $I/(I_u − I)$	0.25…∞	0.67…∞
y_I	Log-linear: = ln(P_I)	−1.39…	−0.40…
a_{gI} (or a_{gt})	Slope = $1/ν$[AC/2 min]$^{-1}$	0.0277	0.0277
b_{gI} (or b_{gt})	Intersection = − ln(k)	−ln(4) = −1.3863	−ln(1.5) = −0.405
RC Task Load	**Logistic model (R_u, R_d, ρ)**		
<RCD> [s]	Radio call duration, <> = average	5	5

(continued)

Table 3 (continued)

Model variables and parameters	Description	Num. prediction (experimental condition, independent variable factor 2: $c = 0$ & 1, 2)	
		Baseline: $c = 0$ & 1, two ATCos, 2 airports	MRT: $c = 2$, one ATCo, 2 airports
RC or R [calls/2 min]	ATCO's radio calls RC per 2 min	0…10	0…8
RC_u or R_u [calls/2 min]	Upper limit (asymptote; also RC_u)	10–11	8
R_d [calls/2 min]	Lower limit ($n = 0$)	0	0
R_0 [calls/2 min]	Lower asymptote (model)	-10	-8
μ [AC/2 min]	shift (inversion point at $\{0, 0\}$)	Def: $= 0$	Def: $= 0$
ρ [AC/2 min]	Scaling $\approx n_c/2$	3.5	3
$dR/dn = R'(n = 0)$	Slope (max.) at inversion $= R_u/(2\rho)$	$10/7 = 1.4$	$8/6 = 1.3$
p_R	Normalized RC: $= R/R_u$	0…$p_R(n = 10) =$ $R(10)/10$	0… $p_R(10) =$ $R(10)/8$
$dp_R / dn = p_R'$	Normalized slope (sensitivity)	0.14	0.16
P_R	Transformed: $= (1 + p_R)/(1 - p_I) = (R_u - R)/(R_u - R)$	1…$P_R(n = 10) =$	1…$P_R(n = 8) =$
y_R	Log-linear: $= \ln(P_R)$	0…$\ln\{P_R(10)\}$	0…$\ln\{P_R(8)\}$
a_{gR}	Slope (inversion) $= 1/\rho$ [AC/2 min]$^{-1}$	0.29	0.33
ISA(RC) power law	**Stimulus R(n) – response I(n) (Psychophysics) Stevens law**		
γ	Stevens' Exponent $= \rho/\nu$	0.97	0.83

If we select for the shift parameter μ the center of the nearly linear range around the inversion point of the logistic characteristic as definition of operational traffic flow, i.e. center between zero load $n = 0$ (with $I_d = 1$), and e.g. $n_{max} = 10$ AC/2 min as high or excessive load (domain experts info), i.e. $\mu := \mu_0 = 5$ AC/2 min, we obtain a theoretical value for the scaling parameter under baseline conditions $\nu = \mu /\ln(4)$ = 3,606 (AC/2 min)$^{-1}$, with slope at inversion point (linearized load sensitivity) I' $= I_u/4\nu = 0.347$ (ISA increment per 1 AC/2 min traffic flow increase).

While for underload (e.g. $n < 3$ AC/2 min) under baseline conditions ($c = 0$&1) it seems plausible that $\lim I(n \to 0) = I_d = 1$, for $c = 2$ in contrast it appears plausible

Fig. 2 Theoretical workload characteristic according to logistic resource limitation model for ISA(n) predicting average behaviour (means across participants). Solid line: condition $c = 0$ & 1 (baseline); $\mu = \mu_0 := 5$, $\nu = \mu \ln(4)$. Dashed line ($c = 2$, single ATCo controlling 2 AP). Assumption: doubled WL-offset for low n due to doubled visual scanning (intersection at ISA = 2, scaling parameter $\nu(c = 2)$: $= \nu(\mu_0)$). For details see text
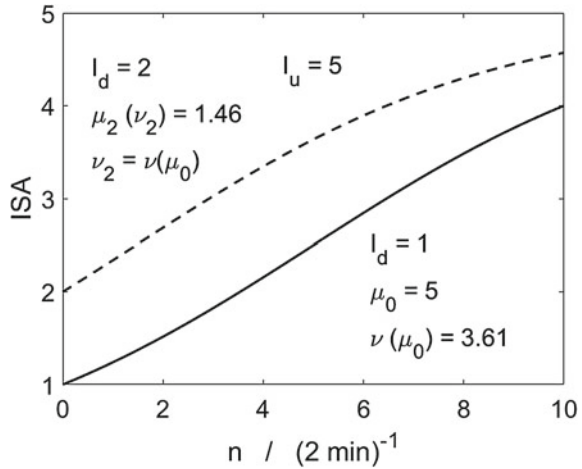
that with low traffic the single ATCo (subjectively) experiences doubled WL with dual AP surveillance: scanning two video panorama systems instead of only one, with only little radio contact at low n (Wickens & Hollands, 2000a, b; Wickens, 2002). Consequently we assume a theoretical offset of lim ISA(n → 0, $c = 2$) = $I_{min}(c = 2) = 2$, $I_{min}(c = 0$&$1) = 1$, resulting in k($c = 2$) = $(5/2 - 1) = 1.5 = \exp(\mu/\nu)$.

For $c = 2$ as a plausible guess we assume for the scaling parameter the same exponential approach to subjective cognitive resource limit ISA = I_u, i.e. $\nu(c = 0$&$1) = \nu(c = 2) = 3.61$ so that we obtain as new shift parameter $\mu(c = 2) = \nu \ln(1.5) = 1.46$. The following Fig. 2 depicts the theoretical prediction for the two ISA(n) characteristics of baseline and $c = 2$ condition.

Both characteristics exhibit only weak nonlinearities within the operational ranges centered around shift values $\mu(c = 0$&$1) := 5$ and $\mu(c = 2)$. However, for condition 2 (single ATCO controls two APs) a significant nonlinearity of the $c = 2$ characteristics is now predicted for $n > \mu_0$. If we assume the nearly linear part of the logistic characteristic between the exponential sections of the sigmoid to be sufficiently represented by a linearized approximation (for derivation see Appendix D of the volume), the corresponding slope and intercept parameters $\{a_l, b_l\}$ are approximated by $dI/dn|_m \approx a_{l} = I_u/(4 \nu) = I_u \ln(4)/(4 \mu) = 0.347$ and intercept $b(n = 0) = b_{l} = I_u/2 (1 - \ln(k)/2) = 0.7671$ (prediction for baseline condition ($c = 0$&1), valid in the vicinity of the logistic shift parameter $\mu = \mu_0 = 5$. For baseline conditions ($c = 0, 1$) $k = 4$, for single controller condition ($c = 2$ with offset $I_{min} = 2$) $k = I_u/I_{min} - 1 = 1.5$. The traffic flow at the mean of the ISA-range $I_{1/2}$ is then given by $n_{1/2} = \mu/\ln(I_u - I_d)[1 + \ln\{(I_u + I_d)(I_u - I_d)/2I_u\}] = 1.353\mu = 6.765$ with $I_{1/2} = (I_u + I_d)/2 = 3$.

Transformation of the ISA variable I(n) into the normalized exponential form $P_I = p_I/(1 - p_I) = I(n)/(I_u - I(n)) = k^{-1} \exp(n/\nu)$, with $I(n)/I_u: = p_I$ allows for linear regression methods with the generalized linear model (GLM) in semilog coordinates:

$$y_I(n) = \ln\left[\frac{p_I}{1 - p_I}\right] = \frac{n - \mu}{\nu} = \frac{1}{\nu}n - \ln(k) \tag{3}$$

with theoretical slope for baseline condition (for transformed ISA variable $y_I(n) = \ln(P_I)$ in semilog coordinates) $a_g = 1/\nu = \ln(4)/\mu = 0.0277$ (AC/2 min)$^{-1}$ and predicted intersection $b_g = -\mu/\nu = -\ln(k) = -1.3863$.

## *4.2 Logistic Model for Rate of Radio Calls RC(N)*

Like for the ISA-WL measure we can define a logistic (or sigmoid) characteristic and predict theoretical parameter estimates (see Table 3) for ATCo's radio communication activity (radio calls RC/2 min), based on plausibility assumptions and prior information. It appears plausible to assume a linear growth with n at low traffic (i.e. inflexion point $n = \mu = 0$), with exponential approach to the asymptote R_u representing the maximum possible call rate of calls. Using average radio call duration of <RCD> = 5 s/call as prior information (see Table 5, Sect. 5.1, from (Lange et al., 2011)) and assuming equal call duration of ATCO and pilot with minimum interval between calls = 1 s we obtain an upper asymptotic limit of frequency (rate) of radio calls as $R_u = 120$ s/$(2 \times 5 + 1)$ s $= 10\ldots11$ calls/2 min for baseline condition $c = 0$ & 1 (2 ATCos for 2 AP). The logistic function model with sensitivity parameter ρ and shift parameter $\mu = 0$, i.e. maximum slope at inflexion point $(0, 0)$ is represented by:

$$R(n) = RC_u \tanh\left(\frac{n}{2\rho}\right) \tag{4}$$

with traffic flow $n \geq 0$. $\tanh(x)$ alternatively may be written in shifted logistic form $\left[2/(1 + \exp\{-2x\}) - 1\right]$ depicting the exponential approach to RC_u with scaling ρ.

For $c = 2$ (single ATCo with dual AP control) it appears reasonable to assume reduced communication activity with higher traffic and a reduced upper limit RC_u due to doubled visual scanning tasks for two AP panorama and radar displays. According to (Sperandio, 1978) a strategy change of single ATCo under transition from normal (operational) traffic to traffic peak situation ($n >$ nominal capacity n_c ($c = 0$ & 1) ≈ 7, the latter obtained from domain experts knowledge) may be assumed: e.g. if two AC arrive simultaneously at both AP, for $c = 2$ the ATCo will delay (put into holding) one for decreasing communication load (basically through rate of calls) in order to gain time for controlling final approach and landing of the other AC and for additional tasks (visual scanning of videopanorama, flight strips, weather display). Also it appears plausible to assume for $n_c(c = 2) < n_c(c = 0 \& 1) := 7$ because for the latter conditions two ATCos share the task load of controlling two APs. As initial guess for single operator 2-AP control we set $n_c(c = 2) := 6$. It appears plausible that for $c = 2$ besides the operational traffic capacity limit n_c, also the asymptotic

Fig. 3 Theoretical RC(n) characteristics for baseline condition c = 0 & 1 (solid curve, 2 ATCos for 2 AP), and for single ATCo condition ($c = 2$, dashed curve). For The asymptotic value RC_u we use the prior estimate of 10 calls/2 min for baseline and $RC_u = 8$ for $c = 2$. Scaling parameter estimate ρ calculated from linear extrapolation using operational traffic capacity limit (for details see text)
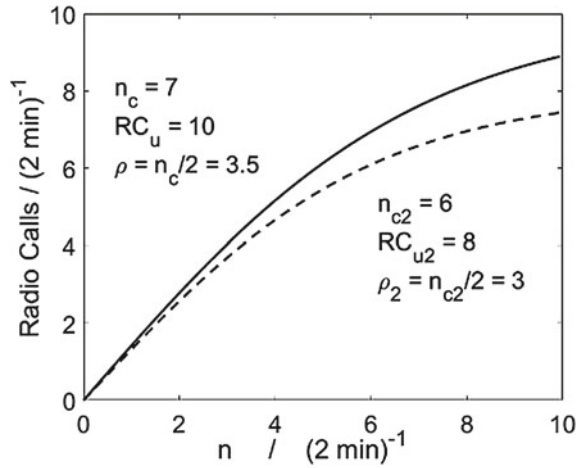
limit RC_u is reduced (corresponding to the hypothesized increased ISA bias $I_{min}(c = 2)$, i.e. $\Delta ISA/I_u = 1/5$), with $R_u(c = 2): = 2/10 = 8$ selected as initial guess. The following Fig. 3 depicts the two theoretical RC(n) curves according to Eq. (4), for baseline ($c = 0 \& 1$) and $c = 2$.

Significant deviation from the nearly linear section for the baseline model ($c = 0 \& 1$) starts for $n > 6$. Also, a significant difference between the baseline characteristic and the $c = 2$ condition (single controller for 2 APs) is predicted for $n > 6$, i.e. for the highest traffic load level only used in the simulation experiments.

The maximum value of the slope $R'(n) = dR/dn$ at $n = 0$ is derived as $R' = R_u/2\rho \approx \Delta R/\Delta n$ (for mathematical details see Appendix 4 of the volume). With linear extrapolation of the maximum slope (tangent of Eq. (4) at $n = 0$) $\Delta p_R = \Delta n/2\rho = n_0/2\rho$ (with $p_R = R/R_u$, $\Delta R(n_0) = R_u$, $n_0: =$ operational capacity limit under baseline ($c = 0 \& 1$) conditions $= n_c = 7$, and $R_0' = \Delta R(n = 0)/n_0 = R_u/n_c$) this yields as rough theoretical estimate for the scaling parameter $\rho : \approx R_u/2R_0' = n_c/2 = 3.5$ (and $\rho(c = 2) = 3$ for single operator condition).

The generalized linear model is again obtained through transformation of R(n) into a normalized function $P_R(p_R)$. Via definition of the transformed and normalized radio-contact variable $P_R = (1 + p_R)/(1 - p_R)$ we obtain the explicit exponential relationship $P_R = \exp(n/\rho)$ in correspondence to the ISA characteristic $P_I = 1/k \exp(n/v)$ and in semilog coordinates:

$$y_R(n) = \ln\left[\frac{1 + p_R}{1 - p_R}\right] = \ln\left[\frac{R_u + R(n)}{R_u - R(n)}\right] = \frac{1}{\rho}n \qquad (5)$$

with GLM-slope $a_R: = 1/\rho = 0.29$ as taskload sensitivity parameter for baseline condition.

4.3   Power Law Model ISA(RC)

With lower limits $I(n=0) = I_{min} := I_d = 1$, $RC(n=0) = 0$ the normalized nonlinear ISA(RC) characteristic $p_I(p_R)$ with $p_I = I/I_u$, $p_R = RC/R_u$ is obtained as a 2-parametric model with $\gamma = \rho/\nu$ and $\mu/\nu = \ln(k)$ (see Eq. (3)) by combining Eqs. (1) and (4):

$$p_I(p_R) = \frac{1}{1 + k\left[\frac{1-p_R}{1+p_R}\right]^\gamma} \tag{6}$$

Fig. 4 Prediction of power law with Stevens parameter γ. Solid line: baseline load case $c = 0$ & 1 (2 ATCOs for remote surveillance of 2 AP) with $I_{min} = 1$, $R_u = 10$; dashed line: $c = 2$ condition with $I_{min} = 2$, $R_u = 8$. b) dashed line for $R_u = 8$ calls/2 min
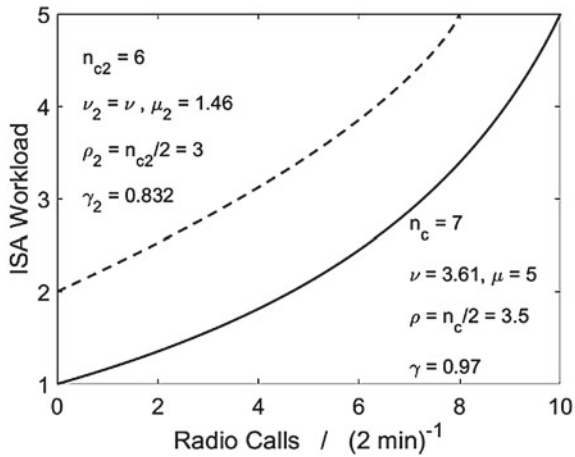
With $k = k_{th} = 4$ for the nominal case ($c = 0$ (and 1)) this is reduced to a model with power γ as the single free parameter. A quantitative estimate for the exponent is obtained from the ratio of logistic sensitivity parameters: $\gamma = \rho/\nu \approx 3.5/3.6 = 0.97$, i.e. a value of the order of 1 consistent with typical Stevens exponents of the psychophysics stimulus – response power law (Fig. 4).

Like for the ISA(n) analysis the nonlinear iterative least squares (LSQ) parameter estimate of experimental data may be compared with a linear regression or maximum likelihood parameter estimate using the generalized linear relationship that is obtained from Eq. (6) after p_I and p_R transformation into the dimensionless functions $P_I(p_I)$, $P_R(p_R)$ as

$$y_I = \ln(P_I) = \gamma \ln(P_R) - \ln(k) = \gamma y_R + b_g \tag{7}$$

with $b_{gt} := -\ln(4)$, i.e. a 1-parametric model for the nominal case ($e = 0$ & 1). The linear increase of the subjective WL function $y_I(ISA)$ with increasing radio communication variable $y_R(RC)$ due to increasing traffic n is characterized through the single psychophysics parameter γ via $y_I(ISA(RC(n)) \mid \gamma)$, and it may also be

written in the well-known (Stevens) power law form $P_I = \frac{1}{k} P_R^\gamma$ with γ of the order of 1 in agreement with typical psychophysics parameters of sensory modalities.

5 Experimental Results

We begin in Sect. 5.1 with a review of the initial ANOCOVA based data analysis, with factors traffic flow and working condition, providing results of F-tests and correlations for initial tests of four hypotheses (Lange et al., 2011). In Sect. 5.2 we extend this evaluation by logistic model based regression analysis using the means across subjects (within fixed traffic levels n) and averaged over work conditions as well as separated for conditions.

5.1 ANOVA Data Analysis: General Linear Model and Multivariate Regression

The data analysis by means of ANOVA was intended to provide initial answers to four hypotheses:

1. ATCOs are expected to provide increased subjective WL ratings with increased traffic flow n (AC per time interval). With the goal of clarifying the WL for the new multiple remote AP control from a center, the three experimentel conditions (Sect. 3.3) were analyzed with the hypothesis that $WL(c = 2) > WL(c = 0$ and $c = 1)$
2. The cumulative duration of ATCos radio call duration per 2 min interval and number of radio calls RC was expected to increase with traffic flow n and again tested for the three conditions.
3. The relationship between WL and cumulative transmission time RD (per 2 min interval) and radio calls RC was expected to exhibit a positive correlation corresponding to results of Manning et al. (2002)
4. Cumulative RD and RC were expected to be sufficient for predicting ATCOs subjective WL. Due to the assumed dependency RD(n), RC(n) a parametric dependence (mediator effect of n) and corresponding correlation of WL(RD, RC) was hypothesized.

Table 4 lists scale limits as prior knowledge that is used for the data analysis.

Figure 5a and b depict overviews on the number N of measurements available for the analysis as derived from the 2 min simulation time intervals. They are ordered for increasing traffic load n with $n > 0$ (independent environmental load variable), and for reported ISA levels respectively (with the same numbers of measured communication data RD, RC). The detailed data set is provided for n in Table A1 of Appendix to the present chapter (separated for conditions c: N(n, c) and for (ISA, RC, RD) in Table A2 (averaged over conditions).

**Table 4** Scale limits as prior knowledge of independent and dependent variables

Variable (units)	Min	Max (theoretical)	Max (empirical)
Independent variable Traffic flow n (AC/2 min)	0	–	10
Subjective ISA-WL ratings (1, 2, 3, 4, 5)	1	5	5
Radio call rate RC (calls/2 min)	0	10–11	10
Cumulative call duration (per 2 min) RD/s	0	120	63
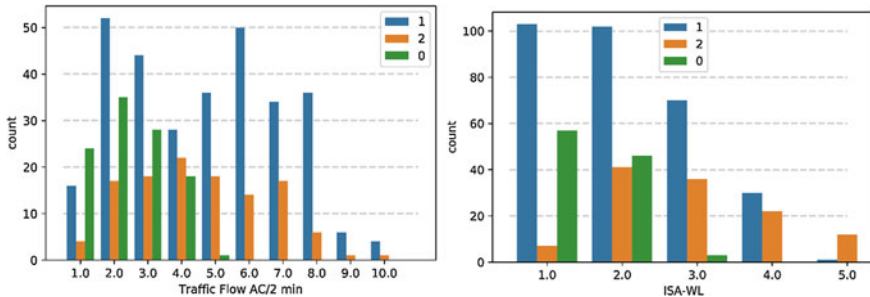
Fig. 5 Histograms, visualizing differences of frequencies of 2 min intervals (number of data points, separated for experimental condition c). **a** (left) independent environmental variable traffic flow n/AC/2 min; **b** (right) dependent variable ISA workload (with the same count applicable to RC, RD)

As shown in Fig. 5 and the corresponding Table A1 in the Appendix only small numbers N of 2 min measurement intervals (cases) were available for extreme traffic flow $n = 8$, 9 and 10 AC/2 min and for $n = 5$ under baseline $c = 0$. In order to improve the statistical comparability of conditions for the ANOVA, for $c = 1$ and $c = 2$ only those traffic flow values n were used where at least ten 2 min intervals were available ($N(c) > 10$). The measurements at $n = 5$ for $c = 0$ and $n = 9$, 10 for $c = 2$ were defined as outliers and omitted for the ANOVA because they are represented by one 2 min-interval only (Appendix, Table A1). This eliminated the traffic flows $n = 0$, 1, 8, 9, 10, yielding a reduced data set of 325 cases for the analysis. So, for the ANOVA test of the basic hypotheses the measurements of dependent variables ISA, RD, RC for six external load levels $n = 2, \ldots, 7$ with conditions $c = 0, 1, 2$ were employed. In contrast, for the model based regression analysis of means in the following Sect. 5.2 we used the full n-range $1 \leq n \leq 10$ according to Tables A2 and A3 in the Appendix.

Table A2 in the Appendix collects the means across $N(n)$ cases for the measured traffic flow range $n = 1\ldots10$ AC/2 min + for ISA-WL, radio calls per 2 min RC, and cumulative transmission time per 2 min as relative value RD/120 s (=time pressure $= 1$ for RD = 120 s) as average across the three conditions. Table A3 lists the measured ISA, RC and RD means across N separated for n (like in Table 2), however now separated for conditions. Figure 6 depicts the ISA mean values with standard
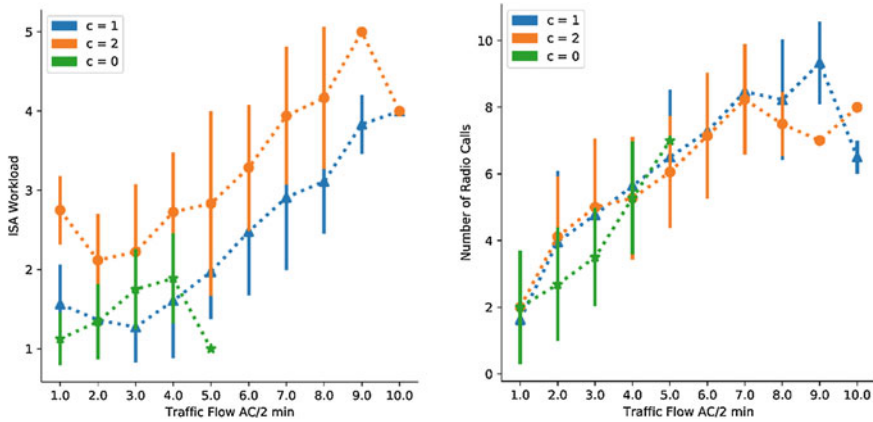
Fig. 6 a Measured ISA Workload (ordinate) as means <ISA> (across all reports per 2 min intervals) with stdev according to Table A3 versus respective traffic flow n (AC/2 min = AC on ATCO communication frequency, abscissa), separated for the three conditions $c = 0, 1, 2$. **b** corresponding radio calls (RC/2 min vs. n)

deviations for each of the three experimental conditions $c = 0, 1, 2$ according to Table A3 (Appendix).

With the exception of the ISA(n) outlier at $n = 5$, for all three conditions the ISA and RC means (across number of 2 min intervals according to Table A3 in the Appendix) the figure clearly depicts the expected increase with n. Furthermore, the theoretically predicted doubling of the subjective WL-level more or less independent of traffic (see Sect. 4.1) is indicated by the ISA(n) results through the evident shift of $c = 0, 1$ data (in the following Sect. 5.2 used as combined baseline $c = 0$ & 1) by $\Delta ISA \approx 1$ into $c = 2$ data. In contrast the RC(n) data exhibit hardly any significant discrimination between baseline and $c = 2$ condition. If at all, only a weak $c = 1$ versus $c = 2$ separation is indicated for large RC(n) values, and in agreement with the theoretical prediction with R_u decreased for the single controller condition $c = 2$ (see Fig. 3).

ANOCOVAs of the data provided information on significance of hypothesized linear relationship (increase) of dependent subjective WL variable ISA on traffic flow n as environmental load variable and work condition c: (hypothesis H1). A corresponding analysis was performed on the hypothesized increase with (n, c) of objective communication load variables RD and RC (hypothesis H2). The correlations ISA(n), RC(n), RD(n) suggested corresponding ANOCOVA based tests of the hypothesized positive correlation between subjective ISA-WL and objective task load variables (H3): ISA(RD), ISA(RC). As hypothesis H4 the sufficiency of communication variables RD, RC for determining ISA-WL, mediated through the common parametric dependence on traffic flow n was investigated.

Evidence for H1 was provided by the ANOCOVA results which analysed subjective ISA-WL as dependent on traffic flow n and condition c:

A two-factor ANOVA for the ISA(n) WL dependency (independent variables n, c) with the participants as covariates showed a significant (linear) increase of WL with n: $F(5, 309) = 38.4$, $p < 0.01$, $\eta^2 = 0.38$. The corresponding linear correlation separately for each condition yielded: $c = 0$: $r = 0.37$, $c = 1$: $r = 0.69$, $c = 2$: $r = 0.57$; on average $r = 0.64$, that appears to provide some evidence for the linear model assumptions. The 2-factor ANOCOVA for the mean ISA dependence on condition c showed a significant increase with c from baseline $c = 0$ to $c = 1$ and to $c = 2$ (single ATCO/2-AP control): $F(2, 309) = 37.6$, $p < 0.01$, $\eta^2 = 0.2$).

Evidence for H2 was provided by the 2-factor ANOCOVA results which analysed the variation of the cumulative radio transmission duration RD and the number of radio transmissions or ATCO's calls RC per 2 min interval on (n, c). As a result RD(n) increased significantly, with $F(5, 309) = 34.4$, $p < 0.01$; $\eta^2 = 0.36$, and RD ~ n correlations separated for conditions: $c = 0$: $r = 0.51$, $c = 1$: $r = 0.68$, $c = 2$: $r = 0.55$; total $r = 0.69$. The ANOCOVA also showed significant increase of RD with condition: $F(2, 309) = 16.7$, $p < 0.01$; $\eta^2 = 0.09$), specifically means of RD($c = 2$) > RD($c = 0$), RD($c = 1$) > RD($c = 0$).

Also for RC, the ANOCOVA indicated highly significant increase with n, quantified by $F(5, 309) = 31.5$, $p < 0.01$, $\eta^2 = 0.34$), with correlations RC ~ n: $c = 0$: $r = 0.49$, $c = 1$: $r = 0.64$, $c = 2$: $r = 0.58$, total $r = 0.67$. Comparison of RC means for different conditions (Bonferroni post hoc test) showed significant increase of RC with condition $c(F(2, 309) = 8.01$, $p < 0.01$; $\eta^2 = 0.05$), specifically from baseline ($c = 0$) to $c = 1$ ($p < 0.01$) as well as to $c = 2$ ($p < 0.01$).

In Table 5, mean values of measured RC, RD variables and average call duration RCD = RD/RC (together with measured standard deviations and std errors = stdev/$\sqrt{N}$) are collected, separated for conditions.

For the test of hypothesis H3 concerning increase of ISA-WL with communication variables RD and RC the complete dataset of 405 cases (2 min intervals) was used (i.e. including traffic flow values $n = 0, 1, 8, 9, 10$). The measured RD values (per 2 min = 120 s: $0 \leq RD \leq 63$ s) were collected into eight RD intervals, from 1 to 7 s (56 cases) to 56–63 s (8 cases) and related to the corresponding average ISA-WL values. The 2-factor ANOVA with independent variables (RD, c) showed a significant ISA increase with RD ($F(7, 382) = 17.3$, $p < 0.01$; $\eta^2 = 0.24$) with correlations $r(c = 0) = 0.50$, $r(c = 1) = 0.58$, $r(c = 2) = 0.53$, and total $r = 0.61$. As expected also the

Table 5 Averages of RD, RC and RCD = RD/RC (=call duration per radio call transmission (($\pm$ stdev; sterr) across n, separated for conditions

Condition c	N	<RD>/(s per 2 min) (stdev; sterr)	<RC> (2 min)$^{-1}$ (stdev; sterr)	<RCD>/s (sterr)
0	89	15.80 (8.53; 0.9)	3.55 (1.88; 0.2)	4. 5 (0.4)
1	128	30.63 (12.62; 1.1)	6.11 (2.46; 0.2)	5.0 (0.3)
2	108	29.34 (11.28; 1.1)	5.86 (2.29; 0.2)	5.0 (0.3)
Total (sterr)	325	25.3 (1.1)	5.2 (0.2)	4.8 (0.2)

factor c again had a significant effect on ISA-WL ($F(2, 382) = 22.3, p < 0.01; \eta^2 = 0.1$).

Also for the ANOVA of ISA-WL(RC, c) the complete dataset with 405 cases (2 min intervals) was used. Number of calls/2 min covered the range $0 \leq RC \leq 12/2$ min, with mean ISA-WL range $1.0 \leq ISA \leq 3.0$. WL increase with RC was significant with $F(12, 371) = 10.6, p < 0.01; \eta^2 = 0.26$ and correlations $r(c = 0) = 0.41, r(c = 1) = 0.60, r(c = 2) = 0.49$, and total $r = 0.60$. The 2-factor ANOVA with (n, c) as independent variables showed a significant effect of c with $F(2, 371) = 29.2, p < 0.01; \eta = 0.14$. The Bonferroni test indicated significant differences between the three conditions.

Concerning hypothesis H4 a mediator effect of traffic flow n on the ISA(RD, RC) correlation was analyzed with reference to (Baron & Kenny, 1986). By means of correlation analysis according to these authors a small mediator effect was found for the ISA(RD(n)) as well as ISA(RC(n)). Conclusions concerning the sufficiency of communication variables RC and RD on WL were drawn from the comparison of correlation coefficients r_I(ISA, n), r_C(ISA, RC), r_D(ISA, RD), and WL averages <ISA> (across traffic load n), separated for conditions, as listed in Table 6.

The correlation coefficients are in the "high" range according to Cohen's definition and support the observed ISA (n) increase for $c = 1$ and $c = 2$, as depicted in Fig. 6. For baseline $(c = 0)$ the correlation is small due to the traffic load limited to the low n range with only three levels $(n = 2, 3, 4)$ where the data indicate decrease of slope as predicted by the theoretical ISA(n) model for low traffic n in Fig. 2. The WL means <ISA> across n increase from baseline condition to single ATCO condition $c = 2$ (significant within stderr of means), although evidently the low <ISA> $(c = 0)$ is mainly due to the low n range. Means <ISA> across RD and <ISA> across RC are not significantly different (within sterr) from the mean < ISA > across n in Table 6. The regression analyses below confirm a significant ISA offset difference between combined conditions $(c = 0, c = 1$: dual operator control used as baseline condition), and single operator $(c = 2)$ as predicted also by theoretical considerations in Sect. 4.1.

**Table 6** Summary of ANOCOVA correlation coefficients and ISA means across n ($\pm$stderr)

Condition c	N	r_I(ISA, n)	n range	r_C(ISA, RC)	r_D(ISA, RD)	<ISA> (sterr)
0	89	0.37	2–4	0.41	0.50	1.6 (0.06)
1	128	0.69	2–7	0.60	0.58	2.1 (0.08)
2	108	0.57	2–7	0.49	0.53	2.8 (0.10)
Total	325	0.64		0.60	0.61	2.2 (0.06)

5.2   Logistic Model Based Regression Analysis of ISA(n) Workload and RC(n) Taskload Means

In contrast to the ANOVAs in the previous Sect. 5.1 which consider the variance within the 2 min simulation time intervals and participant sample based on the reduced raw data set ($2 \leq n \leq 7$), we use in this section for the nonlinear logistic model based regression analysis means across WL measurements and participants for given traffic load n (AC/2 min), i.e. the preprocessed data of the full dataset (see Appendix of the Chapter with Tables A1, A2, A3). We use a bisquare weighting procedure for reducing the effect of outliers to generate robust nonlinear model parameter estimates (Matlab® statistcs toolbox).

As mentioned above for $c = 0$ only measurements with $1 \leq n \leq 5$ were obtained and ISA values within stdev overlap with those of condition $c = 1$ (see Fig. 6). That is why we combine $c = 0$ and $c = 1$ data into a new baseline condition, for the regression analysis in order to improve the statistics, i.e. stdev of means = standard error (sterr) = stdev/$\sqrt{N}$ and confidence intervals (CI) of logistic parameter estimates. This appears justified also because our focus is on quantifying the WL difference between standard work conditions as baseline and new MRTO work condition with single ATCo controlling multiple APs ($c = 2$) so that we perform regressions with baseline defined by "2 ATCOs control two remote APs". With our theoretical model, we aim to quantify hypotheses H1, H2 for ISA-WL and communication task load variable RC dependencies on traffic load n via the corresponding logistic sensitivity parameters. Furthermore, the above hypothesized (H3, H4) mediator relationship is quantified by the formal combination of ISA-WL and RC-TL (ISA(RC(n))), providing the psychophysics (Stevens) power law exponent $\gamma = \rho/\nu$ as quantitative index for the WL—TL dissociation.

In Sect. 5.2.1 we report nonlinear (NL) iterative weighted least squares (LSQ) regressions for logistic parameter estimates of ISA(n|ν) and RC(n|ρ) (see Eqs. (2), (4) in theory Sects. 4.1, 4.2) averaged across conditions as well as separated for conditions. In Sect. 5.2.2 we present the results of one-parameter regressions with the generalized linear model (see Eqs. (3), (5)) applied to the normalized and transformed WL and TL data y_I(ISA(n)), y_R(RC(n)). They are used for improvement of confidence and test of consistency with parameter estimates obtained with the nonlinear iterative procedure in the previous section. Finally, in Sect. 5.3 we present the results of nonlinear iterative and generalized linear regression analysis using the psychophysics power law of the ISA(RC) dependency for estimating the (Stevens) stimulus—response exponent $\gamma = \rho/\nu$ that formalizes hypotheses H3. The Stevens exponent puts the mediator hypothesis H4 on a formal theoretical basis.
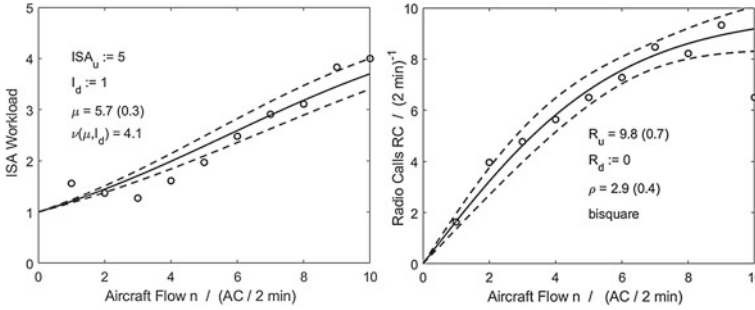
Fig. 7 Logistic model regressions with 95% CI with baseline ($c = 1$) ISA(n) data (Table A3): **a** 1-parameter (μ) fit (with $I_{min} := I_d = 1$, i.e. $\nu = \mu/\ln(4)$), **b** radio calls RC(n) with 2-parameter (R_u, ρ) fit. Effect of RC($n = 10$) = 6.5 minimized through bisquare weighting (represented through single 2 min interval only). Prior knowledge (I_u, I_d, R_{min}) and parameter estimates as inset

5.2.1 Nonlinear Logistic Least Squares Regressions for ISA(N) and RC(N)

In what follows we provide logistic parameter estimates using the robust iterative least squares Levenberg–Marquardt algorithm with bisquare residuals weighting of data to minimize the effect of outliers (Matlab Statistics Toolbox: Nlinfit) applied to the ISA and RC means of Tables A2, A3 in the Appendix. We compare ISA-WL and RC-TL data averaged across the three different conditions ($c = 0, 1, 2$) with analysis for separate conditions, i.e. baseline $c = 0$ & 1 with task sharing between two controllers for two remote APs, and single ATCO control of the two APs ($c = 2$).

Logistic Regressions for Baseline and Averages Across Conditions

In order to check the baseline definition $c = 0$ & 1 (as compared to $c = 0$ used for the ANOVAs in Sect. 5.1) we compare logistic parameter estimates with $c = 1$ and $c = 0$ & 1: they should be equal within sterr, and ideally sterr and CI should be smaller for $c = 0$& 1 due to increased number of data points. Figure 7a, b depicts the nonlinear regressions for $c = 1$ data.

NL-regression with the one-parameter model (Eq. (2), Sect. 4.1 using $\mu = \nu\ln(4)$) yields as fit parameter estimate for ISA(n) $\mu = 5.7$ (± 0.3), with derived scaling parameter $\nu(\mu) = 4.1$ (± 0.2), reasonably close to the theoretical shift value $\mu_t = 5$ derived as prior (rough) estimate by plausibility argument. The observed weak ISA(n) nonlinearity (in agreement with prediction, Fig. 2) contrasts the experimental communication load curve RC($n|R_u$, ρ) (radio calls TL/(2 min)$^{-1}$). The two-parameter regression with Eq. (4) exhibits the theoretically predicted asymptotic
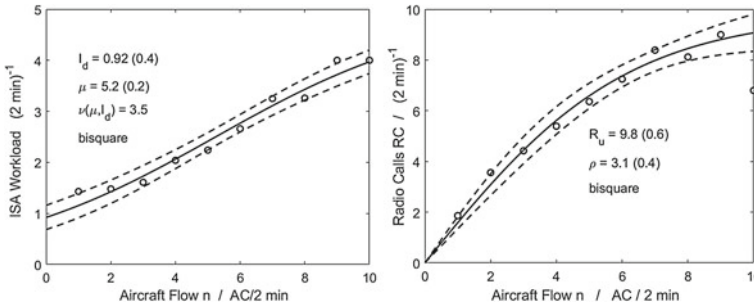
Fig. 8 Data averages across conditions (Table A2). Two-parameter logistic model fits with 95% confidence intervals and parameter estimates (with standard errors of means) of the **a** left: subjective ISA workload vs. traffic flow n with two-parameter (I_d, ν) fit; **b** right: (objective) rate of radio calls $RC/(2\ min)^{-1}$ versus $n/(2\ min)^{-1}$, with two parameter (R_u, ρ) fit

approach to the estimated upper limit $R_u = 9.8$ (± 0.7), $\rho = 2.9$ (± 0.4). The parameter estimates are reasonably close to the theoretical values $R_{ut} = 10$–11, $\rho_t = n_c/2 = 3.5$.

Figure 8a, b depicts the averages across conditions ($c = 0, 1, 2$; Table A2) of the subjective ISA-WL data and the radio call rate (task load, RC/2 min) as dependent on environmental traffic load n, again with logistic regression curves according to Eqs. (2), (4).

The results are very similar to the baseline ($c = 1$) curves in the previous Fig. 7. Because baseline data for $c = 0$ were available only for the traffic range $n \leq 5$ (with $n = 5$ represented only by one 2 min interval) the curves basically represent averages of $c = 1$ (statistically dominating) and $c = 2$ (the latter represented by mostly less than half as much cases). The good correlation with reduced sterr and improved CI for ISA(n) justifies the baseline definition $c = 0\&1$ and provides initial evidence for validity of the logistic models for the traffic dependence. The theoretically predicted parameter estimates ν_t (μ_t: $= 5$) $= 3.6$, $R_{ut} = 10$–11, $\rho_t = 3.5$ of Sect. 4 are valid for baseline conditions (2 ATCOs for 2 airports) and exhibit close agreement with experimental estimates for $c = 1$ (Fig. 7) and for averaged conditions ($\pm$sterr, Fig. 8): $(\nu, R_u, \rho)_{estimate} = (3.5\ (0.1), 9.8\ (0.6), 3.1\ (0.4))_{c=0,1,2}$. These preliminary conclusions are suggested by the dominating effect of $c = 1$ for the averaging across the three conditions.

Logistic Regressions for Separate Baseline and RTC Work Conditions

In this section we provide iterative nonlinear regression analyses separated for baseline ($c = 0\&1$) and single operator conditions $c = 2$ (Table A3 data). Regressions with estimates of two free parameters of the logistic model are compared with 1-parameter regressions where prior information is included to restrict degrees of
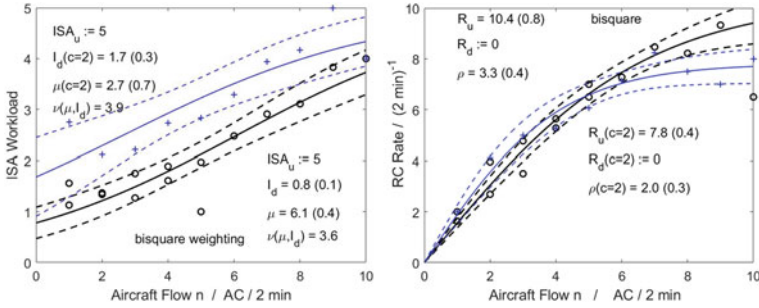
Fig. 9 Data averages separated for conditions with baseline ($c = 0$ & 1, circles) and single operator ($c = 2$, crosses). Two-parameter logistic model fits with 95% confidence intervals. Parameter estimates as inset (with standard errors of means), from left to right: **a** subjective ISA workload versus n with two-parameter (I_d, v) fit; **b** (objective) rate of radio calls $RC/(2\ min)^{-1}$ versus traffic flow $n/(2\ min)$, with two parameter (R_u, ρ) fit. Outliers at $ISA(n = 5, c = 0) = 1$, $ISA(n = 9, c = 2) = 5$, $RC(n = 10, c = 1) = 6.5$. Bisquare residuals weighting minimizes outlier effect of $ISA(n = 5, c = 0; n = 9, c = 2)$), $RC(n = 9, c = 1)$

freedom for model adaptation. If 1-parameter regressions exhibit improved confidence of parameter estimates (reduced sterr) this is taken as evidence for correct model assumptions.

The two independent model parameters to be estimated are: (a) (I_{min}, μ) for ISA(n), for testing the theoretical lower limits (logistic intercept I_d) for baseline ($I_{min}: = I_d = 1$) and $I_{min} = 2$ for $c = 2$; (b) (R_u, ρ) for RC(n) for testing the asymptote predictions in Sect. 4.2 $R_u(c = 0 \& 1) > R_u(c = 2)$, with lower limit (intercept) set to $RC(n = 0): = 0$ for all conditions.

By ignoring the fact that the logistic model allows for a 1-parameter fit to the ISA(n) data through the dependency $\mu = v \ln(k)$ (with $I_d = 1$ and $k(I_d) = 4$ for baseline) we can test with a 2-parameter fit the logistic intercept assumption $I_{min}(n = 0 \mid c = 0 \& 1): = I_d = 1$. With regard to the hypothesised WL increase under the $c = 2$ condition the data based estimate of effective $I_d(= I_{min})$ allows to check the plausibility argument of Sect. 4.1 leading to $I_{min}(c = 2): = 2$ and to quantify H1 stating $WL(c = 2) > WL(c = 0 \& 1)$. Of course, ignoring prior information goes at the cost of increased uncertainties of estimates (assuming the theory to be correct), i.e. increased CI and standard errors. The ISA(n) and RC(n) data together with logistic regressions (with 95% CI) separated for conditions $c = 0 \& 1$ and $c = 2$ are depicted in Fig. 9a, b.

The parameter estimates are summarized in Table 7 together with estimates of 1-parameter fits with WL scaling parameter for ISA(n) calculated via $v = \mu/\ln(k(I_u, I_{min})$ (see Fig. 10). For the RC(n) fit the logistic model assumption $RC(n = 0) = 0$ for all conditions evidently is verified by the data.

For ISA(n|v) the one-parameter model was formalized for baseline condition $c = 0 \& 1$ by Eq. (2) (with prior knowledge $I_{min}: = I_d = 1$ and relationship $\mu(v) = v \ln(k)$, $k = 4$). For $c = 2$, by theory, $I_{min} = 2$ (with $k = 1.5$), with experimental evidence provided by the above 2-parameter fit. For RC(n|ρ) the 1-parameter fit requires the

Table 7 Parameter estimates (with sterr of means) of two-parameter fits to the ISA(n) and RC(n) data separated for baseline ($c = 0$ & 1) and single controller condition ($c = 2$), compared with 1-parameter regression estimates (Fig. 10)

Cond	2-parameter regression (Fig. 9)					1-parameter regression (Fig. 10)				
	I_{min}	μ	$v(I_{min}, \mu)$	R_u	ρ	I_{min}	μ	$v(\mu)$	R_u	ρ
$c = 0$ & 1 (sterr)	0.83 (0.1)	6.1 (0.4)	3.6 (0.5)	10.4 (0.8)	3.3 (0.4)	1	6.1 (0.5)	4.4 (0.3)	10	3.1 (0.1)
$c = 2$ (sterr)	1.7 (0.3)	2.7 (0.7)	3.9 (1.0)	7.8 (0.4)	2.0 (0.3)	2	2.0 (0.3)	4.9 (0.7)	8	2.5 (0.2)
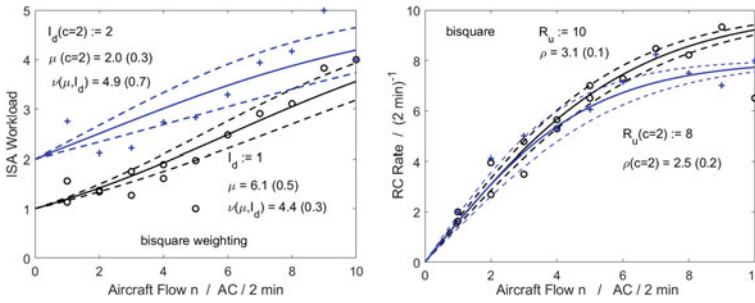
Fig. 10 One-parameter (ν, ρ) logistic model fits (solid curves) with 95% confidence intervals (dashed, dotted) of the data averages (across 2 min cases) separated for baseline ($c = 0$ & 1, circles) and single ATCo ($c = 2$, crosses) conditions: **a** ISA workload vs. traffic flow n (AC/2 min); **b** frequency of radio calls RC/(2 min)$^{-1}$ versus n/(2 min)$^{-1}$. Insets: parameter estimates (with standard errors of the means); prior knowledge: I_{min}, I_u. Outliers at ISA($n = 5$, $c = 0$) = 1, ISA($n = 9$, $c = 2$) = 5, RC($n = 10$, $c = 1$) = 6.5. Bisquare residuals weighting minimizes outlier effect on regression

assumption of prior knowledge for asymptote values (theoretically predicted and supported by 2-parameter fits) R_u: = 10 for $c = 0$ & 1 and R_u: = 8 for $c = 2$. Results of the one-parameter regressions are depicted in Fig. 10.

The 1-parameter fits for RC(n) exhibit improved confidence intervals and standard errors (derived from the regression) as compared to the 2-parameter fits however not for ISA(n), probably due to large data scattering in the low n-range. In agreement with the 2-parameter fit significantly increased subjective workload is confirmed for $c = 2$ condition as compared to baseline $c = 0$ & 1, i.e. a single ATCo (2 airport control) bias $I_{min}(c = 2) = 2 > I_{min}(c = 0\&1) = I_d = 1$. Radio call rates RC($n$) converge to the predicted significantly lower asymptote for $c = 2$. Corresponding t-Tests are provided in the following section using the generalized linear models for regression with normalized and (logarithmic) transformed variables.

Figure 10 with the nonlinear regressions of the subjective ISA(n) and objective radio calls data RC(n) and logistic parameter estimates allow for direct comparison with the theoretical predictions depicted in Figs. 2 and 3 of Sect. 4. For a corresponding discussion see Sect. 6.

5.2.2 One-Parameter Regressions with Generalized Linear Models

The possibility of using a generalized linear version of the logistic model according to Eqs. (3) (5) allows for application of established linear regression methods with t-tests of the transformed normalized dependent variables into normalized semilog-scale. It requires however, inclusion of prior knowledge on the asymptotic limit (i.e. $I_u = 5$, $R_u \approx 10$ (2 min)$^{-1}$) for normalization and logarithmic transformation into the new WL and TL variables variable $y_I = \ln(\text{ISA}/(I_u - \text{ISA}))$ and $y_R = \ln(R_u + R)/(R_u - R)$ respectively. The following Fig. 11 depicts $y_I(n)$ as calculated with ISA-data
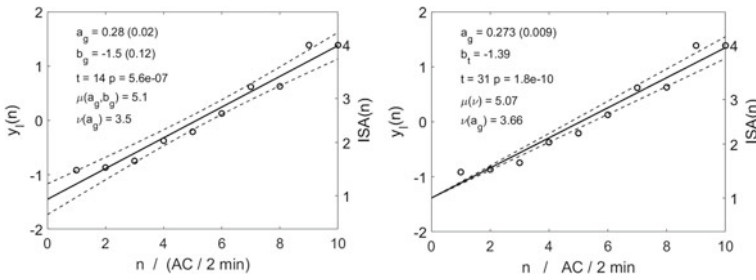
Fig. 11 Normalized and log-transformed WL data $y_I(n)$ with ISA-WL means across conditions. Linear regressions using GLM. **a** 2-parameter fit (slope a_g, intercept b_g). **b** 1-parameter fit with $b_g := b_t = -\ln(4)$ (baseline, $I_d = 1$). Inset: slope and intercept estimates ($\pm$stderr) and derived logistic scaling and shift parameter. 95% CI and t-test shows improvement with theoretical b_t ($\equiv b_{gt}$) as prior information

averaged across conditions (Table A2 in the Appendix of this chapter) together with the linear regression using the generalized linear model (GLM) (3).

As expected the logistic parameter estimates $\nu_g = 1/a_g$ of the $y_I(n)$ fits confirm those of the above nonlinear iterative least squares regression within the given uncertainties (inset of Fig. 8). Relative standard error of sensitivity parameter ν_g corresponds to that one of slope a_g ($0.01/0.3 \approx 3\%$). The t-tests and corresponding confidence intervals show high significance ($p < 2\ 10^{-10}$ with $t = 31$) that supports the significance of logistic nonlinearities in the model.

Figure 12 depicts the normalized and log-transformed variables separated for baseline ($c = 0\ \&\ 1$) and single ATCO ($c = 2$) conditions, again using model Eq. (3) and (5) respectively. The regression models require as prior knowledge asymptote
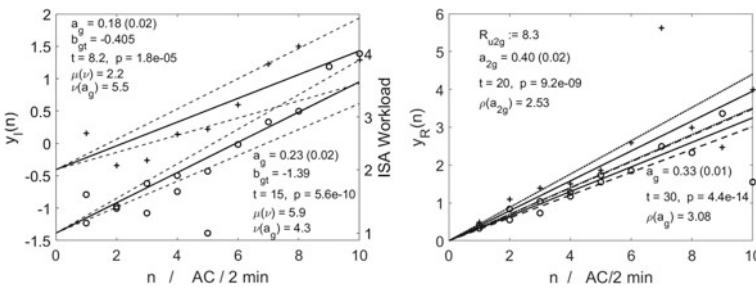
Fig. 12 Normalized and logarithmic transformed yI(ISA)(n) and yR(RC)(n) means separated for condition $c = 0\ \&\ 1$ with $k = 4$ ($I_{min} = 1$) and $c = 2$ with $k = 1.5$ ($I_{min} = 2$), with least squares 1-parameter (robust) linear regressions using generalized linear (semilog) model for slope estimates (slope a_g ($\pm$stderr)), with theoretical intercept $b_{gt} = -\ln(k)$. CI and t-test for 95% confidence. **a** $y_I(n)$ averages with model Eq. (3) for lin. regression. **b** $y_R(n)$ means with model Eq. (6). Regressions are clearly separated with no overlap of 95% CI for $c = 0\ \&\ 1$ (dashed), $c = 2$ (dotted). Insets with parameter estimates and t-test. Residuals bisquare weighting for outlier rejection as described above for nonlinear regressions

Table 8 Sensitivity parameter estimates (with sterr. of means) of generalized linear one-parameter fits to the transformed variables $y_I(ISA(n))$, $y_R(RC(n))$ using prior knowledge for normalization and calculation of y_I-intercept and asymptotes (I_u, I_d, R_u), compared with theoretical predictions μ_t, $\nu_t(\mu_t)$, ρ_t. Parameters separated for baseline ($c = 0$ & 1) and single controller condition ($c = 2$). $p(t) = t$-test probability at 5% level

| Cond	ν_t	ν(sterr)	μ_t	$\mu(\nu)$	p(t)	ρ_t	ρ(sterr)	p(t)
0 & 1	3.6	4.3 (0.3)	5	5.9 (0.4)	$5.6 \cdot 10^{-10}$ (15)	3.5	3.1 (0.1)	$4.4 \cdot 10^{-14}$ (30)
2	3.6	5.5 (0.7)	1.5	2.2 (0.3)	$1.8 \cdot 10^{-5}$ (8.2)	3.0	2.5 (0.1)	10^{-9} (20)

value I_u (:= 5), $I_{min} := I_d$ for calculation of k, and R_u for normalization of RC with $R_{min} = R_d = 0$ for both conditions. Corresponding estimates are provided by theoretical considerations in Sect. 4 (Table 3) and they were confirmed by the nonlinear two-parameter fits in the previous Sect. 5.2.1.2 (Fig. 10, Table 7): $R_{ug}(c = 0$ &1): $= 10$, $R_{ug}(c = 2)$: $= 8$.

The logistic sensitivity parameter estimates ν_g and ρ_g (=inverse of the slope parameter estimates a_g with sterr.) are collected in Table 8, together with theoretical predictions and t-tests. As expected parameters ν_g and ρ_g of the generalized linear $y_I(n)$ and $y_R(n)$ regressions in Fig. 12a and b respectively confirm those of the above nonlinear logistic regressions within the given uncertainties. This agreement of results obtained with original and transformed normalized variables and different regression methods is taken as a further support for the logistic model. The t-tests and corresponding confidence intervals show high significance for the $y_I(ISA)$ baseline condition ($p < 6 \cdot 10^{-10}$ with $t = 15$) whereas for single operator control ($c = 2$) the confidence is smaller ($p < 2 \cdot 10^{-5}$ with $t = 8.2$). Nevertheless CI of regressions are sufficiently separated for clear discrimination between conditions when using the theoretical predictions as prior knowledge for intercept $b_g = -\mu/\nu = -\ln(k)$, with $b_{gt}(c = 0$ & $1) = -\ln(4) = -1.39$ and $b_{gt}(c = 2) = -0.4$.

For the generalized linear RC-regressions, besides $R_{ug}(c = 0$ & $1) = 10$ the asymptote $R_{ug}(c = 2)$: $= 8.3$ (required for the $y_R(RC)$-transformation) is selected as minimum value > maximum call rate (for $n = 7$, see A3 in the Appendix). This adjustment, as compared to the 2-parameter estimate $R_u = 8$, is required for avoiding negative log-arguments and complex numbers during the regression calculations. As expected, also the logistic radio calls sensitivity parameters ρ_g agree with those of the nonlinear regressions within the standard errors. 95% CI and t-tests confirm the significance of $c = 0$ & 2 and $c = 2$ sensitivity difference characterized by scaling parameters ρ.

5.3 Psychophysics Power Law Model for ISA(RC) Regression Analysis

In this section we analyze the correlation between the dependent TL and WL variables based on the power law relationship Eqs. (6) and (7). The ISA(RC) power law
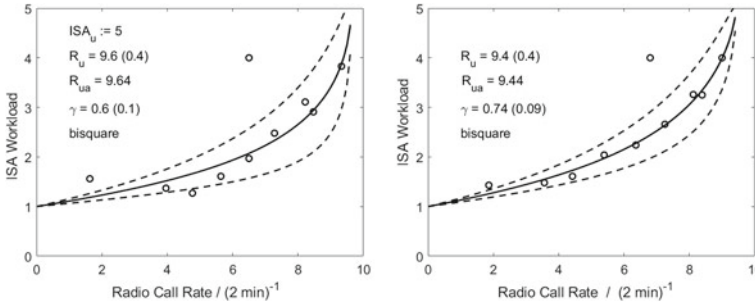
Fig. 13 Measured subjective ISA-WL (ordinate) vs. objective RC-TL (abscissa) data (Tables A2, A3) with 1-parameter power law regression (Eq. (6), solid lines; baseline assumption: $I_{min} := I_d = 1$; dashed curves: 95% CI for estimation of exponent γ (with sterr). **a** left: baseline data, $c = 1$; **b** right: average across conditions. R_{ua} = adjusted asymptote for 1-parameter fit, derived from prior 2-parameter estimate R_u. Outlier RC = 6.5 (ISA = 4, $c = 1$) represented by only 4 cases with traffic load $n = 5$ (Table A1, Bisquare weighting applied for minimizing outlier effect)

analysis of the experimental data allows for an additional estimate of the sensitivity parameters ν, ρ independent of the environmental traffic load n due to the theoretical relationship $\gamma = \rho/\nu$ between power law exponent and logistic scaling parameters. The observed dissociation of the logistic ISA-WL and RC-TL characteristics can be discussed in terms of Stevens psychophysics stimulus (RC)—response (ISA) paradigm (Fürstenau & Radüntz, 2021), see Discussion Sect. 6).

Here we provide the results that formalize the hypotheses H3 and H4 of Sect. 5.1 by means of the power law model Eq. (6) for the ISA(RC) relationship and its generalized linear (log–log) form (7). It combines the measured subjective ISA(n) workload data with the objective task load (communication) variable RC(n) by elimination of the environmental load variable n. As an initial test for the validity of the power law in the present context the experimental results obtained for condition $c = 1$ and for the average across all three conditions are depicted in the following Fig. 13a, and b, together with the nonlinear robust least squares regression (bisquare weighting applied for minimizing outlier effect) using two- (R_u, γ) and one-parameter (γ) fits. In a first step two parameter fits provide estimates for R_u and γ. Because R_u estimates agree within standard errors with the logistic estimates of Sect. 5.2.1.1 we use these values as prior information for succeeding 1 parameter fits with improved CI for estimates of exponent γ(R_{ua} = adjusted R_u from prior 2-parameter fit to avoid negative numbers in the power argument of Eq. (6) during execution of the 1-parameter fit procedure):

The ISA(RC|R_u, γ) regression estimates (independent from environmental traffic load n) are given in the insets of Fig. 13a), with $\gamma(c = 1) = 0.64\ (\pm 0.1)$ and Fig. 13b), and γ(average across conditions) = 0.74 ($\pm$0.09). The results agree within sterr with the logistic 1-parameter regression estimates (ρ, ν) of Sect. 5.2.1.1 (Fig. 7, traffic load n as independent variable: $\gamma(c = 1) = \rho/\nu = 2.9/4.1 = 0.71$), and for the average across conditions $c = 0$ & 1 and $c = 2$ (Fig. 8) $\gamma = 3.07/3.53 = 0.87$. Also the present 2-parameter power law estimates of the asymptotes $R_u = 9.6 (\pm 0.4)$ for $c = 1$, and
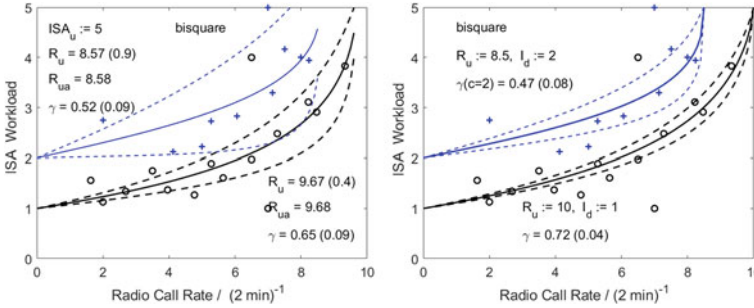
Fig. 14 Subjective ISA-WL vs. RC task load data from Table A3 (Appendix), separated for baseline condition (circles: $c = 0$ & 1) and single ATCo (crosses: $c = 2$). **a** Left: two parameter (R_u, γ) power law fit (solid curve) with 95% CI (dotted). **b** Right: 1 parameter regressions with asymptote R_u from 2-parameter fit (rounded). Intersection values as prior knowledge: I_{min} (baseline): $= 1$, $I_{min}(c = 2)$: $= 2$. Bisquare residuals weighting applied for minimizing effects of outliers: ISA $= 1$ ($c = 0$, RC $= 6.5$) represented by only a single case with traffic $n = 5$; RC $= 6.5$ ($c = 1$, ISA $= 4$) by 4 cases with $n = 10$; ISA $= 5$ ($c = 2$, RC $= 7$) by 1 case

$R_u = 9.4(\pm 0.4)$ for average across conditions is consistent with the independent estimate by means of the logistic RC($n|R_u$, ρ) model in Sect. 5.2.1.1 (see Table 9).

In the following Fig. 14a, b the measured ISA(RC) data are depicted, separated for conditions together with (a) two-parameter (R_u, γ) and (b) 1-parameter (γ) power law fit ISA(RC) using Eq. (6) with prior information $I_{min}(c = 0\&1)$: $= 1$, $I_{min}(c = 2)$: $= 2$ (for calculating $k = I_u/I_{min} - 1$). The 2-parameter fit allows for an independent test of the logistic R_u asymptote estimate with RC($n|R_u$, ρ) in Sect. 5.2.1.2, together with estimates of the power law exponents. In an iterative procedure R_u is then defined as prior information for the 1-parameter fit in Fig. 14b), resulting in improved confidence intervals if the theoretical assumptions are correct.

The independent two-parameter power law fit to the ISA(RC) data in Fig. 14a for estimating (R_u, γ) confirm the R_u estimates of RC(n) (Sect. 5.2.1.2, Fig. 9 and provide dimensionless power law exponent values of the order of 1 (rel. sterr ca 15%) with $\gamma(c = 0$ & 1$) = 0.65$ (0.09) $> \gamma$ ($c = 2$) $= 0.52$ (0.09). For the ISA(RC) characteristic and condition $c = 2$ we got to increase the asymptote to R_u: $= 8.5$–8.6 to avoid complex values during the iterative regression procedure. Power law curves clearly indicate the difference of RC_u asymptotes between conditions and exhibit significantly different exponents $\gamma(c = 0$ & 1$) = 0.72$ (0.04) and $\gamma(c = 2) = 0.47$ (0.08) which is compatible with the rough theoretical predictions (using $\gamma_t = \rho/\nu$ Sect. 4.3): $\gamma_t = 0.97$ for $c = 0$ & 1 and 0.83 for $c = 2$.

Like for the nonlinear model version also the generalized linear one allows for an estimate of γ independent from the external load variable n through the characteristics of the transformed normalized variables (ISA, RC) in log–log scale $y_I(y_R)$ (Eq. (7)). The following Fig. 15 depicts the corresponding 1-parameter fits (exponent γ) with 95% CI, separated for conditions. We again use prior knowledge (from theory and nonlinear 2-parameter fits, Fig. 10) for asymptotes and for ISA intercept parameter (see Table 3). Like before also for $y_R(n)$ the RC-asymptote has to be somewhat
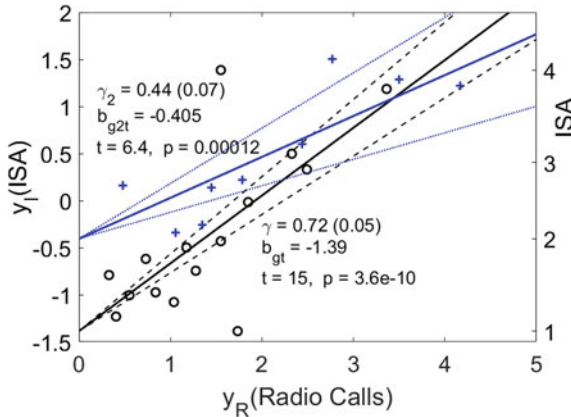
Fig. 15 Transformed, normalized y_I(ISA) versus y_R(RC) data with linear scale for log- transformed variables (abszissa and left ordinate). Logarithmic ISA scale at right ordinate. One-parameter robust linear regression (bisquare weighting for outlier rejection) using generalized lin. model Eq. (7) separate for baseline (circles, $c = 0$ & 1) and single controller (crosses, $c = 2$). 95% confidence intervals for baseline $c = 0$ & 1 (dashed) and $c = 2$ (dotted)

increased for the regression as compared to the nonlinear 2-parameter estimate in order to avoid complex numbers during the calculation (keep all RC data $< R_u$ for the transformation).

As expected parameter estimates γ again confirm the results of the nonlinear fits in Fig. 10, with significantly different γ for the two conditions. Consequently, both the nonlinear regression with the subjective ISA WL-means and the objective RC communication means data depicted in Fig. 10 and the linear regression with the generalized linear model provide evidence for a psychophysics power law relationship.

For comparison with the estimates in the previous sections and with the theoretical predictions (ν_t ($\mu_t := 5$) = 3.6, $\rho_{t =} 3.5$, $\gamma_t = \rho/\nu = 0.97$, for baseline condition) the exponent estimates of the nonlinear ISA(RC) power law and the transformed generalized linear model $y_I(y_R)$-regressions (GLM) are summarized in the following table:

6   Discussion

In the present work we analyzed experimental data of a Multiple Remote Tower human-in-the-loop simulation to explore the effect of features of the task and environment on subjectively experienced workload. The goal of this analysis was to quantify the effect of dual tasks (controlling two airports (AP) at the same time by a single operator) on experienced workload as compared to standard procedures (single AP control). For this purpose, we tested four hypotheses (H1–H4, Sect. 5.1) concerning

Table 9 Power law exponent estimates (with sterr of means) of ISA(RC) (Eq. (6)) using prior information (I_u, I_d, R_u) for comparison of nonlinear 2- and 1-parameter fits, separated for baseline (c = 0 & 1) and single controller condition (c = 2). Comparison with GLM one-parameter fits (Eq. (7), prior information b_{gt}) of the log_log transformed data for variables $y_I(y_R)$, p(t) = probability of t-test at 5% rejection level

Parameter estimates (I_u = 5)	Power law, 2-param. Fit Fig. 14a		Power Law, 1-param. Fit Fig. 14b	95% CI 1-param. fit	GLM, 1-param fit	
Prior inform.:	I_d(c = 0 & 1 / 2): = 1 / 2		I_d: = 1 / 2; R_u: = 10 / 8.5		b_{gt} = ln(k): = −1.39 / − 0.405	
Condition c	γ (sterr)	R_u	γ (sterr)	[$\Delta\gamma$−CI]	γ (sterr)	t-test p (t)
0&1 (baseine)	0.65 (0.09)	9.7 (0.4)	0.72 (0.04)	[0.6, 0.8]	0.72 (0.05)	$3.6\ 10^{-10}$ (15)
2 (single ATCO)	0.52 (0.09)	8.6 (0.9)	0.47 (0.08)	[0.3, 0.6]	0.44 (0.07)	10^{-4} (6.4)

the effect of environmental (traffic) load n on dependent variables workload (ISA-WL) and communication task load TL (radio call duration RD and frequency RC), and the ISA—RC interdependence. In addition to two-factor variance analysis (ANOVA, linear model assumption with F-test and Pearson correlation coefficients) we applied a new nonlinear (logistic) resource limitation model for regression based model parameter estimates which can be compared with theoretical parameter predictions based on prior information. The logistic resource limitation model was designed and validated within a previous simulation experiment addressing approach sector radar control WL measurements (Fürstenau et al., 2020; Fürstenau & Radüntz, 2021). That experiment was comparable to the present MRT-simulation experiment with regard to independent and dependent variables and the general simulation setup (simulation of communication between ATCOs and trained pseudo pilots who replaced the real pilots) although the detailed work environments were different: RTC videopanorama with standard tower control equipment providing the data for the present analysis vs. approach sector radar control workplace for the previous model validation.

In what follows we discuss and compare the original ANOVA based evidence for the four hypotheses (Sect. 5.1) with results gained from the logistic resource limitation (LRL, Sect. 5.2) and stimulus–response power law (SRP, Sect. 5.3) regression analysis, and with the theoretical model parameter predictions based on prior information in Sect. 4.

6.1 H1: Dependence of Subjective WL on Traffic Load and Impact of MRTO

Hypothesis H1 concerning the increase of subjective WL with traffic load n is clearly depicted in Fig. 6 both for $c = 1$ and $c = 2$, the latter exhibiting a shift towards higher WL with an ISA difference of around $\Delta ISA \approx 1$. Initial quantitative evidence for this observation was provided within a linear approach in Sect. 5.1 by variance analysis (2-factor ANOVAs with F-tests and $p < 0.05$) and Pearson correlation coefficients (Table 6) mostly determined as high ($r > 0.5$) according to Cohens' definition. Moreover, the new RTC work condition with a single controller responsible for two airports (condition $c = 2$) was confirmed to generate significantly higher workload than baseline condition where two controllers managed the traffic on the same two remote airports (with two alternatives of task sharing for $c = 0$ and $c = 1$) with $p(F = 38)$ < 0.01. The extreme n values were disregarded for the ANOVAs (limited to $2 \leq n \leq 7$) generating comparable statistics for all n levels.

Obviously, a linear model for the ISA(n) characteristic can only be valid for a limited (operational) range of the independent variable due to boundaries of the ISA scale and the fact of cognitive resource limitation (see Introduction, Sect. 2 and Appendix C). Analysis of ISA-WL vs. traffic load in previous work (Fürstenau et al., 2020) has provided the inter-individual variation of slope and intercept parameters of a linearized version of the logistic model (Eq. (1)) for a sample of 21 participants. The average of intercept values across participants was smaller than the lower ISA scale limit ($I_d = 1$) so that a WL underload interval with slope = 0 for traffic $0 \leq n \leq n_0$ (ISA(n_0) = 0) was obtained. Consequently, a more realistic model has to consider a deviation from linearity of ISA(n) for the low traffic range, i.e. gradually increasing slope or concave side of the characteristic. This was suggested also by the corresponding data and plots of the means vs. n (Sect. 5.1, Tables A2, A3 and Fig. 6) when considering also the extreme traffic load values n (AC/2 min) that included only a low number of cases (i.e. 2 min data intervals), extending the traffic range from $3 \leq n \leq 7$ as used for the ANOVAs, to $1 \leq n \leq 10$ for the nonlinear models.

The logistic resource limitation model described in Sect. 4 provides the most simple approach for quantifying the expected nonlinearities (see Appendix D for details). Our logistic modelling approach may be compared with results of a simulated ATC HitL-simulation experiment of Lee (2005) (Lee et al., 2005). They used a heuristic sigmoid function based four-parameter regression for ATWIT-WL data analysis that exhibited significant parameter estimates (see Sect. 2.2.4 and Appendix 3).

The linear ANOCOVA correlations of the cleaned data set ($2 \leq n \leq 7$, see above and Sect. 5.1) was improved by means of logistic parameter estimates in Sect. 5.2 with the full data set ($1 \leq n \leq 10$) using robust nonlinear regressions (bisquare weighting for outlier suppression) with highly significant t-values (see Tables 7, 8). As mentioned already in Sect. 5, for improving statistics with the nonlinear regressions it appeared justified (supported by theory) to define as baseline condition (2 ATCOs controlling the two airports) the combined set of $c = 0$ and $c = 1$ data of

Table A3. The $c = 0$ baseline data are limited to the lower half of the traffic range and as such are not sufficient for a model-based test of H1 via comparison with $c = 2$ data. Future MRTO HitL validation should take into account this problem.

Regressions were based on the logistic model (Eq. (1)–(3)) for the ISA means averaged across conditions (Fig. 8), and separated for baseline ($c = 0$ & 1) and single operator MRT condition $c = 2$ depicted in Figs. 9, 10, 11. In contrast to the ANOVA analysis, the theoretical model allows the integration of prior information so that 1-parameter fits to the data can be used to estimate the remaining logistic scaling (inverse sensitivity) parameters ν for baseline and $c = 2$ condition. This reduced the standard error to half of the uncertainty of the 2-parameter fits with corresponding improvement of t-test and CI (see Table 7).

For ISA-WL(n) data deviation from linearity was most pronounced in the low traffic range ($n \leq 2$ AC/2 min, i.e. ISA underload, see Fig. 6: $c = 1$, 2) and it is evident, e.g. in Figs. 7 and 8 in Sect. 5.2.1. depicting the $c = 1$ condition and the average across conditions with logistic 2-parameter regressions for ISA(n|I_{min}, μ) respectively. This nonlinearity was theoretically predicted for the baseline sigmoid ($c = 0$ & 1) in Fig. 2 (Sect. 4.1) and quantified through the estimates of logistic scaling parameter ν. For ISA(n), in contrast to RC(n) (see below Sect. 6.2) the experimental results did not show the theoretically proposed deviation from linearity for the high traffic range for baseline conditions, however a weak effect as depicted in Fig. 10 for the $c = 2$ condition. This may be expected due to its global shift to higher WL by ΔISA $\approx + 1$. It has to be considered that the original study was not designed to analyze formal relationships between environmental load, objective TL and subjective WL responses. As pointed out above, future research should design studies to allow for a representative data set for all working conditions and a realistic traffic range. For this paper, the results obtained from applying the non-linear model to the data at hand demonstrate the application of these models for workplace and working method design.

The contribution of the $c = 2$ condition to the I_{min} estimate in Fig. 8 evidently is limited due the fact that the number of $c = 2$ cases (2 min intervals) is smaller for each n and particularly for the low *n*-range than the number of combined $c = 0$ & 1 cases (Table A1). The regression based shift estimates $\mu = 5.7$ (± 0.3) and 5.2 (± 0.2) in Figs. 7 and 8 respectively support our assumption that condition $c = 1$ with two ATCOs sharing the two-airport control, and only one communicating with pilots is reasonably well modeled by the baseline ISA characteristic, with ISA(n = 0): = $I_{d = 1}$, $k(I_d) = 4$. Also, it is in reasonable agreement with the theoretical characteristic Eq. (2), depicted in Fig. 2 (solid curve) using shift parameter μ: = 5 as plausible prior information that represents the inversion point of the logistic curve as center of the nearly linear n-range, allowing for optimum operational conditions.

Within standard errors the lower ISA range limits for both conditions, as estimated from 2-parameter regressions for separate conditions (Fig. 9a and Table 7) are compatible with scale limits, prior information and theoretical predictions $I_{min = I_d} = 1$ for $c = 0$ & 1 and $I_{min}(c = 2) = 2$ (see Sect. 4 and Table 3). Consequently, these values were used as prior knowledge for 1-parameter regressions depicted in Fig. 10a. Within sterr also the logistic shift parameter μ for baseline corresponds

to the theoretical prediction (see Table 7). Because sterrs. of the scaling parameters ν for both conditions overlap (indicating more or less equal slope at the inversion point) the single controller characteristic ($c = 2$) appears to correspond to a shifted version of the baseline condition ($\mu(c = 0$ & $1) \rightarrow \mu(c = 2)$). So, the subjective WL intersection I_{min} ($c = 0$ & $1) = 1$ as prior knowledge for the 1-parameter fit is shifted from 1 to 2 under MRT (single controller) condition $c = 2$, in agreement with the theoretical prediction. Evidently this contrasts with the objective radio calls communication load (see Figs. 9b, 10b, and below Sect. 6.2).

The generalized linear model (3) allowed for testing the consistency of the above iterative lsq. regression based parameter estimates (Fig. 10a) with standard linear regressions for the correspondingly transformed normalized ISA and RC variables y_I(ISA) as depicted in Fig. 12a with semilog. coordinates. Using the theoretical regression line intercept parameter values as obtained from prior knowledge ($b_{gt} = -\ln(k)$) the slope parameter estimates obtained from linear 1-parameter regressions as shown in the insets of the figures together with t-tests indicate a high significance at the 5% level. Sensitivity parameter ν is given by the inverse of slopes $1/a_g$ (Eq. (3)) and regression estimates are summarized in Table 8 together with the theoretical predictions. The theoretical shift parameter μ_t: $= 5$ as plausible prior assumption (Sect. 4.1) yields a sensitivity parameter $\nu_t = 3.6$ via the theoretical relationship $\nu = \mu/\ln(k)$ with $k(c = 0$ & $1) = 4$. Theoretical arguments in Sect. 4.1 together with the experimental results (Fig. 6) suggested for the single ATCO(2 airports) MRTO condition an intercept $I_{min}(c = 2)$: $= 2 = I_d + 1$ as a plausible value yielding $k(c = 2) = I_u/I_{min} - 1$: $= 1.5$. As a first approximation we assumed this offset shift $\Delta I_{min} = 1$ to result from a simple shift of the baseline characteristic from μ_τ: $_= 5$ to $\mu_t(c = 2)$: $= 1.5$, with unchanged sensitivity parameter $\nu_t(c = 2) = \nu_t(c = 0$ & $1)$. Consequently for $c = 2$ the "subjective overload" (decreasing slope) starts at lower traffic n. Evidently, the theoretical shift values μ_t are reasonably close to the measured ones (reflecting the difference between conditions), and the regression based scaling parameter estimates in fact confirm the theoretical prediction of nearly equal sensitivity or slope $a_g = 1/\nu$ for both conditions (Tables 7 and 8).

To summarize, the new MRTO work condition ($c = 2$) with a single controller responsible for two airports was confirmed to generate significantly higher workload than baseline condition. The GLM regression in agreement with the iterative nonlinear LSQ-regressions above provided evidence that the subjectively experienced workload was raised about one level on the ISA scale. This increase was predicted theoretically based on plausibility arguments as depicted in Fig. 2.

6.2 H2: Dependence of Objective Communication TL on Traffic Load and Impact of MRTO

The ANOVAs of cumulative call duration RD and frequency of calls RC (per 2 min interval) supported the hypothesis H2 of increase with n (AC/2 min), quantified by

linear correlation coefficients around $r = 0.5\ldots0.7$ with F-test indicating significance at the 5% level (Sect. 5.1, Table 6). Table 5 provides a list of averages of RC, RD and derived mean call durations including uncertainties, separated for conditions. Within sterr the baseline $c = 0$ condition exhibits a clearly lower mean RD and RC as compared to MRT condition for $c = 1$ and $c = 2$. However, due to the limited n range for baseline $c = 0$ ($n < 6$, see above) this cannot be taken as evidence for a significant TL difference, compared to $c = 1$ and 2. Interestingly, the average call duration RCD agrees with results of a similar simulation experiment addressing ATCOs WL in a standard approach sector setting (terminal radar control (TRACON), typically 30–50 nautical miles distance from the AP Fürstenau & Radüntz, 2021), despite the different work environment: RCD decreases from 4 to 3.6 s with traffic flow increasing from extremely low to extremely (non-realistic) high values. This value is also close to results mentioned for sector control by (Djokic et al., 2010). These similarities in communication parameters can be explained by the highly standardized ATC communication.

In contrast to the ANOVAs (i.e. linear model), for high traffic load ($n > 5$ per 2 min) the theoretical prediction of radio call rates RC in Fig. 3 (RC model Eq. (4)) indicated even more pronounced nonlinear deviation near the theoretical maximum R_u as compared to the weak convergence of ISA(n) to I_u (for $0 < n \leq 10$). Consequently, logistic model-based regression analysis of objective RC(n) data was expected to provide significant scaling parameters ρ with slope $1/\rho$ (in semilog ordinate) $> 1/v$. This is confirmed in Figs. 7b and 8b) for $c = 1$ and for the average across conditions respectively, as well as separated for conditions in Figs. 9b, 10b, 11b. For RC(n) both under baseline and single controller condition ($c = 2$) the characteristic starts linear at the logistic inversion point $(n, RC) = (0, 0)$ and within confidence intervals, for lower load $n < 6$ the characteristics of both conditions overlap. The $c = 2$ curve begins to significantly deviate from the common low load section when approaching the capacity limit defined as $n_c \approx 7$, again in agreement with the theoretical prediction. The sensitivity (scaling) parameters between conditions are significantly different (within sterr and t-test at 5% level), with $\rho(c = 2) < \rho(c = 0 \ \& \ 1)$ quantifying the expected more rapid approach to the maximum value for $c = 2$ (Tables 7, 8). In Sect. 4.2 (Fig. 3) the theoretical model for RC(n) also correctly predicts the observed asymptote of single operator condition to exhibit a lower value ($R_u \approx 8$) as compared to baseline ($R_u \approx 10$). The inverse behavior for large traffic is a an important additional difference between subjective ISA-WL and objective RC-TL and it agrees with results of a comparable HitL simulation experiment reported in a recent publication (Fürstenau & Radüntz, 2021).

The generalized linear model (5) allows for testing the consistency of the above iterative lsq. regression based parameter estimates of RC(n) data (Fig. 10b) with standard linear regressions for the correspondingly transformed normalized RC variables $y_R(n)$ as depicted in Fig. 11b with semilog. coordinates. Using the theoretical regression line intercept parameter values as obtained from prior information (RC: $b_{gt} = 0$) the slope parameter estimates obtained from linear 1-parameter regressions as shown in the insets of the figures together with t-tests indicate a high significance at the 5% level. Sensitivity parameter ρ is given by the inverse of slopes $1/a_g$ (Eq. (5)) and

are summarized in Table 8 together with the theoretical predictions. The theoretical predictions for the RC-TL sensitivity index ρ_t are reasonably close to the measured ones.

When comparing regression based estimates of ISA-WL and RC-TL sensitivities (Table 7), i.e. normalized slopes $dp_R/dn_= p_R$' (normalized: $p_R = RC/R_u$) at the logistic inversion points, the functional differences of Eq. (1) and (4) give for ISA(n) (baseline condition $c = 0\&1$): p_I'$(n = \mu = 6) = 1/(4\nu) = 0.06$ and for RC(n) p_R'$(n = 0) = 1/(2\rho) \approx 0.16$. Together with the different intercepts and asymptotic behavior of subjective WL and objective communication TL this sensitivity difference quantifies the dissociation between TL and WL through the ratio p_I'$/p_R$' $= \rho/2\nu = 0.38$. The logistic parameter ratio $\rho/\nu \approx 0.7$ defines the power law or Stevens exponent γ derived in Sect. 4.3 (theoretical prediction ≈ 1) that is discussed with regard to hypotheses H3, H4 in the following two sections.

6.3 H3: Positive Correlation Between WL and Communication TL Variables RD, RC

Positive correlations, i.e. a linear dependency between RD, RC-TL and ISA-WL as initial evidence for the hypotheses H3 were determined within the 2-factor ANOVA (Table 6, Sect. 5.1) for the different work conditions, with significant (according to Cohen definition) Pearson coefficient $r =$ between 0.5 and 0.6. This result is in agreement with a principal components analysis of sector (radar) control data correlating 7-level ATWIT online WL ratings (Stein, 1985) with seven significant (task) load variables including traffic count per 4 min, communication rate and duration as reported by Manning et al. (2002). Obviously however, and as extension of the linear correlations, the different nonlinear characteristics of ISA(n) and RC(n), i.e. the theoretically predicted and experimentally observed WL—TL dissociation as discussed above, leads to a significant nonlinear relationship between subjective WL and objective TL.

For the following discussion of the results in Sect. 5.3 we refer to the power law Eqs. (6) and (7) relating communication load RC (calls/2 min) as TL-stimulus to subjective ISA-WL as response. The power law with exponent $\gamma = \rho/\nu$ as derived in Sect. 4.3 can be taken as a formal expression for the proposed ISA(RC) dependency that quantifies hypothesis H3 and the hypothesized mediator role (H4) of traffic variable n. Traffic load n as environmental simulation variable is eliminated when combining both dependent variables {RC-TL, ISA-WL} or {$y_R(n)$, $y_I(n)$} into the power law.

Figure 13a and b depict estimates with model Eq. (7) of the power law exponent for the $c = 1$ condition ($\gamma = 0.6$ (± 0.1) and for averages across conditions respectively (0.74(± 0.1)). The latter does not differ significantly from the former due to the dominating amount of $c = 1$ data (Table A1), however yielding improved 95% CI). Within sterr. the estimates agree with the value derived from the ratio of logistic

sensitivities (Fig. 7) so that this result can be taken as initial evidence for the power law as consequence of the resource limitation assumption and the logistic dependencies on environmental load. These results are also compatible with the theoretical prediction based on the theoretical parameter estimates $(v, \rho)_t$ of Sects. 4.1, 4.2, yielding as rough numerical exponent predictions $\gamma_t(c = 0 \& 1) = 3.5/3.61 = 0.97$ for the baseline condition and $\gamma_t(c = 2) = 3.0/3.61 = 0.83$. Both values are of the order of 1 as typically observed for the Stevens exponents in psychophysics.

Table 9 collects results from three different power law regressions using the ISA and RC data of Table A3 (Appendix), with direct parameter estimates separated for the two conditions: 2-parameter nonlinear fit for (γ, R_u) estimates, 1-parameter nonlinear γ estimate, and 1-parameter GLM γ estimate (95% CI with t-test). The best direct estimates of $\gamma(c = 0 \& 1) = 0.72 (0.05)$ and $\gamma(c = 2) = 0.44 (0.07)$ compare well with the above mentioned logistic sensitivity ratios (Table 8) $\gamma(c = 0 \& 1) = 0.72 (0.06)$ and $\gamma(c = 2) = 0.45 (0.06)$, the latter with roughly 10% relative uncertainty obtained from error propagation. Consequently, also this comparison of power law estimates with logistic parameter ratios using the environmental traffic load sensitivity, is taken as evidence for the resource limitation hypothesis as basis for the logistic models with power law as formal consequence.

The above results are also in agreement with those of a recently published similar HITL experiment using the same model equations for data analysis and parameter prediction, with exponent values in the range $0.8 < \gamma < 1$ (Fürstenau & Radüntz, 2021). Recent experimental results on flight control simulations with pilots with psychophysics-based data analysis of workload ratings and task load in terms of a manual control tasks using Stevens' stimulus—response relationship was published by (Bachelder & Godfroy-Cooper, 2019), with experimental Stevens exponents in the range $0.24 \leq \gamma \leq 0.41$.

6.4 H4: Mediator Effect and Sufficiency of Communication Load for Explaining Workload

Lange et al. proposed a model that integrated external (traffic) load, task load as mediator and experienced subjective workload (Lange et al., 2011). It was justified by the fact, that for the ATC task, external load as number of aircraft on frequency, causes the ATCOs to issue clearances to these a/c, measured as rate of radio calls (RC) and accumulated duration of radio communication (RD). The subjectively experienced workload should be caused by traffic load mediated by radio communication. This appears plausible, as for instance a traffic situation with two a/c being in conflict causes the ATCO to issue clearances that increases workload. WL-models used for sectorizing the air spaces are based on these mediating effects of ATCOs tasks, with radio communication one major contribution (compare Sect. 2.2). The mediator effect of communication load (communication time and radio call frequency) between traffic load n and ISA-WL, i.e., the interdependences ISA(RD) and ISA(RC) was

analyzed. As expected from comparison of Fig. 6a and b, for both cases a high and significant direct correlation was determined: $r = 0.63$ for ISA(RD) and $r = 0.60$ for ISA(RC). Not surprising, also a small although also significant mediating effect of traffic n on both pairs of variables (RC(n), ISA(n)) was determined: $r = 0.22$ for ISA(RD) and $r = 0.16$ for ISA(RC). It has to be noted that these results were obtained with a linear correlation analysis The difference between both correlation analyses can be explained by the evident deviations from linearity in Fig. 6. The nonlinear behavior of the data, for ISA(n) dominating in the lower range, for RC(n) in the upper range suggest the nonlinear analysis to provide even better results.

As mentioned above, by taking into account the fact of cognitive resource limitation (related to the dynamic energy budget of the brain) within the power law model the environmental load variable n was eliminated through combination of logistic ISA(n) and RC(n) functions yielding the exponent γ as ratio of TL/WL sensitivities ρ/ν (Eq. (6)). The high significance of the (nonlinear) logistic and power law parameter estimates (Table 9) together with the reasonable agreement with quantitative theoretical predictions based on prior information (Table 3) provide further quantitative evidence for the mediator effect of the communication load, originating from ATCO's task load due to the external traffic load. This conclusion appears valid for the average across work conditions (see Fig. 13b), and it is even more significant for well defined work conditions (Fig. 14b and Table 9 with reduced sterr and CI).

The clear separation of regression curves in Figs. 14 and 15 for $c = 0$ & 1, $c = 2$, i.e. ambiguity of the ISA(RC) and GLM $y_I(y_R)$ workload functions due to the dependence on work condition as second independent variable proves that hypothesis H4 concerning sufficiency of communication load to explain WL is valid only for single well defined work conditions. Communication load despite its strong contribution to WL (see Sect. 2), is not sufficient to explain the subjective workload without considering work condition. This appears not surprising when reminding the arguments for the ISA-bias increase from baseline $c = 0$ & 1 to $c = 2$ condition in Sect. 4.1 (increased visual scanning). Despite the evidence for a power law relationship ISA (n, c) contrasts to communication load RC(n, c) with the possibility of communication strategy change under high load (Sperandio, 1978, see Sect. 4.2), and high ATC task complexity in general (e.g. Manning et al., 2002).

7 Conclusion

For the new remote tower center work environment (MRT) a HitL simulation experiment with 12 tower controllers had been performed. Workload is a crucial factor for workplace design and defining limits of operational concepts, for example airspace capacity or maximum number of aircraft allowed at one MRT working position. Based on the data gathered in this study, four hypotheses concerning dependence of subjective ISA workload and objective radio communication task load on the two independent variables traffic load n and work conditions c were tested. Two-factor ANOVAs with F-tests and Pearson correlation coefficients provided initial

evidence on: (1) increase of ISA-WL and radio communication load (radio calls frequency (RC) and cumulative call duration RD)) with environmental load *n*, and with transition from baseline work condition to MRT operation by a single controller (hypotheses H1, H2); (2) significant positive correlation between ISA-WL and communication load (H3), and (3) a mediator effect of radio communication (TL-variables RC, RD) between common environmental independent load variable *n* and WL, together with sufficiency of communication load to explain the reported subjective ISA-WL level (H4).

The explanatory power of ANOVAs is limited, because they are based on linear models. They do not consider psycho-physiological and theoretical models of workload which provide clear indications of nonlinear TL—WL interdependencies. Consequently these analyses were extended and put on a detailed nonlinear work and task load model based on the cognitive resource limitation hypothesis (Kahnemann, 1973; Wickens, 2002). The functional relationships of this non-linear workload model were derived and previously validated in (Fürstenau et al., 2020; Fürstenau & Radüntz, 2021) by an independent HitL approach-sector radar work place simulation with 21 domain experts. Specifically, the theoretical predictions together with regression based estimates of logistic sensitivity and offset parameters of the nonlinear ISA(n) and RC(n) models put the initial evidence for hypotheses H1 and H2 on a quantitative level. A specific feature of the two-parametric model based approach lies in the fact that the inclusion of prior knowledge (e.g. ISA scale limits, asymptotic load boundaries, capacity limit) besides theoretical parameter prediction also allows confidence increase of regression based estimates through reduction to a single free (sensitivity) parameter to be estimated.

The present results of the nonlinear (logistic) and generalized linear ISA(n) and RC(n) analysis quantify the ISA workload increase after transition from baseline to multiple airport single operator control in terms of an effect of WL-bias increase as dominant contribution and only a small change of WL-sensitivity, i.e. basically constant WL difference between conditions with increasing n. A plausible explanation is the subjective experience of doubling of visual scanning of two video-panoramas and other basic tasks like approach radar and weather display observation at the second AP which is nearly independent of the number of aircraft under control,

This agrees with the predicted and measured objective communication load RC (calls/2 min) (H2) that follows the logistic model even closer with regard to dependence on traffic load, however does not change with transition to single ATCO(2-airport) control up to a traffic level where the logistic nonlinearity due to the communication time limit becomes significant. When approaching the radio calls limit under *c* = 2 condition, operators apparently change communication strategy and reduce calls frequency, approaching a somewhat lower asymptote as compared to baseline condition when communicating either with pilots at the same airport or being supported by a coordinator (Sects. 4.2, 5.2.1, 5.2.2, and Manning et al., 2002; Sperandio, 1978). A similar observation was reported in (Fürstenau & Radüntz, 2021).

Concerning hypothesis H3 initial ANOVAs indicated significant linear correlations. Within the resource limitation model the possibility for combining logistic characteristics of communication task load and subjective ISA-WL into a single

model in terms of Stevens' stimulus—response power law was of particular interest (Sects. 2.2.5, 4 and Appendix D). In fact, Lehrer suggested in (Lehrer et al., 2010) the combined use of different measures due to well known large inter-individual differences in sensitivities because "it is known that some individuals respond more sensitively to task load changes in self-report measures, others in specific physiological measures". The resulting power law ISA(RC) dependency is quantified by the exponent $\gamma = \rho/\nu$, as ratio of logistic scaling (TL/WL-sensitivity) parameters. The fact that the ISA(RC) relationship can be formally transformed into the Stevens form of power law $P_I(ISA) \sim P_R(RC)^\gamma$ (Eq. (7)), with prediction γ of the order of 1, is taken as further support for a psychophysics character of the relationship between subjectively experienced workload and objective (communication) task load. The psychophysics approach to subjective workload was originally proposed by (Gopher & Braune, 1984) and recently applied to HitL simulations of aircraft control tasks by (Bachelder & Godfroy-Cooper, 2019).

The hypothesized mediator effect of communication TL (H4) between traffic load n and ISA-WL was supported within the initial ANOVAs by linear Pearson correlation coefficients. The detailed nonlinear power law analysis provided initial evidence for the mediator effect because the derivation of the Stevens stimulus (TL)—response (WL) function involved the elimination of the parametric dependence on external load *n*. For fixed work condition the measured (nonlinear) correlation between the physical (communication) stimulus and the subjective (ISA-WL) response support the assumption that the communication load is sufficient for explaining the subjective load. However as depicted clearly in Fig. 14, in contrast to ISA, RC(n) exhibits no bias difference between work conditions at low traffic load, and on the other hand at high traffic load it approaches a lower asymptote for single ATCO(2 airport) control as compared to baseline. ISA(RC|condition) is an ambiguous function that obviously requires measurement of additional activities or tasks in order to disambiguate the ISA(RC)-WL level with regard to work condition. A suitable measure for considering WL change due to work conditions (under constant traffic load) could be eye tracking for quantifying the visual scanning behavior or online analysis of communication content (speech analysis).

In conclusion, besides formal definition and quantification of characteristic WL and communication load parameters for different work conditions of the new MRTO work environment as extension of ANOVAs, the present model based data analysis of RTC human-in-the-loop simulations provided additional evidence for the psychophysics power law dependency between subjective ISA-WL and objective communication TL (rate of radio calls), as derived in (Fürstenau & Radüntz, 2021). Although there appears to exist sufficient quantitative evidence for the conclusions concerning WL increase under single-operator MRTO conditions, additional experiments are certainly required for improving significance of parameter estimates due to the relatively small participant sample, and for quantifying additional (TL) variables which might contribute to the measured WL. Furthermore, the integration of prior knowledge possibly could be improved by formally including distributions of prior parameter estimates using maximum likelihood estimates of posterior parameter distributions by means of the Bayesian method of conditional probability densities

(for a brief introduction into basics of the Bayes method as used in Chaps. 10 and 16 of the book, see Appendix 2).

Appendix

Here we present the measured pre-processed ISA and RC data as dependent on traffic load n. The numerical values represent averages across participants which are clustered with regard to equal traffic load n within 2 min time intervals (Table A1).

Because ATCO team 1 had to be excluded due to missing ISA data and in addition some individual cases (2 min intervals) had to be excluded due to missing data the original data volume of 728 distinct measurement pairs of ISA (AC, RC, Load) per 2 min interval was reduced to 405 cases for the initial ANOVA data analysis (Lange et al., 2011). For the shared task condition with two controllers ($c = 1$) only the workload ratings from the ATCO responsible for the communication with pilots was included in the evaluation (Tables A2, A3).

Table A1 Number of available measurement intervals N (2 min) per traffic flow value n (AC on ATCOs transmission frequency AC/ $(2\ min)^{-1}$) separated for conditions $c = 0, 1, 2$

n (AC / $(2\ min)^{-1}$)	$c = 0$	$c = 1$	$c = 2$	$\Sigma(c = 0,1,2)$
1	24	16	4	44
2	35	52	17	104
3	28	44	18	90
4	18	28	22	68
5	1	36	18	55
6	0	50	14	64
7	0	34	17	51
8	0	36	6	42
9	0	6	1	7
10	0	4	1	5
Σ	106	306	118	530

Table A2 Traffic flow values $n = 1\ldots10$ AC/2 min with means over N(n) individual cases (from 532 individual cases) averaged across conditions for dependent variables ISA, RC and relative cumulated radio call transmission time (= time pressure / % of 2 min interval) SD = standard deviation based on N(n) cases (2 min simulation intervals)

Traffic (AC/2 min	Number of cases	Workload ISA (1,…,5)		Radio call rate RC/(2 min)$^{-1}$		Time pressure RD/120 s (%)	
n	N	Mean	SD	Mean	SD	Mean	SD
1	44	1.43	0.62	1.86	1.47	0.05	0.06
2	104	1.48	0.57	3.56	2.06	0.14	0.07
3	90	1.61	0.68	4.42	2.08	0.16	0.07
4	70	2.04	0.85	5.39	1.63	0.21	0.08
5	55	2.24	0.94	6.36	1.94	0.26	0.09
6	64	2.66	0.88	7.25	1.65	0.31	0.09
7	51	3.25	1.04	8.39	1.50	0.35	0.07
8	42	3.26	0.80	8.12	1.76	0.35	0.09
9	7	4.00	0.58	9.00	1.53	0.37	0.11
10	5	4.00	0.00	6.80	0.84	0.24	0.05

Table A3 Mean values of 2 min intervals across N(n) measurements separated for experimental condition $c = 0, 1, 2$, of dependent variables ISA, RC, RD

n [AC/2 min]	Workload <<ISA>>			Radio call rate <<RC>>[calls/2 min]			Cumulative call time RD [sec/2 min]		
	$c = 0$	$c = 2$	$c = 1$	$c = 0$	$c = 2$	$c = 1$	$c = 0$	$c = 2$	$c = 1$
1	1.13	2.75	1.56	2.00	2.00	1.63	7.58	8.25	4.50
2	1.34	2.12	1.37	2.69	4.12	3.96	11.49	18.18	19.27
3	1.75	2.22	1.27	3.50	5.00	4.77	14.96	25.22	22.50
4	1.89	2.73	1.61	5.28	5.27	5.64	23.33	27.27	27.50
5	1.00	2.83	1.97	7.00	6.06	6.50	21.00	29.78	32.83
6		3.29	2.48		7.14	7.28		35.21	38.20
7		3.94	2.91		8.24	8.47		39.59	43.59
8		4.17	3.11		7.50	8.22		38.00	42.78
9		5.00	3.83		7.00	9.33		28.00	48.67
10		4.00	4.00		8.00	6.50		30.00	32.50

References

Bachelder, E., & Godfroy-Cooper, M. (2019). Pilot workload esimation: Synthesis of spectral requirements analysis and Weber's law. 322514/620191228

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Brennen, S. D. (1992). An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. Internal Report, DRA/TM/CAD5/92017, Defence Research Agency, Portsmouth.

Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary task, physioplogical, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors, 25*, 623–641. https://doi.org/10.1177/001872088 302500602

Djokic, J., Lorenz, B., & Fricke, H. (2010). Air traffic control complexity as workload driver. *Transportation Research Part C, 18*, 930–936.

Edwards, T. (2013). *Human performance in air traffic control*. University of Nottingham.

Flynn, G., Benkouar, A., & Christien, R. (2003). *Passimistic sector capacity estimation.* Eurocontrol.

Friedrich, M., Biermann, M., Gontar, P., Biella, M., & Bengler, K. (2018). The influence of task load on situation awareness and control strategy in the ATC tower environment. *Cognition, Technology, and Work, 20*(2), 205–217. https://doi.org/10.1007/s10111-018-0464-4

Fürstenau, N., & Radüntz, T. (2021). Power law model for Subjective Mental Workload and validation through air-traffic control human-in-the-loop simulation. *Cognition, Technology, and Work*. Retrieved 05 30, 2021. https://doi.org/10.1007/s101011-021-00681-0

Fürstenau, N., Radüntz, T., & Mühlhausen, T. (2020). Model based development of a mental workload sensitivity index for subject clustering. *Theoretical Issues in Ergonomics Science.* https://doi.org/10.1080/1463922X.2020.1711990

Fürstenau, N. (2016). *Virtual and remote control tower.* In N. Fürstenau, (Ed.). Springer.

Gianazza, D. (2017). Learning air traffic controller workload from past sector operations. In *Proceedings 12th ATM research and development seminar.* FAA & Eurocontrol. Retrieved from http://www.atmseminarus.org/seminarContent/seminar12/papers/12th_ATM_RD_Seminar_paper_37.pdf

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures. *Human Factors, 26*(5), 519–532.

Gopher, D., Chillag, N., & Arzi, N. (1985). The psychophysics of workload: A second look at the relationship between subjective measures and performance. In *Proceedings of the human factors society* (29th Annual Meeting), pp. 640–644.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland.

Johannsen, G., Morey, N., Pew, R., Rasmussen, J., Sanders, A., & Wickens, C. (1979). Final report of experimental psychology group. In N. Morey (Ed.), *Mental workload. Its theory and measurement* (NATO Conference Series ed., vol. 8, pp. 101–114). Springer.

Jordan, C. S., & Brennen, S. D. (1992). *Instantanous self-assessment of workload technique (ISA).* Internal Report, DRA/TM/CAD5, Defence Research Agency, Portsmouth, GB.

Jordan, C. (1992). *Experimental study of the effect of an instantaneous self assessment workload recorder.* Tech. rep. DRA/TM/CAD5/92011, DRA Maritime Command and Control Division, Portsmouth.

Josefson, B., Jakobi, J., Papenfuss, A., Schmidt, C., & Sedov, L. (2018). Identification of complexity factors for remote towers. In *Proceedings, SESAR innovation days.* Salzburg, Austria.

Kahnemann, D. (1973). *Attention and effort*. Prentice Hall.

Kahnemann, D. (2011). *Thinking, fast and slow (German edition, 2012, Siedler* (Verlag). Farrar, Straus, Giraux.

Kirwan, B., Evans, A., Donohoe, L., Kilner, A., Lamoureux, Atkinson, T., & MacKendrick, H. (1997). Human factors in the ATM system design life cycle. In *FAA/eurocontrol ATM R&D seminar.* Eurocontrol. Retrieved 30 Sept 2020, from http://www.atmseminar.org/seminarContent/seminar1/papers/p_007_CDR.pdf

Lambdin, C. (2012). Significance tests as scorcery: Science is empirical—significance tests are not. *Theory and Psychology, 22*(1), 67–90. https://doi.org/10.1177/0959354311429854

Lange, M., Möhlenbrink, C., & Papenfuss, A. (2011). *Analyse des Zusammenhangs zwischen dem Workload von Towerlotsen und objektiven Arbeitsparametern.* Internal Report IB112–2011/46, German Aerospace Center (DLR), Inst. of Flight Guidance, Braunschweig.

Lee, P. U., Mercer, J., Smith, N., & Palmer, E. (2005). A non-linear relationship between controller workload, task load, and traffic density: The straw that broke the camel's back. In *Proceedings of the 2005 international symposium aviation psychology*, (pp. 438–444). Retrieved 3 Nov 2021, from https://corescholar.libraries.wright.edu/isap_2005/66

Lee, P. U. (2005). A non-linear relationship between controller workload and traffic count. In *Proceedings of the human factors and ergonomics society, 49th meeting 2005* (vol. 49, pp. 1129–1133). Retrieved from https://doi.org/10.1177/154193120504901206

Lehrer, P., Karavidas, M., Lu, S.-E., Vaschillo, E., Vaschillo, B., & Cheng, A. (2010). Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study. *International Journal of Psychophysiology, 76*, 80–87. https://doi.org/10.1016/j.ijpsycho.2010.02.006

Manning, C. A., Mills, S. H., Fox, C. M., Pfleiderer, E. M., & Mogilka, H. J. (2002). *Using air traffic control task load measures and communication events to predict subjective workload.* No. DOT/FAA/AM 02/4, DOT/FAA.

Möhlenbrink. (2011). *Influence of workplace design on workload.* DLR.

Möhlenbrink, C., Papenfuss, A., & Jakobi, J. (2011). The role of workload for work organisation in a Remote Tower Control Center. In *Proceedings of the 91th USA/Europe air traffic management research and development seminar (ATM2011).*

Moray, N. (1982). Subjective mental workload. *Human factors*, pp. 25–40.

Moray, N. (1988). Mental workload since 1979. *International review of ergonomics*, 123–150.

Papenfuss, A. (2013). Phenotypes of Teamwork—an exploratory study of tower controller teams. In HFES (Ed.), *Proceedings human factors and ergonomics society annual meeting.* San Diego.

Russo, R. (2016). CAPAN methodology: Sector capacity assessment. *Air traffic services system capacity seminar.* ICAO.

Sperandio, J. C. (1978). The regulation of working methods as function of workload among air traffic controllers. *Ergonomics, 21*(3), 195–202.

Stein, E. (1985). *Air traffic controller workload: An examination of workload probe.* DOT/FAA/CT-TN84/24, DOT/FAA.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153–181.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural and social prospects.* Wiley.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as measure of workload. *Ergonomics, 39*(5), 740–748.

Wickens, C. D., & Hollands, J. G. (2000a). Attention, time sharing, and workload. In *Engineering psychology and human performance* (3 edn, pp. 439–479). Prentice-Hall.

Wickens, C., & Hollands, J. (2000b). *Engineering psychology and human performance.*

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. Retrieved from https://doi.org/10.1080/14639220210123806

Changing of the Guards: The Impact of Handover Procedures on Human Performance in Multiple Remote Tower Operations

Anneke Hamann and Jörn Jakobi

Abstract Multiple Remote Tower Operations (MRTO) change the way air traffic is managed. In this concept, air traffic control officers (ATCOs) operate several aerodromes simultaneously from a specially designed working position, also referred to as a multiple remote tower module (MRTM). This change in operations also introduces significant changes in the ATCOs' workflow and cognitive demands. In theory MRTO can facilitate the ATCOs' ability to balance their mental workload through a flexible allocation of aerodromes to each MRTM, but new procedures need to be implemented to enable such flexible allocations: Appropriate handover procedures are needed to transfer aerodromes between MRTMs and their operators. This paper investigated the feasibility of handover procedures during simulated air traffic control as a mitigation to counteract inappropriate mental workload. In a human-in-the-loop real-time simulation, six ATCOs completed traffic scenarios with or without handover via two MRTM, dealing with a total of three aerodromes. Descriptive data showed no adverse short-term effects caused by the handovers and indicated possible beneficial long-term effects on cognitive capacity and safety. The handover procedures were overall feasible and accepted by the ATCOs, as a strategy to better balance mental workload in MRTO.

Keywords Human performance · Mental workload · Situation awareness · Air traffic control · Remote tower · Multiple remote tower

1 Introduction

In recent years, with the development of new technologies, the concept of Remote Tower Operations (RTO; for an overview see Fürstenau, 2016) has gained much

A. Hamann (✉) · J. Jakobi
Human Factors, Institute of Flight Guidance, German Aerospace Center (DLR), Braunschweig, Germany
e-mail: anneke.hamann@dlr.de

J. Jakobi
e-mail: joern.jakobi@dlr.de

attention. RTO describes the remote surveillance of an aerodrome by means of live video feeds, as opposed to the classical out-of-the-window view from the tower. Multiple Remote Tower Operations (MRTO) take this concept a step further. Two or more remotely controlled aerodromes are managed simultaneously by one air traffic control officer (ATCO) from one Multiple Remote Tower Module (MRTM). This offers the possibility of a more flexible ATCO-aerodrome allocation, matching the actual traffic situation, e.g. by combining several smaller aerodromes in one MRTM during times of low traffic (Jakobi et al., 2019). In fact, MRTO is a concept addressing Air Traffic Services (ATS) as a whole, including air traffic advisory service, flight information service and alerting service, but this paper focusses on air traffic control service (ATC) provided by ATCOs. MRTO lead to changes in ATCOs' roles and responsibilities when controlling more than one aerodrome, and adds complexity while the basic cognitive and task demands largely remain the same (Jakobi et al., 2019).

ATCOs are expected to assure a safe, efficient and orderly flow of air traffic (Mensen, 2014)—whether they work in a conventional tower, an RTO or an MRTO workplace. In order to do that, they need to integrate information from various sources, form a mental picture of the situation and its future development, communicate effectively with other stations, such as approach and meteorology service units, and with pilots, and issue timely commands. In sum, ATC requires high levels of situation awareness (SA), and in turn high working memory capacity and attention to constantly process new information (Endsley, 1999). Such cognitive resources, however, are limited (Kahnemann, 1973; Wickens, 2002). The extent to which one's cognitive resources are used up by task demands are defined as mental workload (MWL). Excessive MWL may lead to cognitive overload, a reduction of performance, and eventually errors (Endsley, 1999; Stokes & Kite, 1997). In air traffic research, specifically, increasing MWL has been tied to a performance reduction in ATCOs (Brookings et al., 1996) and in piloting tasks (Causse et al., 2015). MWL should be kept at an acceptable level, and both too high (excessive) and too low (underutilized) MWL should be avoided in order not to impair safety (Weinger, 1999).

While prolonged periods of high MWL should be avoided, ATCOs are trained to use strategies to cope during short periods of high task load (Möhlenbrink et al., 2012). Strategies include prioritizing tasks or applying procedures to reduce the complexity of challenging situations, e.g. deferring departure flights, keeping arrival flights in holding patterns, coordinating with adjacent ATS units for delaying inbound traffic, or asking a supervisor or colleague ATCO for assistance (Möhlenbrink et al., 2012). Especially on bigger aerodromes, however, complex or highly demanding traffic situations cannot be avoided and the ATCOs' personal coping skills and strategies may be insufficient to counteract high MWL. This requires basic organizational countermeasures like shorter shifts, more frequent breaks, splitting the ATC task into several roles (clearance delivery, ground, or runway/local controller), permanent double staffing (executive and planning ATCO), and supervisor positions. In addition, technical support can further reduce task demands, e.g. with approach radar, surface

movement radar, electronic flight strips with planning assistance functionality, or conflict monitoring tools.

All these considerations also apply to MRTO, where controlling several aerodromes adds to the complexity of the ATC task. However, some coping mechanisms that are effective for controlling only one aerodrome may prove ineffective for MRTO. A complex situation or incident occurring at one aerodrome may indirectly influence all other aerodromes controlled by the same ATCO, because of the necessity to focus their limited cognitive resources on a single problem. Coordination with adjacent ATS units or aerodrome services is multiplied by the number of aerodromes in the MRTM, and delays due to holding traffic may affect all aerodromes instead of only one. There is a need for appropriate procedures to ensure that ATCOs working in MRTO can balance their MWL and assure a safe, efficient and orderly flow of air traffic just as well as ATCOs working in RTO or conventional towers. One possible strategy is handover procedures. Handover procedures, in this paper "handovers", allow the swift transfer of one or more aerodromes from one ATCO and MRTM to another. This will enable ATCOs to reduce their task load in order to counteract high MWL or concentrate on one incident, as well as ensure that the capacity of, or service level provided to the other aerodromes will not be reduced.

The aim of this paper is to assess the general feasibility of handovers in MRTO by investigating their impact on ATCOs' mental capacity. In this study, we focused on the ATCOs handing over aerodromes to a colleague ATCO. We hypothesized that the procedures would not induce negative short-term effects on safety and ATCOs' MWL and SA during handover. Furthermore, we expected beneficial long-term effects for the ATCOs who handed over aerodromes, indicated by lower MWL and higher SA after handover compared to without.

In the present paper, we present questionnaire data. Concurrent eye tracking data are analysed and presented separately in Friedrich et al. (2020).

2 Method

2.1 Participants

Six active Air Traffic Control Officers (ATCOs) from a Northern European air navigation service provider took part in the experiment. All were male, aged between 25 and 37 years ($M = 29.6$, $SD = 3.9$) and with job experience ranging between 1.5 and 8 years ($M = 3.9$, $SD = 2.4$). The ATCOs participated voluntarily during their working hours. The study was performed in accordance with the General Data Protection Regulation (EU) 2016/679.

2.2   Design and Material

The study was conducted in the Tower Lab research facility at DLR Braunschweig. Two MRTMs as shown in Fig. 1 were provided. Up to three aerodromes can be operated from one MRTM. Each MRTM consisted of the following parts: panoramic view (208° horizontal and 32° vertical) and panel for a pan-tilt-zoom camera for each aerodrome, stacked on top of each other; radar, and electronic flight strip system (Frequentis AG, Vienna, Austria) for each aerodrome in corresponding order from left to right; radio communication with coupled frequency for all three aerodromes; separate telephone connection for each aerodrome for local aerodrome services. The experiment was performed as a human-in-the-loop real time simulation on an NLR Air Traffic Control Research Simulator (NARSIM; Have, 1993).

Two independent variables (IV-A and IV-B) were varied in two levels each (see Table 1). IV-A "Non-Nominal Situation" varied in A1 "Increased Traffic Load", and A2 "Emergency". IV-B "Handover" varied in B1 "Without", and B2 "With". Every

Fig. 1 Multiple remote tower module (MRTM) as used in the study (two active ATCOs with an observer right beside them)

Table 1 Experimental design with independent variables (IV-A, IV-B) forming Four experimental conditions (EC)

		IV-A Non-nominal situation	
		A1 *Increased traffic load*	A2 *Emergency*
IV-B Handover	B1 *Without*	EC_{11}	EC_{21}
	B2 *With*	EC_{12}	EC_{22}

participant completed every experimental condition. This resulted in a complete 2×2 factor within-subject design. To reduce learning effects, aircraft call signs were varied and arrival and departure times shifted slightly between conditions.

The ATCOs controlled up to three aerodromes in parallel, two small and a medium-sized one. In each experimental condition the ATCOs had to control a traffic scenario with an average traffic load of 28 movements per hour (ground vehicles included). The traffic was composed of 90% IFR (instrument flight rules) and 10% VFR (visual flight rules) traffic. The overall traffic load was unevenly distributed between the aerodromes with the larger aerodrome accounting for approx. 50% of the traffic and the smaller aerodromes for 25% each. The simulated weather always met visual meteorological conditions (VMC) with no clouds. The time of day was always day time.

In order to enable handover procedures, the six participants worked together in groups of two, forming three dyads. Each participant completed all experimental conditions as Lead ATCO (the one handing over traffic) and as Support ATCO (the one receiving traffic). For conditions without handover ("*Without*"; EC_{11}, EC_{21}) only the participant performing the role of the Lead ATCO was present and no traffic was handed over. The conditions were presented in a pseudo-randomized order (see Sect. 2.2.3), for a detailed description). The MRTM station on the right side of the room was always the main MRTM occupied by the Lead ATCO. The left MRTM was only opened and manned with the Support ATCO in experimental conditions involving a handover ("*With*"; EC_{12}, EC_{22}).

2.2.1 Implementation of the UV-B "Handover"

The handover procedures were designed in a human-centred approach with feedback from ATCOs prior to the study, and provided a scaffold for all relevant information to be shared. The participants were asked to perform handovers in a standardized way: The Lead ATCO was permitted to request a colleague to take over aerodromes. If no request was made by a pre-defined time, the decision to open the second position was made by a confederate supervisor in order to ensure the experimental procedure was carried out as planned. The aerodromes to be handed over were fixed pre-experiment. The Support ATCO was then fetched from the break room by the confederate supervisor. The Support ATCO sat down at their work station and would turn on the panoramic view and radio communication for the aerodromes the Lead ATCO was working on. This way the Support ATCO was able to gather information on the situation in order to build up a mental traffic picture prior to the actual handover. The Lead ATCO initiated the handover and ensured that the Support ATCO was ready to take over. Both went through the standardized handover checklist to ensure all information was shared, with the Lead ATCO giving information and the Support ATCO confirming the correct understanding. At the end of the handover, the Lead ATCO gave control to the Support ATCO, who confirmed the takeover. The Lead ATCO then turned off their panoramic view and radio communication for the aerodrome handed over to Support. If two aerodromes were to be handed over, this

was done in consecutive order and both ATCOs would run through the checklist twice. Handovers could always be interrupted by incoming radio communication. One full handover procedure took approximately 20–30 s, depending on traffic and runway conditions.

2.2.2 Levels of the UV-A "Non-nominal Situation"

Level "Increased Traffic Load"

One experimental run lasted approx. 55 min. It included a traffic peak around minute 15 (increasing and coinciding traffic on all three aerodromes) and a bird strike (i.e. collision between a bird and a departing aircraft) on one aerodrome around minute 40. In the "*With*" handover experimental condition, the Lead ATCO was informed about the traffic peak by a confederate Supervisor (DLR member) five minutes in advance and asked to hand over one predefined aerodrome to the Support ATCO. This aerodrome was the one with the bird strike. Around minute 45 the Lead ATCO was told to prepare to receive back the aerodrome, when the bird strike incident was solved. The Lead ATCO continued working on three aerodromes until the end of the run. In the "*Without*" handover condition, no handover took place and the ATCO had to manage both the traffic peak and bird strike.

Level "Emergency"

One experimental run lasted approx. 40 min. Around minute 10, the Lead ATCO received a call informing them about an incoming aircraft with an emergency. The aircraft was scheduled for minute 25 so the Lead ATCO had time to prepare. In the "*With*" handover condition they should request the Support ATCO and hand over the two aerodromes that were not affected by the emergency. The aerodromes were not handed back later and the Lead ATCO kept working with only one aerodrome until the end of the run. In the "*Without*" handover condition, no handover took place. The nature of the emergency was varied between "*With*" and "*Without*" handover to counteract learning effects. For "*With*", an engine failure and for "*Without*" a medical emergency was announced, both situations in which the pilots were forced to send a "Pan-Pan" urgency signal call. The affected aircraft, timing and response vehicles (fire trucks for the engine failure, ambulance for the medical emergency) were varied and the nature of the respective emergency required the ATCOs to ask the pilots for different information. Both situations, however, resulted in temporary closure of the runway and the need for coordination with aerodrome and approach service officers.

2.2.3 Experimental Protocol

The ATCOs participated in fixed dyads, i.e. the same two ATCOs worked together for the duration of the experiment. They completed all experimental conditions on two consecutive days, three runs on day 1 and five runs on day 2. On day 1 the ATCOs received information about the aim of the study and gave informed written consent. They were then instructed how to operate the MRTM and received 80 min of training including handovers (40 min each as Lead and Support). They then proceeded to the experimental conditions. In total, every participant completed six runs: Two "*Without*" handover (i.e. as the only ATCO), and four "*With*" handover (i.e. two as Lead ATCO and two as Support ATCO). In a "*Without*" handover condition only one participant would complete the run while the other one could take a break. The order of the conditions was pseudo-randomized, such that one participant would never be Lead ATCO (or only ATCO in "*Without*" handover conditions) in consecutive runs. This resulted in an alternating sequence of being Lead/only ATCO and Support ATCO/having a break.

On day 2 the experiment continued and was concluded with a debriefing session. At the beginning of each run the ATCOs controlled all three aerodromes. The ATCOs were told they would encounter special situations during the experiment and could request their colleague as Support. A DLR member served as a Supervisor and informed the ATCO when a handover was necessary or, in case of a "*Without*" handover condition, not possible.

2.2.4 Human Performance Assessment

Mental Workload

Mental workload (MWL) was assessed with the short version of the Assessing the Impact on Mental Workload questionnaire (AIM-s; Dehn, 2008), Bedford scale (Roscoe, 1984; Roscoe & Ellis, 1990) and the Instantaneous Self-Assessment of Workload technique (ISA; Tattersall, 1994; Tattersall & Foord, 1996). The AIM-s assesses the impact of automation on MWL. It is rated on a 0–6 scale ("strongly disagree" to "strongly agree") with 6 indicating the highest MWL. The questionnaire consists of 16 items, of which only 14 were used in the experiment. Two items were excluded because they focused on team interactions that were not part of the experiment. The remaining 14 items were then averaged to provide a final score.

The Bedford scale promotes self-assessment of the experienced MWL on a 1–10 scale (1 = insignificant; 10 = unable to perform task).

The ISA scale is used to assess current MWL during the task. It is answered on a 0–5 scale (0 = underutilized; 5 = excessive). Every five minutes the participants were asked to rate their MWL of the previous five-minute period.

Situation Awareness

Situation awareness (SA) was assessed with the China Lake Situation Awareness scale (CLSA; Adams et al., 1998) and the Situation Awareness for SHAPE questionnaire (SASHA; Dehn, 2008). The CLSA assesses SA on a 1–10 scale (1 = far too low; 10 = excellent). The SASHA questionnaire addresses SA in six items on a 0–6 scale ("strongly disagree" to "strongly agree") with 6 indicating the best SA. The six items are then averaged to provide a final score.

Safety

The Cooper-Harper scale (Cooper & Harper, 1969) adapted to fit the ATC context was used for retrospective self-assess of safety impairment. ATCOs were asked to rate the most challenging or critical situation from the previous run. The adapted Cooper-Harper scale is rated from 1–10. Values 1–3 indicate no to minor impairment (low to slightly increased MWL). Values 4–6 indicate an impairment of efficiency (having caused "minor unpleasant" to "very disturbing" traffic delays). Values 7–10 indicate an impairment of safety ("loss of ability to plan ahead" to "not being able to control the traffic any more").

Impact of Handover Procedures

Additional tailored questions were asked to assess the ATCOs' experience of the handover procedures.

After "*Without*" handover conditions, ATCOs were asked if a handover would have helped them to balance their WL and SA. After "*With*" Handover conditions, ATCOs were asked which impact the handover procedure had on their WL and SA, and whether they considered a handover an appropriate measure during traffic peaks or emergency situations. Answers were given on a 0–6 scale ("strongly disagree" to "strongly agree").

"*Without*" handover conditions:

1. I felt confident I could handle the traffic on my own.
2. Handing over traffic to a colleague would have helped me maintain my situation awareness.
3. Handing over traffic to a colleague would have helped me balance my workload.

"*With*" handover conditions—Increased Traffic Load/[Emergency]:

1. I was able to hand over an aerodrome [two aerodromes] to my ATCO colleague in a safe and efficient way.
2. During the handover procedure I lost track of the traffic.
3. Handing over the aerodrome to my colleague was demanding.

4. Handover is an appropriate measure to counteract <u>high task load</u> [emergency] situations.

2.2.5 Analysis

For this paper, only data from the "*Without*" handover conditions and the Lead ATCOs' data from the "*With*" handover conditions were evaluated as the aim was to assess the impact of the handover procedure on ATCOs handing over aerodromes to a colleague. Comparisons are made between the two levels of IV-B ("*Without*" vs. "*With*" handover) in each IV-A level ("*Increased Traffic Load*" and "*Emergency*"). The data from the tailored questionnaires on handover procedures are evaluated separately for "*With*" and "*Without*" handover conditions as the questions differed between the conditions. Because of the low sample size, only descriptive data are reported.

3   Results

Due to the small sample size the data are not normally distributed and show high variance, as can be seen in high standard deviations for most of the results. Between-subject outliers were plotted and visual inspection showed that the high variance was not due to the answer tendencies of particular participants. Therefore, no outliers were excluded. Because of these limitations, the results should be interpreted cautiously.

3.1   Mental Workload

The success of the experimental manipulation of MWL was checked using the ISA data. The introduction of special situations (traffic peak and bird strike for "*Increased Traffic Load*"; aircraft emergency in "*Emergency*") is associated with rising MWL levels in the "*Without*" handover conditions (see Table 2 and Fig. 2). This can be interpreted as a successful induction of an increase in MWL.

For possible beneficial effects of handovers on MWL, the conditions "*With*" and "*Without*" are compared. In the "*Increased Traffic Load*" conditions (Fig. 2, upper panel), mean values for "*With*" handover were constantly lower than for "*Without*". Additionally, in the "*With*" condition the MWL peak in minute 45 (bird strike) is not visible: The incident did not affect the Lead ATCO as it happened on the aerodrome given over to the Support ATCO. In the "*Emergency*" conditions (Fig. 2, lower panel), mean ISA scores did not differ between "*With*" and "*Without*" handover until the onset of the special situation (i.e. information about the emergency aircraft). After situation onset and handover, the mean scores were higher in the "*Without*" handover condition compared to "*With*".

Table 2 ISA data on mental workload (lead ATCOs only)

Measure (values)	IV-A	IV-B	Mean ISA values after x minutes (SD)										
			05	10	15	20	25	30	35	40	45	50	55
ISA (1–5)	Increased traffic load	Without	1.60 (0.55)	3.20 (0.84)	3.00 (0.71)	3.80 (0.84)	3.40 (1.14)	2.80 (0.84)	2.50 (1.05)	3.40 (1.14)	3.83 (0.75)	3.17 (1.17)	2.60 (1.14)
		With	1.00 (0.00)	2.67 (0.52)	2.33 (0.52)	3.00 (0.63)	2.67 (0.82)	1.83 (0.41)	2.00 (0.63)	2.60 (0.89)	2.50 (1.05)	2.67 (0.52)	2.00 (0.71)
	Emergency	Without	1.00 (0.00)	2.75 (0.50)	3.20 (0.84)	3.40 (0.55)	3.00 (0.82)	2.80 (0.84)	3.25 (0.50)	–	–	–	–
		With	1.17 (0.41)	3.00 (0.63)	3.33 (0.82)	3.00 (0.89)	2.40 (0.55)	2.33 (0.82)	2.17 (0.41)	–	–	–	–
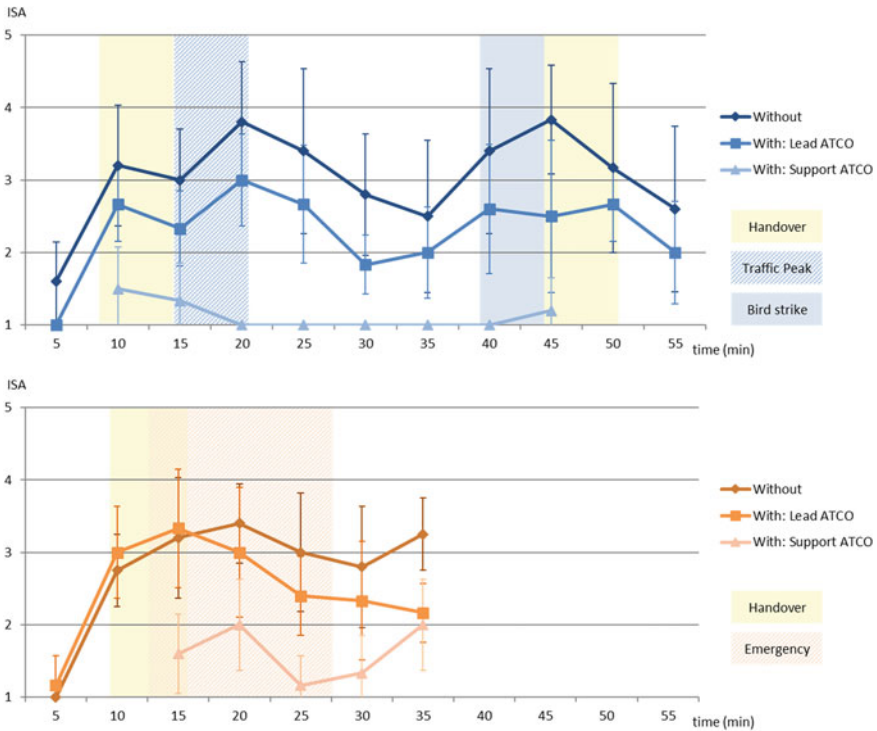
Fig. 2 Mean ISA values with standard deviations over time for "Increased Traffic Load" (upper panel) and "Emergency" (lower panel) conditions. Highlighted areas indicate time windows in which special situations took place. Onset and duration of the special situations could vary depending on ATCOs' actions and the simulated traffic

Exploratory evaluation of corresponding Support ATCOs' ISA data indicates overall low MWL levels (Fig. 2). In the "*Increased Traffic Load*" condition, slightly higher values can be observed during the initial handover, followed by a period of very low MWL. The special situation (i.e. bird strike) did not inflict a meaningful increase in MWL. In the "*Emergency*" condition, ISA scores varied more and were higher than in the "*Increased Traffic Load*" condition, indicating a possible main effect of IV-A (non-nominal situation) due to the amount of traffic the Support ATCOs received.

MWL ratings collected post-run indicate medium to slightly elevated MWL levels for all experimental conditions (Table 3). In the "*Increased Traffic Load*" condition, data of the Bedford scale show a lower mean and lower standard deviations in the "*With*" handover condition compared to "*Without*". This could indicate a possible beneficial effect of handover in terms of lower and more equal MWL ratings. This effect cannot be found in the "*Emergency*" conditions (Fig. 3). AIM-s data show no noticeable differences between conditions (Fig. 4).

Table 3 Questionnaire data on mental workload

Measure (values)	IV-A	IV-B	Min	Max	M	SD
Bedford scale (1–10)	Increased traffic load	Without	3.00	9.00	6.00	2.76
		With	3.00	5.00	3.67	0.82
	Emergency	Without	4.00	8.00	5.17	1.47
		With	3.00	8.00	5.33	1.63
AIM-s (0–6)	Increased traffic load	Without	2.29	5.21	3.70	1.15
		With	1.14	5.00	3.38	1.33
	Emergency	Without	2.36	4.93	3.49	0.95
		With	3.21	4.36	3.73	0.47

Fig. 3 Mean values of the Bedford Scale MWL assessment (error bars indicate SD)
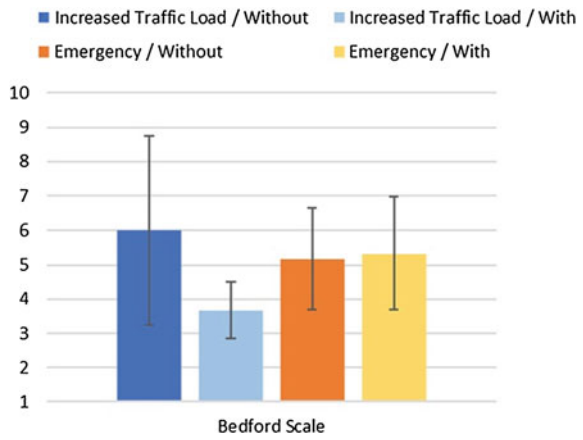
Fig. 4 Mean values of the AIM-s MWL assessment (error bars indicate SD)
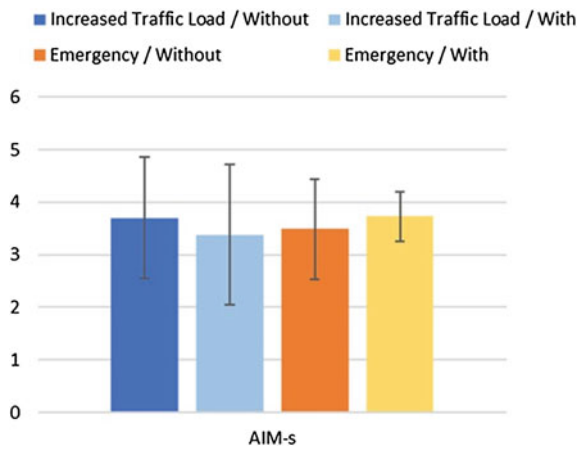
In sum, these findings show successful MWL manipulation and hint to a main effect of IV-B (handover) on MWL.

3.2 Situation Awareness

Retrospective SA ratings indicate medium to high SA levels for all experimental conditions (Table 4). In the "*Increased Traffic Load*" conditions, CLSA data indicate slightly higher mean values and lower standard deviations for "*With*" than "*Without*" (Fig. 5). This could indicate a possible beneficial effect of handover in terms of higher and more equal SA ratings. This effect cannot be found in the "*Emergency*" conditions.

Data of the SASHA show very similar means and standard deviations for both "*Increased Traffic Load*" conditions (Fig. 6). In comparison, the

Table 4 Questionnaire data on situation awareness

Measure (values)	IV-A	IV-B	Min	Max	M	SD
CLSA (1–10)	Increased traffic load	Without	5.00	8.00	6.83	1.47
		With	7.00	9.00	7.67	0.82
	Emergency	Without	5.00	8.00	6.67	1.03
		With	5.00	8.00	6.83	0.98
SASHA (0–6)	Increased traffic load	Without	3.00	4.83	3.92	0.81
		With	2.17	5.00	3.94	0.95
	Emergency	Without	2.83	4.83	3.92	0.74
		With	2.83	4.50	3.56	0.66

Fig. 5 Mean values of the China Lake SA assessment (error bars indicate SD)

Legend: ■ Increased Traffic Load / Without ■ Increased Traffic Load / With ■ Emergency / Without ■ Emergency / With

China Lake Situation Awareness Scale
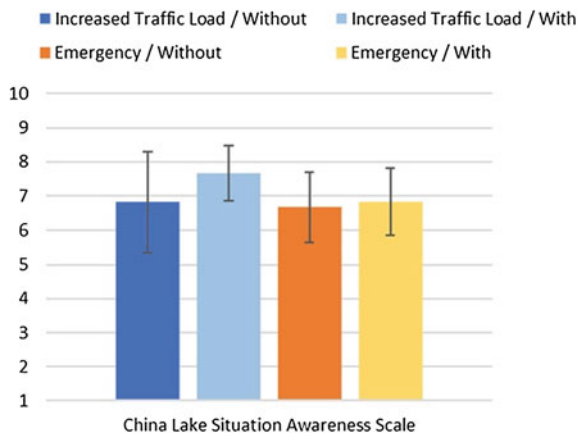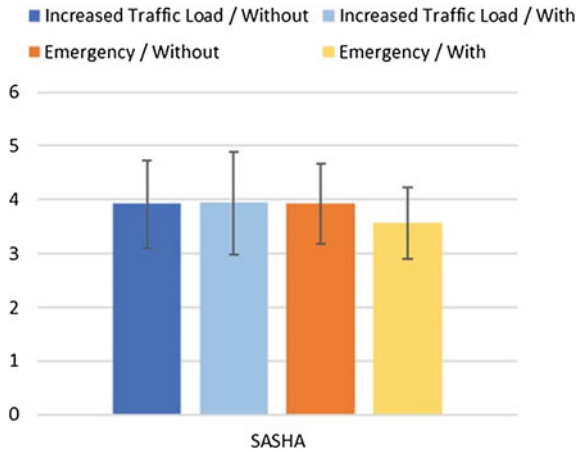
**Fig. 6** Mean values of the SASHA SA assessment (error bars indicate SD)

"*Emergency/With*" condition displays a slightly lower mean and standard deviation than "*Emergency/Without*". In sum, no clear effect of handover on SA could be found.

3.3 Safety

Retrospective self-assessment of critical situations shows no impairment of safety. Maximum values indicate impairment of efficiency (values 4–6), with "minor unpleasant delays" for the "*Emergency/Without*" condition and "very disturbing delays" for all other conditions. Regarding mean values, the "*Emergency/With*" condition shows the highest efficiency impairment rating. This could indicate a possible negative effect of handover in emergency situations in terms of efficiency. In sum, no safety impairments and no clear effect of handover on safety could be found (Table 5 and Fig. 7).

Table 5 Adapted Cooper-Harper scale data on safety

Measure (values)	IV-A	IV-B	Min	Max	M	SD
Adapted Cooper-Harper scale (1–10)	Increased traffic load	Without	3.00	6.00	4.17	1.17
		With	2.00	6.00	3.83	1.33
	Emergency	Without	3.00	4.00	3.67	0.52
		With	3.00	6.00	4.67	1.21
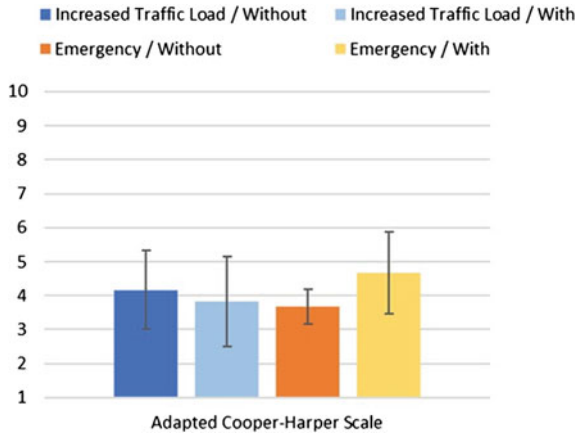
Fig. 7 Mean values of the adapted Cooper-Harper scale for all four experimental conditions (error bars indicate SD)
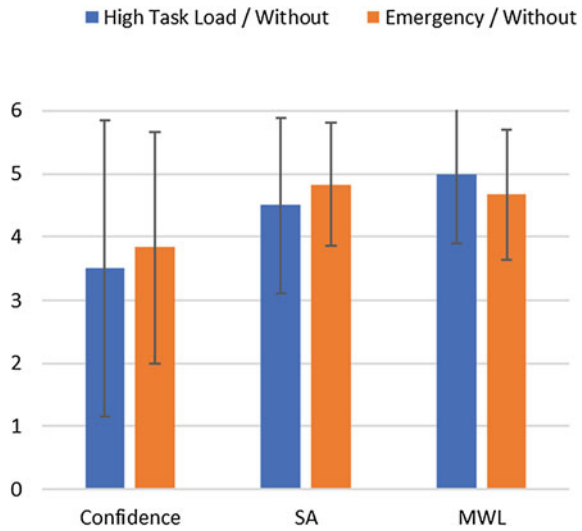
3.4 Impact of Handover Procedures

In "*Without*" handover conditions, mean values indicate a medium to high confidence in the participants' own ability to handle the traffic on their own (see Table 6 and Fig. 8). A high standard deviation and large data range indicate large individual differences, with some participants stating very low confidence. Concerning the questions regarding a possible benefit of handover on SA and MWL, data show medium to high agreement. This indicates the participants would have welcomed the possibility of a handover. The findings did not differ between IV-A (non-nominal situations) conditions.

In "*With*" handover conditions (see Table 7 and Fig. 9), the ratings indicate large individual differences between participants for SA and MWL, regardless of the IV-A level. The ability to hand over aerodromes, too, shows large differences in "*Increased Traffic Load*" but not in "*Emergency*" that poses an exception with strong agreement between participants. Participants unanimously viewed handover procedures

Table 6 Questionnaire data on handover procedures—"Without" handover

Measure (values)	IV-A	Min	Max	M	SD
Confidence handling traffic alone (0–6)	Increased traffic load	0.00	5.00	3.50	2.35
	Emergency	1.00	5.00	3.83	1.84
Benefit of handover on SA (0–6)	Increased traffic load	3.00	6.00	4.50	1.38
	Emergency	3.00	6.00	4.83	0.98
Benefit of handover on MWL (0–6)	Increased traffic load	3.00	6.00	5.00	1.10
	Emergency	3.00	6.00	4.67	1.03

Fig. 8 Mean values of the tailored questions on handover procedures—"Without" conditions (error bars indicate SD)

**Table 7** Questionnaire data on handover procedures—"With" handover

Measure (values)	IV-A	Min	Max	M	SD
Ability to hand over aerodrome(s) (0–6)	Increased traffic load	1.00	6.00	4.17	1.72
	Emergency	4.00	5.00	4.50	0.55
Loss of SA during handover (0–6)	Increased traffic load	0.00	4.00	1.83	1.47
	Emergency	1.00	4.00	2.00	1.27
Increase in MWL during handover(0–6)	Increased traffic load	1.00	5.00	3.17	1.72
	Emergency	2.00	5.00	3.17	1.17
Handover as appropriate mitigation (0–6)	Increased traffic load	4.00	6.00	4.83	0.75
	Emergency	4.00	6.00	5.17	0.75

as an adequate countermeasure for situations involving increased traffic load and emergencies, as seen by high mean values and comparably low standard deviations.

In sum, tailored questions show that the impact of handovers on participants differed substantially, yet all participants considered handovers an appropriate mitigation for demanding situations.

4   Discussion

The aim of this paper was to assess the general feasibility of handover procedures in an MRTO setting by investigating their impact on safety, and ATCOs' MWL and SA. This paper focused on the ATCOs dealing with three aerodromes of which one or two
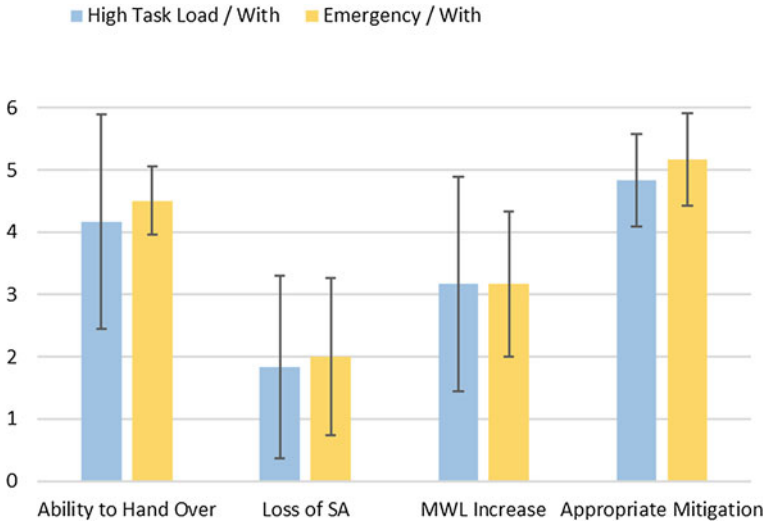
Fig. 9 Mean values of the tailored questions on handover procedures—"With" conditions (error bars indicate SD)

could be handed over to a colleague ATCO. We hypothesized that handover procedures would not cause immediate negative effects on MWL, SA and safety during handover (no negative short-term effects), and would in turn lead to a lasting reduction in MWL and increase in SA after handover procedures were applied (beneficial long-term effects).

Given the small sample size, no inferential statistical analyses were performed and descriptive data was presented instead. Overall, high standard deviations were observed, indicating large individual differences. The results therefore likely reflect participants' individual skills, attitudes, and a complicated interaction between experimental conditions, own actions and interaction with the team partner. We therefore very cautiously draw conclusions and give recommendations for further studies.

Results of the ISA data indicate that our manipulation of MWL by means of special situations worked. Yet, we did not induce excessive MWL in conditions without handover. Especially the emergency situation seems not to have elicited excessive cognitive demands. The hypothesized beneficial effect of handovers should manifest itself most prominently in situations where the ATCOs' mental capacity is exceeded and they need to use a coping strategy. It is therefore possible that our MWL manipulation was not strong enough to show the hypothesized handover effects. Future studies should use a higher traffic volume and/or more challenging situations in order to induce higher cognitive demands.

The handover procedures performed in this study did not seem to have interfered with participants' ability to perform their tasks. On average, their SA and MWL remained on a medium level during handover, but showed high individual differences. This could be explained with traffic at hand during handover that may have

disrupted the procedure and caused additional cognitive load in some situations. Yet, the participants were able to hand over the aerodromes in a safe and efficient way and viewed handovers as an appropriate mitigation to counteract demanding situations. Furthermore, they would have welcomed the possibility of a handover in runs where this was not possible. These findings highlight that while the handover process itself may add additional load in certain situations, ATCOs consider handovers a useful strategy to balance their MWL in demanding situations.

Concerning beneficial long-term effects of handovers, our data show mixed results. There is first evidence that handovers may be able to reduce MWL by reducing task load in the long term. These effects seem stronger for increased traffic load than emergency situations. Our data did not show a clear direction of the effect of handovers on SA. Similar to the MWL effects, handovers may be able to increase and equalize SA across participants for increased traffic load situations. By contrast, SA was reduced in emergency situations with handover. This pattern could also be observed in the safety ratings. Even though safety was never impaired in any condition, efficiency was impacted most in the "*Emergency/With*" handover condition.

This could point towards an adverse effect of handover in emergency situations or an unfavourable interaction of both variables. An alternative explanation lies in the nature of the emergency situations used in this study. As pointed out in section "Level "Emergency"", we used two different emergencies to counteract learning effects. The emergency used in runs without handover was a medical emergency aboard an aircraft, while the emergency in runs with handover was an aircraft with an engine failure. Even though the actions to be undertaken by the ATCOs were largely the same (runway closure, coordination with the pilot and other units, etc.), the engine failure emergency may have been perceived as more difficult. With the possibility of a fire and casualties upon landing, the consequences of this emergency, even though simulated, may have seemed more severe. This could have induced more stress and forced the ATCO to direct more cognitive resources towards the situation than in the medical emergency situation. This way the nature of the emergency would have impacted the retrospective SA and safety assessments more than the handover, and positive handover effects could have been masked. This explanation is also supported by the MWL findings: While the on-task assessment (ISA) showed lower MWL levels following handovers in all conditions, the retrospective assessments only show this effect for the increased traffic load situation and no difference between the emergency conditions. The comparability of our emergency conditions may therefore be limited. We advise that for further investigations of the effect of handovers in emergency situations, the nature of the situation should be kept constant, and learning effects should be counteracted by other measures like greater variation in aircraft arrival and departure times.

Taken together, this study provides hints towards the usefulness of handovers as a mitigation for demanding situations in MRTO. Even though handovers might induce additional MWL during the procedure, our data suggests beneficial long-term effects. The high approval from the participants is an additional benefit. Apart from objective positive effects, handovers have the potential to positively influence

ATCOs' confidence. The knowledge that they can get support and give away traffic if they feel the need to do so could reduce stress levels and improve the acceptance of the transition from conventional towers or RTO to MRTO.

Nevertheless, more empirical evidence is needed in order to assess effects of handover procedures on ATCOs' MWL and SA, especially during the handover itself. In addition to subjective and retrospect ratings, physiological measurement of MWL could shed light on the impact on cognitive resources during and after handover. Electroencephalography (EEG), for example, has proved useful for the classification of MWL (Causse et al., 2015; Radüntz, 2017; Radüntz et al., 2020), providing both high temporal resolution and an objective assessment. In addition to MWL and SA, future research should also investigate the effects of handovers on the development of mental fatigue (e.g. Fatigue Instantaneous Self-Assessment F-ISA; Hamann & Carstengerdes, 2020) and sleepiness (e.g. Karolinska Sleepiness Scale KSS; Akerstedt & Gillberg, 1990) during ATS shifts. We therefore encourage research on this topic with additional (physiological) measurements, larger samples, as well as more challenging traffic scenarios including a wider variety of unexpected situations and weather conditions. The interaction between ATCOs and our dyadic team approach pose another methodological challenge: A nested design. Interactions and learning curves might differ between ATCO dyads and could influence the efficacy of communication during handovers. A multi-level analysis approach could shed light on team dynamics and improve handover procedures further.

In the future, handovers could not only prove to be a useful strategy to reduce cognitive demands. This study focused on inflicting high levels of MWL with a high traffic volume and special situations in order to see if handing over traffic to a colleague could mitigate these effects. Having shown that aerodromes can be handed over safely and efficiently, we hypothesize that this strategy could be used to balance ATCOs' MWL in both directions: from excessive down and from underutilized up to an acceptable level. The ISA data we collected from the Support ATCOs showed a pattern that could be interpreted as a floor effect. The ATCOs were underutilized most of the time, especially with only one aerodrome in the "*Increased Traffic Load*" condition. In conventional towers the ATCOs can only work with the traffic at hand. In case of a small, low frequented aerodrome the task load may prove insufficient and ATCOs could be underutilized for long periods. This poses a safety risk since ATCOs are prone to lose vigilance and risk a slow but steady decline of their ATS skills when they do not train them in day-to-day operations (Weinger, 1999). MRTO offer the possibility to enlarge and enrich chronically underutilized ATCOs' work. The flexible addition of aerodromes and thereby traffic as well as responsibility could foster job satisfaction and counteract the detrimental effects of long periods of being underutilized. In ATC, task load and complexity do not remain the same; instead they oscillate during the day. MRTO provide an opportunity to flexibly allocate aerodromes to ATCOs and hand over aerodromes when needed. During peak times or highly complex situations, ATCOs could hand over an aerodrome when they feel the need to, while underutilized ATCOs could accept another aerodrome to increase their MWL to a comfortable level. Therefore, future research should focus on both, MWL reduction and increase depending on the situational needs and ATCOs' level

of comfort. If handovers can be used to regulate ATCOs' MWL in both directions, they will become a powerful strategy for ATCOs. Making handovers an inherent part of the concept would give MRTO the potential to let ATCOs regulate their MWL depending on the situation, whilst ensuring a safe, efficient and orderly flow of air traffic at the same time.

References

Adams, S. R., Kane, R., & Bates, R. (1998). *Validation of the China Lake Situational Awareness scale with 3D SART and S-CAT (452330D).* China Lake, CA. Naval Air Warfare Center Weapons Division.

Akerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *The International Journal of Neuroscience, 52*(1–2), 29–37. https://doi.org/10.3109/002074590 08994241

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology, 42*(3), 361–377. https:// doi.org/10.1016/0301-0511(95)05167-8

Causse, M., Fabre, E., Giraudet, L., Gonzalez, M., & Peysakhovich, V. (2015). EEG/ERP as a measure of mental workload in a simple piloting task. *Procedia Manufacturing, 3*, 5230–5236. https://doi.org/10.1016/j.promfg.2015.07.594

Cooper, G. E., & Harper, R. P., JR. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities (TN-D-5153).* Moffet Field, California, USA. National Aeronautics and Space Administration, Ames Research Center. https://ntrs.nasa.gov/citations/19690013177

Dehn, D. M. (2008). Assessing the impact of automation on the air traffic controller: The SHAPE questionnaires. *Air Traffic Control Quarterly, 16*(2), 127–146. https://doi.org/10.2514/atcq.16. 2.127

Endsley, M. R. (1999). Situation awareness in aviation systems. In D. J. Garland (Ed.), *Human factors in transportation. Handbook of aviation human factors.* Erlbaum.

Friedrich, M., Hamann, A., & Jakobi, J. (2020). An eye catcher in the ATC domain: Influence of multiple remote tower operations on distribution of eye movements. In D. Harris & W.-C. Li (Eds.), *Lecture Notes in Computer Science. Engineering Psychology and Cognitive Ergonomics. Cognition and Design* (17th Ed., Vol. 12187, pp. 262–277). Springer Nature. https://doi.org/10. 1007/978-3-030-49183-3_21

Fürstenau, N. (Ed.). (2016). *Research topics in aerospace. Virtual and remote control tower: Research, design, development and validation.* Springer International Publishing. http://gbv.eblib. com/patron/FullRecord.aspx?p=4528419. https://doi.org/10.1007/978-3-319-28719-5

Hamann, A., & Carstengerdes, N. (2020). *Fatigue Instantaneous Self-Assessment (F-ISA): Development of a short mental fatigue rating* (DLR-IB-FL-BS-2020–64). Braunschweig. Deutsches Zentrum für Luft- und Raumfahrt, Institut für Flugführung.

ten Have, J. M. (1993). The development of the NLR ATC Research Simulator (Narsim). *Simulation Practice and Theory, 1*(1), 31–39. https://doi.org/10.1016/0928-4869(93)90009-F

Jakobi, J., Hamann, A., Filipp, M., Kaufhold, R., & Priboj, O. (2019). Final project report: Project "PJ05 Remote Tower". Grant no.: 730195, Call H2020-SESAR-2015–2, Topic SESAR.IR-VLD.Wave1–08–2015. https://www.sesarju.eu/sites/default/files/documents/ projects/FPR/PJ05_D2.1_Final%20Project%20Report.pdf

Kahnemann, D. (1973). *Attention and effort*. Prentice-Hall.

Mensen, H. (2014). Moderne Flugsicherung: Organisation, Verfahren, Technik (4. Aufl.). VDI-Buch. Springer Vieweg. https://doi.org/10.1007/978-3-642-54294-7

Möhlenbrink, C., Papenfuss, A., & Jakobi, J. (2012). The role of workload for work organisation in a remote tower control center. *Air Traffic Control Quarterly, 20*(1), 5–26.

Radüntz, T. (2017). Dual frequency head maps: A new method for indexing mental workload continuously during execution of cognitive tasks. *Frontiers in Physiology, 8*, 1019. https://doi.org/10.3389/fphys.2017.01019

Radüntz, T., Fürstenau, N., Mühlhausen, T., & Meffert, B. (2020). Indexing mental workload during simulated air traffic control tasks by means of dual frequency head maps. *Frontiers in Physiology, 11*, 300. https://doi.org/10.3389/fphys.2020.00300

Roscoe, A. H. (1984). Assessing pilot workload in flight. In: *AGARD Conference Proceedings Flight Test Techniques*, Paris, France.

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale of assessing pilot workload in flight: A decade of practical use: A technical report* (TR 90019). Farnborough, Hampshire. Procurement Executive, Ministry of Defence.

Stokes, A., & Kite, K. (1997). *Flight stress: Stress, fatigue, and performance in aviation* (Repr.). Avebury.

Tattersall, A. J. (1994). Practical guidelines for workload assessment. In J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Aviation human factors series. Human factors certification of advanced aviation technologies: Proceedings of Human Factors Certification of Advanced Aviation Technologies Conference held at the Château de Bonas*, near Toulouse, France, July 19–23, 1993 (pp. 193–205). Embry-Riddle Aeronautical University Press.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics, 39*(5), 740–748. https://doi.org/10.1080/00140139608964495

Weinger, M. B. (1999). Vigilance, boredom, and sleepiness. *Journal of Clinical Monitoring and Computing, 15*(7–8), 549–552. https://doi.org/10.1023/a:1009993614060

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Which Minimum Visual Tracking Performance is Needed in a Remote Tower Optical System?

Jörn Jakobi and Kim Laura Meixner

Abstract To maintain or even increase the ATCO's situation awareness at a remote tower working position, augmentation features are introduced, such as automated tracking of objects. Moving objects (aircraft, vehicles, persons, etc.) the ATCO is interested in are tracked and augmented by this function. However, a tracking function is never reliable by 100% and nuisance tracking information occur, information, which is not of operational relevance and is disturbing to the ATCO. Not only human performance gains by "wanted" tracking information but also losses due to "nuisance" tracking information are to be expected. This paper investigates the effect of visual tracking function in a Remote Tower Optical System on ATCO's acceptance and effects on situation awareness and workload. Erroneous tracking performance will be discussed within the framework of a response matrix (e.g. Wickens, Elementary Signal Detection Theory, Oxford University Press (2002) and Appendix B). It collects the correct and false (system) responses as conditional probabilities for two alternative situations (object to be tracked/object not to be tracked), or for signal and noise. In a human-in-the-loop real time simulation, seven ATCOs performed a realistic traffic scenario. The study was conducted at Remote Tower laboratory at DLR in Braunschweig. As an experimental condition the performance of the visual tracking was operationalized by the number of nuisance tracking indication: (1) *none* (no visual tracking (baseline)), (2) *low*, (3) *medium*, and (4) *large* number. The results show that ATCOs very much appreciate visual tracking information. ATCOs can more easily detect critical traffic situation, which increases their situation awareness and safety. Further on, acceptance is rather high and workload on a moderate level, and both parameters behave rather robustly, even when the number of nuisance tracking information increases.

Keywords Remote tower · Tracking · Nuisance · Signal detection theory

J. Jakobi (✉)
German Aerospace Center, Institute of Flight Guidance, Lilienthalplatz 7, 38108 Braunschweig, Germany
e-mail: joern.jakobi@dlr.de

K. L. Meixner
Hochschule Fresenius, University of Applied Science, Im MediaPark 4d, 50670 Köln, Germany
e-mail: kim-laura-meixner@web.de

1   Introduction and Background

Remote Tower Optical Systems, which provide a video panorama of the conventional out-of-the-window Tower view, can offer an opportunity conventional out-of-the-window environments are hardly able to provide: ATC relevant information can be superimposed, or augmented, on the real scenery.[1] Information, related to weather or wind for instance, or even flight plan or track data are usually available head down only. With Remote Tower this visually presented information can now be spatially and timely linked to the video scenery. Hence, head-down times can be reduced and the ATCO can capture all relevant ATC information with one gaze. One of the most promising augmentation features is visual tracking of objects with high operational interest.

Such an automatic tracking function is an image processing-based function. It reads out the camera sensor information, identifies groups of moving pixels and augments them by e.g. a so called bounding or tracking box capturing the moving object. The system attempts to maintain the association with the moving pixels between successive video frames ("tracking"). The intention is not to use it as a surveillance tool but only to support operator situation awareness. (EUROCAE, 2021). Initial research on augmentation features in a remote tower environment had been performed at DLR since 2004 and tested in the experimental DLR Remote Tower System (see chapter "Remote Tower Experimental System..." (RTO Exp Syst. With Augm.Vision Videopan.): Sect. 3.4, chapter "Remote Tower Prototype System..." "Integration of (Surveillance)Multilateration Sensor Data into a Remote Tower System" (RTO Prototype & Automation Perspect.: Sect. 4: detection, classification, tracking), and chapter "Assessing Operational Validity..." "Remotely-Operated AFIS in Japan" (Operational Validity of RTO in HITL Simul: Sects. 3.3, 4.3: feasibility, usability of augmented vision and tracking tools). Technical details and initial experimental results including HITL-simulations are published in Fürstenau et al. (2007) (Virtual Tower Patent), Schmidt et al. (2007), Möhlenbrink et al. (2011).

If the aerodrome is equipped with additional non-optical surveillance sensors, like approach or surface radar, the quality of the tracking function can be further improved by correlation with surveillance information from those additional sensors. Particularly in far distance when neither ATCOs nor camera sensors are able to detect and track traffic, the radar captures the traffic and track information can be indicated on the video panorama. When the aircraft comes closer also the camera sensors would capture the aircraft and camera and radar track information are to be correlated to provide associated track information from different sensors. Radar and camera sensors can complement each other, because radar usually has a higher coverage

---

[1] This chapter is a revised version of the conference paper *Jakobi, Jörn and Meixner, Kim Laura (2018). Effects of Unwanted Tracking Boxes in a Remote Tower Control Environment. International Journal For Traffic And Transport Engineering (ICTTE Belgrade 2018), 27–28. Sep 2018, Belgrade. ISBN 978-86-916,153-4-5.* Revisions include additional results and an update of the official EUROCAE ED-240A tracking terminology that progressed with its change 1 release from 2021.

range (up to 25 miles) than cameras. Cameras instead usually provide better spatial accuracy and a higher update rate. That is, even when provided with a low update rate and low position accuracy, ATCOs see traffic right in advance in far distance on the panorama display. When the ATCO becomes in charge of traffic in closer vicinity, cameras provide more accurate position information. Another advantage of a tracking function is labelling of objects. As soon as the position and identification of an aircraft or vehicle is known to the system, label information stemming from cooperative radar or flight plan information can be linked to the object. Call sign information, aircraft type, altitude, speed or destination information could be provided by a label attached to the object. But such additional label information is only feasible when additional cooperative radar information is available on an airport. Mostly, small, remotely operated airports are solely equipped with optical surveillance sensors and tracking bases on optical information only.

In the past, such automated tracking functions were investigated in several national and European research projects with concordant results: Tracking information is very much appreciated by the ATCOs. In the project "Advanced Remote Tower" (ART) of the 6th European Framework Program (Van Schaik et al., 2016a and 2016b) and with the SESAR project P06.09.03 (SESAR, 2014a) ATCOs in an active shadow mode setting were presented tracking information captured from an aerodrome 100 km away from their ATCO working position under test. Within the ART project, ATCOs complained about some quality issues, e.g., jumping tracking information, but the majority of the ATCOs appreciated tracking information. In the SESAR project 06.09.03, ATCOs' opinions resulted in an above-average acceptance score of 4.03 on a Likert scale from 1 "strongly disagree" to 5 "strongly agree". Also, subjective "trust" and "situation awareness" (SA) measurements scored above-average. Regarding safety, the ATCOs admitted that "Remote Tower environment can have features that a local tower does not have, including the addition of technical enablers such as tracking overlays". In SESAR project 06.08.04 (SESAR, 2014b), tracking and labelling information was investigated in a multiple Remote Tower setting. In a real-time simulation, ATCOs worked a very heavy traffic scenario with 30 movements per hour via a single airport (baseline condition) and the same traffic scenario spread over two airports on a multiple Remote Tower CWP (treatment factor) via two factor levels, with and without automated tracking function. In accordance to SESAR (2014a), the tracking was appreciated as 'nice to have', but was, contrary to SESAR (2014a) results, hardly been used. The authors explained this effect by the high traffic load that led ATCOs to work head down predominately to get ATC relevant information and to operate the electronic flight strips. Papenfuss and Möhlenbrink (2016) tested 12 ATCOs in a simulated Remote Tower environment with automated tracking and assessed their eye-point of regards. By the means of tracking, head-down times could be significantly reduced. ATCOs also regarded the tracking functionality as helpful, particularly in critical situations.

Even when past results could show that augmented tracking information are much appreciated by the ATCOs to support their situation awareness (SA), there is also a risk that ATCOs get distracted by tracked objects that are not of interest to the ATCO or that the tracking function misses to track an object that is of interest to the

Table 1 Visual tracking classification scheme—four tracking result pattern and their anticipated negative effects with respect to the signal detection theory

	Object tracked	Object not tracked
Object-to-be-tracked	HIT (H) (wanted)	MISSED (M) (unwanted)
Object-*not*-to-be-tracked	FALSE ALARM (FA) or NUISANCE (unwanted)	CORRECT REJECTION (CR) (wanted)

ATCO. Both effects decrease the ATCO's SA and can cause additional workload and even safety risks. Green and Swets (1988) describes those subjective discrimination effects related to a radar environment by applying signal detection theory (SDT) (see also Appendix B). Core of SDT is a "signal", which is to be detected, and can be present or absent when it is detected. If the signal is really present, SDT speaks of a "hit" (H); if it is absent, the SDT names it "false alarm" (FA). Present signals that are not detected are "misses" (M) and if a signal is not present and also not detected SDT speaks of "correct rejection" (CR). The SDT result pattern can be applied for our automated tracking function (see Table 1).

To minimise unwanted *nuisance* and *missed* results (lower left and upper right cells in Table 1) and maximise wanted SDT result pattern (upper left and lower right cells in Table 1) the tracking function must be tuned in an optimal way. It must be sensible enough to keep the "missed" on a low level but also sensitive enough to keep the "nuisance" on an operationally reasonable level (see also Friedman-Berg, 2008). In a nutshell, the minimum performance of the tracking function must be high enough to gain an improved ATCO SA. Past research could show that the tracking function can improve SA in a Remote Tower context but the quality of the tracking function has never been varied in an experimental setting to gain knowledge about the minimum operational performance. For the future, this knowledge would help to quantify the minimum needed performance of a visual tracking function in order to support the standardisation process and to reduce the implementation time. The research question of this study therefore reads: *Which minimum visual tracking performance is needed in a remote tower optical system to provide positive effects on SA?*

2 Experimental Design

2.1 Participants

Seven male ATCOs between 29 and 62 years ($M = 42.1$, $SD = 12.8$) took part in the experiment. Except of one they all held an active ATC license and work or worked for various air navigation service providers (Austrocontrol, LFV, AVINOR, NATS and ROMATSA). ATCOs were directly invited by an invitation letter and they supported the study voluntarily and without any monetary compensation. All ATCOs

were familiar with the Remote Tower concept. The written and spoken language was English.

2.2 Experimental Platform

The experimental study was carried out at the Apron Tower Simulator at the Remote Tower laboratory at DLR Braunschweig between 15th and 31st of May 2017. A 45 min lasting traffic scenario at the Braunschweig-Wolfsburg airport (EDVE) was prepared to be used in a real-time simulation. EDVE is a controlled regional airport with class D control zone (CTR), 2200 feet altitude. The simulated scenario reflected typical day-to-day traffic with a total of 17 movements (15 VFR and two IFR movements), amongst them two aircraft performing right-hand traffic patterns, four arrivals, seven departures and four CTR crossers. Furthermore, two abnormal traffic events were induced, a maintenance car, which crossed the runway during a simultaneous departure event without waiting for an ATC clearance. One CTR crosser was flying through the CTR from north to south without receive/transmit radio communication (R/T) and without a clearance. Visibility was CAVOK without any clouds. Wind was calm, 260°, five knots and runway 26 was in use. All communication was conducted via one radio channel and also VFR communication was conducted in English, because of the international ATCO test sample. Each ATCO controlled the traffic alone, no coordination with apron control or approach control was needed. In- and outbound traffic called directly before entering CTR (VFR), or IFR, when established on final 10 miles out, respectively, outbound traffic requested start-up or taxi. The traffic was operated by two blip pilots, one responsible for arrivals and crossers, one for departures and traffic patterns. The ATCOs were provided with 200° northwards panorama view and a 360° pan-tilt zoom camera (PTZ). The CWP should simulate a very basic Remote Tower equipage, therefore they were neither provided with 360° panorama nor with approach or surface radar. Head down, they were provided with a tower flight data processing system to operate the flight strips, a space mouse to operate the PTZ, a communication display and an information data processing system (IDVS), providing visibility, wind and QNH information. The automated tracking function was realised with cyan-coloured tracking boxes, some of them "wanted", others "unwanted" or also called "nuisance" boxes (see also Fig. 1).

2.3 Test Variables and Test Procedures

The reliability of automated tracking function is related to the relation of "wanted" (correct hits) to "missed" boxes and the number of "nuisance" boxes. To keep the experimental design efficient and manageable, the relation of "wanted" to "missed" boxes was kept stable. Instead, the number of "nuisance" boxes was varied over the

Fig. 1 Experimental set up of the Remote Tower Module with Tracking Boxes in the Panorama Visual Presentation *remark: two tracking boxes on the taxiway are "wanted" [green arrows], the one in the back is an unwanted "nuisance" tracking box [red arrow]; the figure below shows a magnification of the scenery above*

experimental conditions. Because a perfect 100% "hits" system is very unlikely in an operational environment, the "wanted" hits were fixed on a reasonable value: 15 out of 17 movements were permanently tracked while two out of 17 movements were permanently untracked, which resulted in an 88% hit rate. This hit rate perfectly corresponds to a recommendation of ATCOs out of the international EUROCAE working group WG-100 (EUROCAE, 2021) who proposed a minimum "hit" performance of 85%. What is "wanted", or in other words, which object-is-to-be-tracked is very dependent on the local operational needs, but for this experiment, it was assumed that all aircraft and vehicles on the movement area and in the vicinity of

Table 2 Independent Variable IV-A and its Variation over the number of "Nuisance" Boxes in the 45 min lasting traffic scenario

IV-A number of nuisance boxes → Area-of-interest → Dwell Time ↓	EC₀ (Baseline)				EC₁				EC₂				EC₃			
	RWY	TWY	Final	TP	RWY	TWY	Final	TP	RWY	TWY	Final	TP	RWY	TWY	Final	TP
30 sec	0	0	0	0	0	0	0	0	0	7	0	0	7	14	7	7
5 sec	0	0	0	0	0	7	7	7	7	14	7	7	14	28	14	14
2 sec	0	0	0	0	14	14	14	70	28	28	28	140	42	42	42	210
	0				133				266				441			

the airport are objects-to-be-tracked. The number of "nuisance" tracking boxes (also called false alerts (FA)) was chosen as the independent variable IV-A and was varied over four experimental conditions EC_0, EC_1, EC_2 and EC_3, which of EC_0 served as a baseline condition in which the automated tracking function was switched off. The occurrence of the "nuisance" boxes was varied over four operational area-of-interests: runway (RWY), taxiway (TWY), final approach area (Final) and traffic pattern (TP) and over their dwell time, the time the "nuisance" tracking boxes coasted until they disappeared on their own again (2, 5, 30 ₛ). Their number was varied over the three experimental conditions: EC_1 should represent a very low and less annoyable number of "nuisance" boxes, EC_2 should represent a medium and EC_3 a very large and annoyable number of "nuisance" boxes. The right choice of the number of "nuisance" boxes was very important to avoid bottom and ceiling effects in order to be able to judge about the minimum acceptable performance of "nuisance" boxes in the end. In a pre-trial with an extra ATCO, who did not participate in the exercise, an appropriate number of "nuisance" boxes was explored to meet this prerequisite. Table 2 represents the final setting of the number of "nuisance" tracking boxes.

All experimental conditions (EC_0, EC_1, EC_2 and EC_3) were run with the same traffic scenario to allow a final deduction on the IV-A "number of nuisance boxes" with respect to measured effects on the dependent variables: acceptance, SA and workload. To avoid ATCOs' recognition and training effects caused by a fourfold repetition of the traffic scenario, the aircraft's call signs were varied over the four different experimental conditions. The timing of the intruders (runway crossing vehicle and CTR crosser) and the two "missed" boxed movements were alternated as well. The sequence of the experimental conditions was randomized over the seven ATCOs to control for learning or fatigue effects.

The study was structured in three parts. The briefing and training phase represented the part in which ATCOs provided demographical data, were informed about the data protection procedure and prepared for the actual experiment. Secondly, the experimental phase corresponded to the conduction of four successive test runs and the third part dealt with completion of the post-run questionnaire and a final debriefing phase. The ATCOs were kept unaware of the actual experimental condi-

tion. During the experimental phase, the ATCOs' workload, SA and acceptance were assessed. Workload was subjectively measured by ISA, a 5-point Likert scale, which popped up head down every 5 min requesting the ATCO to assess his current workload by pressing a button reaching from 1 "under-utilised" until 5 "excessive" (SESAR, 2012). At the same time, the ATCO was requested to subjectively judge about his acceptance of the tracking function by the statement: "During the past five minutes I found the tracking boxes: 1 = "Disturbing", 2 = "Not of interest", and 3 = "Helpful", whereas this mid-run assessment was not used during the EC_0 baseline experimental condition, since tracking was not available. Furthermore, SA and workload were assessed every nine minutes by the objective SARA-T online questionnaire (Kraemer & Süß, 2015). During the exercise at regularl intervals, operational questions are popping up that relate to the current traffic situation, e.g. "Is KLM4123 on your frequency", which have to be answered by the participant. Reaction time and correctness are analysed to conclude on the instant workload and situation awareness. Regarding the abnormal events, the experimenter noted if the intruders were recognised by the ATCO and how they dealt with them. During the debriefing phase, the ATCOs were asked to fill in standard questionnaires with respect to SA (SASHA), workload (AIM) and Acceptance (SATI), a test battery developed by EUROCONTROL in its project "Solutions for Human-Automation Partnerships in European ATM" (SHAPE) (Dehn, 2008). Further on, the ATCOs provided answers to several closed statements regarding SA, workload and acceptance and were given the chance to provide additional comments and remarks which they eventually had in mind after a full day of testing. After performing all test runs, the ATCOs were asked 11 further general statements (closed questions) to the automatic tracking function to be answered on a 6-point Likert scale. Table 3 presents a complete overview of the dependent variables and the applied test battery.

**Table 3** Dependent variables, their respective measurement tools and the time of assessment

Dependent variable	Measurement tool	Timing
Workload	I.S.A	Mid-run
	SARA-T	
	AIM-s	Post-run
Situation awareness	SARA-T	Mid-run
	SASHA	Post-run
Acceptance	3-likert acceptance scale	Mid-run
	ad-hoc comments	
	SATI	Post-run
	Debriefing closed questions	Post-trial
	Debriefing open questions	
	Ad-hoc comments	

3 Results

In order to get reasonable result with an acceptable external validity, a sample size of experts was aimed for. A potential drawback with expert sample sizes is the rare availability of them. By consequent, the sample size often lacks a reasonable size to apply parametric inference statistics with an acceptable internal validity. The analysis therefore focussed on the descriptive analysis of the results and when appropriate, non-parametric inference statistics like the Friedman or binomial test were applied.

Regarding mid-run workload, neither ISA, SARA-T nor the post run AIM provided significant differences by a Friedman test and also the mean values do not draw an easily interpretable pattern. In general, workload was rather below-average and rather unaffected of the treatment. A similar result was obtained with respect to SA measurements. Neither the mid-run SARA-T nor the post-run SASHA provided a significant result pattern. With SARA-T, following mean values (M) and standard deviation (SD) were measured $EC_{0} = 4.39/1.89$; $EC_1 = 3.47/0.76$; $EC_2 = 4.51/2.52$ and $EC_3 = 4.65/2.96$. SASHA measured: $EC_0 = 4.48/0.98$; $EC_1 = 4.74/0.78$; $EC_2 = 4.83/0.77$ and $EC_3 = 4.93/0.71$.

According the ATCOs' acceptance, the mid-run assessment via a three-point Likert acceptance scale with 1 = "Disturbing", 2 = "Not of interest", and 3 = "Helpful", revealed an overall above-average acceptance but without any significant deviation between the experimental conditions: $EC_1 = 2.84/0.2$; $EC_2 = 2.86/0.2$ and $EC_3 = 2.8/0.24$ ($\chi^2(2) = 0.353$, p = 0.838). In each test run each of the seven participants answered this scale 8 times, in total 56 answers per EC. In every EC, independent of the treatment "number of nuisance tracking boxes", more than 80% of the answers referred to „Helpful ". "Not of Interest" was almost equally chosen in between of $EC_1 = 17.9\%$; $EC_2 = 10.7\%$ and $EC_3 = 16.1\%$. "Disturbing" instead was just ticked in the "medium" and "large" conditions $EC_2 = 1.8\%$ and $EC_3 = 3.6\%$. Figure 2 presents a graphical overview of these result patterns.
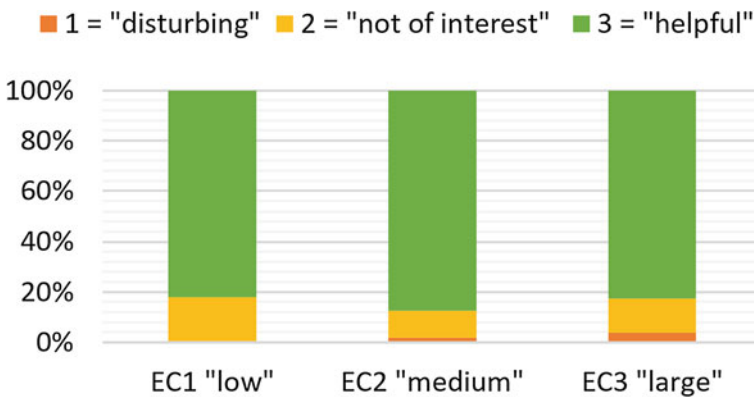
Fig. 2 Percentage values from the three-point likert acceptance scale w.r.t. low, medium and large number of nuisance tracking boxes

Same result pattern with the SATI post run questionnaire: $EC_1 = 4.29/1.27$; $EC_2 = 4.14/0.77$ and $EC_3 = 3.60/1.07$. Actually, descriptivey acceptance decreases with larger number of "nuisance" boxes but not significantly.

Post-run the ATCOs were also asked: "The experienced wanted, missed and nuisance Boxes had an acceptable rate to help me increasing my situation awareness." and "I experienced nuisance boxes but they popped up in an acceptable number that they did not prevent me from working in a safe and efficient manner" to be answered on a 6 point Likert scale from 1 "strongly disagree" to 6 "strongly agree". Both statements were answered over-average positive and after having conducted a binominal test EC_1 and EC_2 with the first question, and EC_1 with the second question, became significant ($p = 0.016$) (see Figs. 3 and 4).

With respect to the two abnormal events: The CTR crosser was not recognised at all it in the EC_0 baseline condition but four of seven ATCOs spotted it in the experimental tracking conditions EC_1, EC_2 and EC_3 (Friedman-Test: $\chi^2 (3) = 12.00$, $p = 0.007$), when it was tracked. Similar results pattern were gained with the vehicle crossing the runway without clearance, which in the EC_0 baseline condition was seen by four out of seven ATCOs but in the tracking conditions EC_1, EC_2 and EC_3, when it was tracked, was always seen by all ATCOs.

After performing all test runs, the ATCOs were asked 11 further general statements to the automatic tracking function to be answered on a 6-point Likert scale from 1 "strongly disagree" to 6 "strongly agree". Figure 5a and b show the mean values and when marked with an asterisk became significant by a binominal test with $p = 0.016$. Nine of the 11 statements became significant proving that the 7 ATCOs agreed with the statements.

Fig. 3 Post-run Question 1 regarding acceptance "The experienced wanted, missed and nuisance Boxes had an acceptable rate to help me increasing my situation awareness." (Rem.: An asterisk symbol marks a significant over-average value tested by a binominal test with $p = 0.016$.)
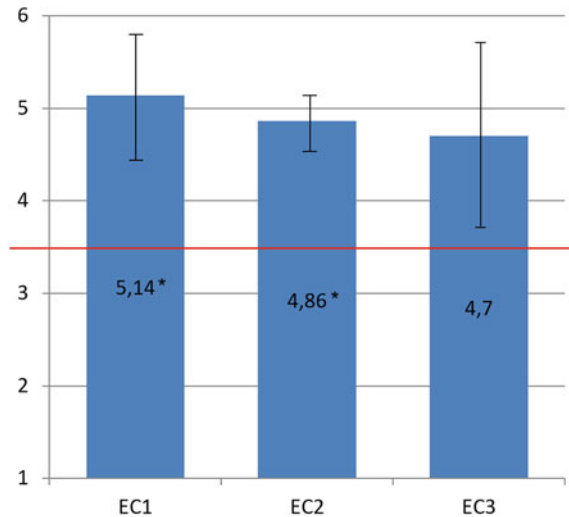
Fig. 4 Post-run Question 2 regarding acceptance "I experienced nuisance boxes but they popped up in an acceptable number that they did not prevent me from working in a safe and efficient manner" (Rem.: An asterisk symbol marks a significant over-average value tested by a binominal test with p = 0.016.)
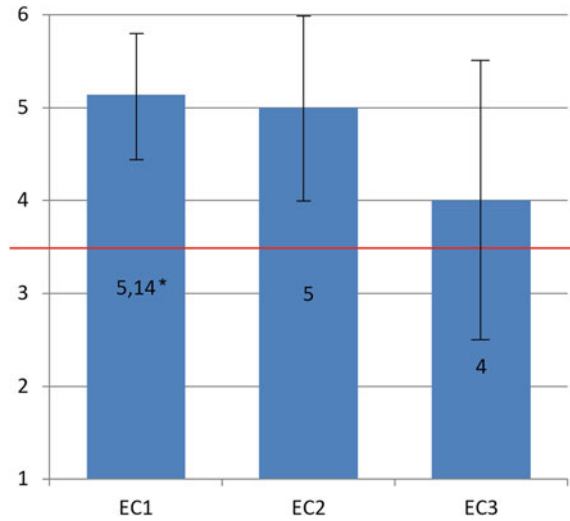
Fig. 5 a Post-trial Questions (1–5 of 11) regarding general Statements to the Automatic Tracking Function (Rem.: An asterisk symbol marks a significant over-average value tested by a binominal test with p = 0.016.; including SD error bars). **b** Post-trial Questions (6–11 of 11) regarding general Statements to the Automatic Tracking Function (Rem.: An asterisk symbol marks a significant over-average value tested by a binominal test with p = 0.016.; including SD error bars)

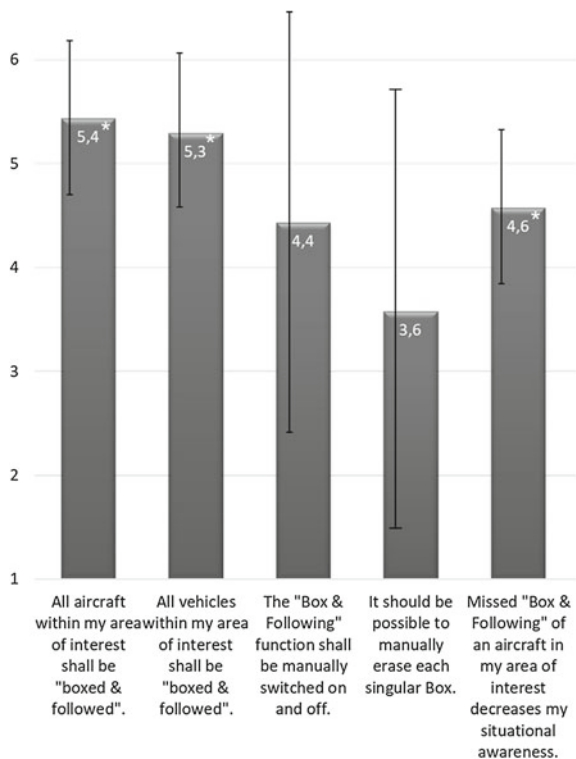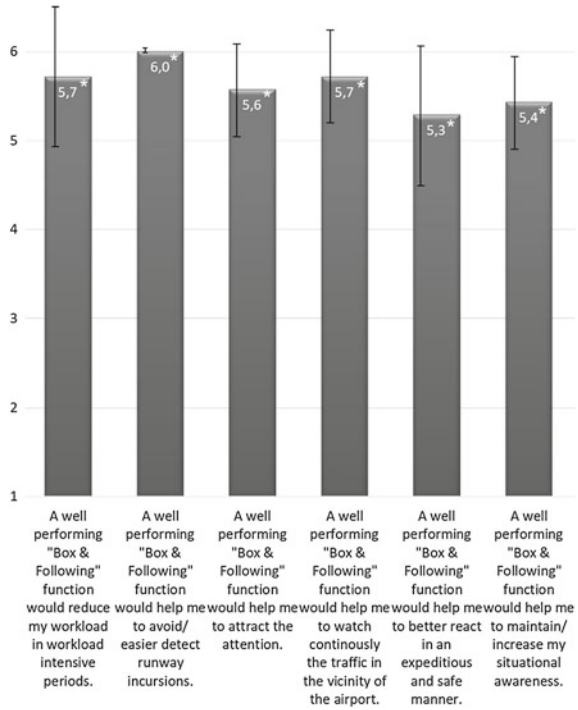## 4 Discussion

The experimental study aimed to get to the bottom of the minimum quality threshold of an automatic tracking function in a Remote Tower context. It fixed the relation of "hits" by 88% and "missed" by 12% respectively, and varied the "number of nuisance tracking boxes" (FA) via three experimental conditions EC_1 = "low", EC_2 = "medium" and EC_3 = "large". A fourth EC_0 served as baseline condition to better compare SA or Workload gains or losses. Even when the ATCOs recognised performance differences in the number of "nuisance" tracking boxes and significantly preferred the lower number of "nuisance" boxes (see Fig. 4), the mid- and post-run standardised measurement tools were not sensitive enough to reveal any significant differences with SA and workload. As a result, it was omitted to perform a quantitative SDT analysis with determination of discriminability d' and decision criterion c. Instead, a dominant ceiling effect was observed: ATCOs felt permanently confident regarding SA and workload rather independent of the treatment "number of nuisance tracking boxes". They appreciated the "wanted" boxes (correct hit (H)) and behaved more or less unaffected from the disturbing "nuisance" boxes (FA).

This result pattern can be explained by three interpretations: Firstly, the traffic scenario was challenging but manageable for the ATCOs and in their subjective perception they must have had a proper SA and workload in order to manage such

a challenging scenario. This phenomenon is quite well known with ATCO expert samples (Badke-Schaub et al., 2012). That is why the experimental design also focussed on the objective SA measurements, like the SARA-T metrics, but unfortunately with the same ceiling effect. Most probably, the SARA-T operational mid-run questions were too easy or not affected by the number of "nuisance" tracking boxes. Secondly, "What the eye does not see, the heart does not grieve over" is a very good proverb to describe the situation when an aircraft crossed the CTR without a clearance. Although always present by some few pixels in the panorama view, in the $EC_{0\text{ "baseline"}}$ condition (no tracking) nobody of the seven ATCOs spotted the crosser. The crosser kept unnoticed by them and they did not interfere with their current tasks, that is, neither workload nor SA was affected, because they did not know that they missed something. By contrast, in the experimental tracking conditions EC_1, EC_2 and EC_3, they detected the crosser through the "wanted" tracking box, which surprised them and they were thinking that they have missed an initial call or a clearance. This situation caused additional taskload by trying to contact the crosser via R/T without any pilot's response and to provide traffic information to the traffic in the vicinity to warn of the intruder. All those actions helped to avoid potential safety critical risks but negatively affected ATCOs' subjectively perceived SA and workload. Third interpretation attempt, why SA and workload remained rather unaffected by the number of "nuisance" boxes, refers to the ATCOs' comments that they could easily distinguish "wanted" from "nuisance" boxes. Particularly large boxes without any contents, stationary boxes or boxes appearing in non-critical areas like on buildings, on the greens or apron could easily been ignored by the ATCOs or even remained unnoticed by them. Critical are "nuisance" boxes with a similar appearance like the "wanted" boxes, e.g., a moving box appearing in the final approach area with size and vector speed similar to a real target. Unfortunately those more complex box behaviors could not be created for this study since it would have exceeded available resources.

Additional comments from ATCOs presume that they used the tracking function like a non-cooperative surface movement radar (SMR). An SMR provides the ATCO with yellow-coloured pixel swarms on a 2D birds view airport surveillance detection equipment (ASDE) display. The radar cannot distinguish between "wanted" or "nuisance" pixel swarms, it simply shows everything that provides a shadow through the radar beam. The ATCOs do not use such an ASDE as a control tool but as an additional means to verify aircraft or vehicle positions seen through the out-of-the-window view or provided by pilot reports. Over the time, using an ASDE ATCOs are becoming very efficient in distinguishing between operational relevant aircraft/vehicle shadows or "nuisance" shadows induced by light masts, buildings, grass, snow banks, Tarmac reflections or flocks of birds. Considering that a visual tracking function is not used as a control tool and when gaining experiences with the tracking function, the ATCO would probably learn to ignore obvious "nuisance" tracking boxes. One could even think about increasing the sensibility of the tracking function to get "hit" rates of up to 95% or 99% and reducing "missed" rates by a simultaneous acceptance of an increased number of "nuisance" tracking boxes. This consideration is also supported

by item 5 in Fig. 5a by which the ATCOs uniformly stated that "missed" tracking boxes may decrease their SA.

Other results refer to the improved recognition of intruders like CTR or runway incursions. With the tracking function intruders are easier to detect and thus the tracking function significantly contributes to safety (see also item 7, 8 and 9 in Fig. 5b). Referring to the statement whether the entire tracking function or single boxes should be able to switch off manually, ambivalent results were gained. Instead, an agreement was gained referring the statement "what is to be tracked": ATCOs want to have tracked all aircraft and vehicles in their area of interest but not on the apron area when the objects are visible anyway (Fig. 5a).

5   Conclusions and Outlook

In accordance to existing studies (see Sect. 1) this experimental study provided proof that an automated tracking function in a Remote Tower environment is very much appreciated by the ATCOs and that they significantly admit its positive contribution to situation awareness and safety. Further on, the tracking function seem to be rather robust against "nuisance" tracking boxes (FA) and probably more focus could be laid on the increase of the sensibility in terms of getting larger percentage rates (H) with the "wanted" tracking boxes. Furthermore, defining a quantitatively expressed minimum performance value for "nuisance" tracking boxes seems to be very difficult or even impossible. This study varied "nuisance" tracking boxes in dwell time and location but there are more attributes, such as size of the tracking box, or if they are stationary or moving. Most probably, each implementation of a tracking function must be repeatedly tuned in line to the local operational needs and must gain acceptance by the local ATCOs. Future studies should focus on a bigger sample size and more realistic "nuisance" tracking boxes to provide further empirical evidence to these conclusions.

References

Badke-Schaub, P., Hofinger, G., & Lauche, K. (2012). *Psychologie des sicheren Handelns in Risikobranchen* (pp. 3–20). Springer.

Dehn, D. M. (2008). Assessing the impact of automation on the air traffic controller: The SHAPE questionnaires. *Air Traffic Control Quarterly, 16*, 127–146.

EUROCAE. (2021). *Minimum Aviation System Performance Specification for Remote Tower Optical Systems*. EUROCAE Working Group (WG) 100, Version ED-240A change 1, (in press).

Friedman-Berg, F., Allendoerfer, K., & Pai, S. (2008). Nuisance alerts in operational ATC environments: Classification and frequencies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*, 104–108.

Green, D. M. & Swets, J. A. (1988). *Signal Detection Theory and Psychophysics*, Peninsula publishing: Reprint edition, Los Altos Hills, CA

Fürstenau, N., Rudolph, M., Schmidt, M., Werther, B., Hetzheim, H., Halle, W., & Tuchscheerer, W. (2007). Flugverkehr-Leiteinrichtung (Virtual Tower). European Patent EP1791364A1. Priority 23.11.2005. granted 30.05.2007

Möhlenbrink, C., Papenfuss, A., & Jakobi, J. (2011). The role of workload for work organization in a remote tower control center. *Air Traffic Control Quarterly, 20*(1), 5–26. https://doi.org/10.2514/atcq.20.1.5

Kraemer, J. & Süß H. -M. (2015). SARA-T: Validierung eines Echtzeit Mess- und Validierungstools für das Situationsbewusstsein. Available from Internet: http://elib.dlr.de/98839/2/DLRK_2015_Kraemer_Süß_20150914.pdf.

Papenfuss, A. & Möhlenbrink, C. (2016). Assessing Operational Validity of Remote Tower Control in High-Fidelity Simulation. In N. Fürstenau (Ed.), *Virtual and Remote Control Tower. Research, Design, Development and Validation* (S. 87–113). Schweiz: Springer International Publishing.

Schmidt, M., Rudolph, M., Werther, B., Möhlenbrink, C., & Fürstenau, N. (2007). Development of an augmented vision video-panorama human-machine interface for remote airport tower operation. In M. J. Smith & G. Salvendy (Eds.), *Lecture Notes Computer Science (LNCS): Human Interface II* (Vol. 4558, pp. 1119–1128). Springer.

SESAR. (2012). Instantaneous self assessment of workload. Available from Internet. Retrieved March 12, 2017, from https://ext.eurocontrol.int/ehp/?q=node/1585.

SESAR JU. (2014a). Remotely Provided Air Traffic Service for Single Aerodrome VALR (Remark: VALR=Validation report), SESAR JU Deliverable D08-02, Edition 00.05.02, 2014-05-01

SESAR JU. (2014b). VALR Multiple Remote TWR V2 (Remark: VALR=Validation report), SESAR JU Deliverable D97, Edition 00.01.00, 2014-06-11

Van Schaik, F. J., Roessingh, J. J. M., Lindqvist, G. & Fält, K. (2016a). Detection and Recognition for Remote Tower Operations. In N. Fürstenau (Ed.), *Virtual and Remote Control Tower. Research, Design, Development and Validation* (S. 53–65). Schweiz: Springer International Publishing.

Van Schaik, F. J., Roessingh, J. J. M., Bengtsson, J., Lindqvist, G. & Fält, K. (2016b). The Advanced Remote Tower System and Its Validation. In N. Fürstenau (Ed.), *Virtual and Remote Control Tower. Research, Design, Development and Validation* (S. 263–278). Schweiz: Springer International Publishing.

Wickens, T. D. (2002). *Elementary Signal Detection Theory*. Oxford University Press.

Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration During Simulated Aircraft Landing

Norbert Fürstenau and Stephen R. Ellis

Abstract   In order to determine the required visual frame rate (FR) for minimizing prediction errors with out-the-window video displays at remote/virtual airport towers, thirteen active air traffic controllers viewed high dynamic fidelity simulations of landing aircraft and decided whether aircraft would stop as if to be able to make a turnoff or whether a runway excursion would be expected. The viewing conditions and simulation dynamics replicated visual rates and environments of transport aircraft landing at small commercial airports. The required frame rate was estimated using Bayes inference on prediction errors by linear FR-extrapolation of event probabilities conditional on predictions (stop, no-stop). Furthermore estimates were obtained from exponential model fits to the parametric and non-parametric perceptual discriminabilities d' and A (average area under ROC-curves) as dependent on FR. Decision errors are biased towards preference of overshoot and appear due to illusionary increase in speed at low frames rates. Both Bayes and A—extrapolations yield a framerate requirement of $35 < FR_{min} < 40$ Hz. When comparing with published results (Claypool and Claypool Multimedia Systems 13:3–17, 2007) on shooter game scores the model based d'(FR)-extrapolation exhibits the best agreement and indicates even higher $FR_{min} > 40$ Hz for minimizing decision errors. Definitive recommendations require further experiments with FR > 30 Hz.

keywords Remote tower · Videopanorama · Framerate · Visual discrimination · Speed perception · Decision experiment · Aircraft landing · Signal detection theory · Bayes inference

N. Fürstenau (✉)
German Aerospace Center (DLR), Institute of Flight Guidance, Braunschweig, Germany
e-mail: norbert.fuerstenau@dlr.de; nfuerstenau@t-online.de

S. R. Ellis
Ames Research Center, NASA, (ret), Moffett Field, Mountain View, CA, USA

1 Introduction

This chapter reviews a two-alternative decision experiment with simulated aircraft landing as dependent on video-framerate (FR) characteristics with the goal of determining the minimum framerate necessary for minimizing decision errors under Remote Tower working conditions. It collects results partially presented in previous publications (Ellis et al., 2011a, 2011b; Fürstenau et al., 2012).

Recent proposals for decreasing cost of air-traffic control at small low-traffic airports have suggested that technology may remove the need for local control towers. Controllers could visually supervise aircraft from remote locations by videolinks, allowing them to monitor many airports from a central point (Hannon et al., 2008; van Scheijk et al., 2010; Schmidt et al., 2007; SESAR-JU Project, 2003). While many current towers on A-SMGCS-equipped airports, even some at busy airports like London-Heathrow, can continue to operate totally without controllers ever seeing controlled aircraft under contingency conditions, although with reduced capacity, it is clear from controller interviews that usually numerous out-the-window visual features are used for control purposes (Ellis & Liston, 2010, 2011; Van Schaik et al., 2010). In fact, these visual features go beyond those required for aircraft detection, recognition, and identification (Watson et al., 2009).

Potentially important additional visual features identified by controllers in interviews involve subtle aircraft motion. These could be degraded by low dynamic quality of remote visual displays of the airport environment. In fact, the dynamic visual requirements for many aerospace and armed forces tasks have been studied, but most attention has been paid to pilot vision (e.g. Grunwald & Kohn, 1994) and military tactical information transmission (e.g. Kempster, 2000). Relatively little attention was paid to the unique aspects of controller vision which, for example, emphasize relative motion cues. Consequently, there is a need to study some of these visual motion cues to understand how their use may be affected by degraded dynamic fidelity, e.g. low visual frame rates. Such low rates could be due to typically low rates of aircraft surveillance systems, e.g. 1–4 Hz, or to image processing loads arising from the very high resolution, wide field of view video systems needed to support human vision in virtual towers (see chapters "Remote Tower Experimental System with Augmented Vision Videopanorama", "Remote Tower Prototype System and Automation Perspectives").

Since preliminary investigation of the role of visual features in tower operations has shown that their principal function is to support anticipated separation by allowing controllers to predict future aircraft positions (Ellis & Liston, 2010) we have begun to investigate the effects of frame rates on the deceleration cues used to anticipate whether a landing aircraft will be able to brake on a runway, as if to make a turn off before the runway end.

Our specific hypothesis is that the disturbance due to low frame rate affects the immediate visual memory of image motion within the video frame. Memory processes classically have an exponential decay. Accordingly, one might expect discriminability of the visual motion associated with aircraft deceleration to reflect

this feature, degrading only a bit for higher frame rates but more rapidly for the longer period, lower frame rate conditions. A possible descriptive function could be of the form: $1 - \exp(-k/T)$. This kind of model captures the likely features that the rate of degradation of motion information increases with greater sample and hold delays T but that there is also an upper asymptote of discriminability corresponding to continuous viewing which is determined by the inherent task difficulty. Significantly, fitting such a model to the drop off in detection performance provides a theoretically based method to estimate that frame rate required to match visual performance out the tower window.

We used two statistical analysis methods for deriving model based framerate requirement estimates via discriminability measurement: Bayes inference and signal detection theory (SDT) with parametric (ROC-isosensitivity-curve index d') as well as non-parametric discriminability (A = average area under all proper ROC-curves). Bayes inference allows for concluding from the measured error probability conditional on the perceived world state, on the probability of this (unexpected) situation conditional on the measurement (see Appendix A2). Measuring these probabilities with different values of the independent variable (i.e. the framerate FR) allows for extrapolation to minimum FR for zero error probability. SDT as an alternative method has the advantage of separating the intrinsic subjective preference (tendency for more liberal or conservative (error avoidance) decisions) by simultaneously separating through the measurement of hit and false alarm rates (= probabilities conditional on the alternative experimental situations) from the decision criterion (or subjective decision bias) index c (for d') and b (for A) respectively.

Experimental Methods and results are provided in Sects. 2, 3. In Sect. 4 the two alternative methods (Bayes inference and detection theory) are used for deriving from the measured response matrices the Bayes inference on risk of unexpected world state, and estimates of discriminabilities and decision criteria d', c and A, b respectively. These in turn are used to provide minimum framerate estimates for maximizing d' and A, and minimizing prediction error risk. We finish with a conclusion and outlook in Sect. 5.

2 Methods

2.1 Subjects

Thirteen active German tower controllers were recruited as volunteer subjects for the experiment. The participants' ages ranged from 25–59 yrs. and were divided into 3 experimental groups of 4, 4, 5. Controllers from small, medium, and large German airports were approximately evenly distributed to the groups.

2.2 Apparatus

The experiment was conducted at a Remote Tower (RTO) videopanorama-console as part of the DLR Apron-and-Tower Simulator (ATS) of the Braunschweig DLR facility. This simulation system was used to generate 60 landings of a lightly loaded A319 transport at the Braunschweig airport with a 1680 m runway 08/86 (Fig. 1, RWY was extended to 2500 m after this experiment). The simulated aircraft would first appear from E on the right most monitor while in the air at 300 m altitude 32 s before touch down (Fig. 2). Then it would fly to touch down seen on the next monitor to the left. Thereafter, it would either roll through to the end of the runway or stop 250 m before the runway end.

The simulator generated 60 1 min landing scenarios with various dynamically realistic deceleration profiles of nominally 1, 2, or 3 m/s^2 maximum (initial) braking and frame rates of either 6, 12, and 24 fps emulating the video signals potentially coming from cameras mounted near the Braunschweig tower. Only the highest deceleration (3 m/s^2) was sufficient to cause the aircraft to stop near the stopping point (Fig. 1) before the end of the runway (leftmost monitor in Fig. 2).The video files were then used in turn as input simulating the actual cameras so the participants could use the video console as if it were connected to actual cameras on the airfield. They present approximately a 180° view as seen from airport tower but compress it to an approximately 120°. Viewing distance between operators and monitors (21″ UXGA: 1600 × 1200 pixels with 4/3 format: 42 × 33 cm, luminousity sufficient for photopic office environment) was ca. 120 cm. An upper array of tiled monitors for a second airport was present but not used during the testing.
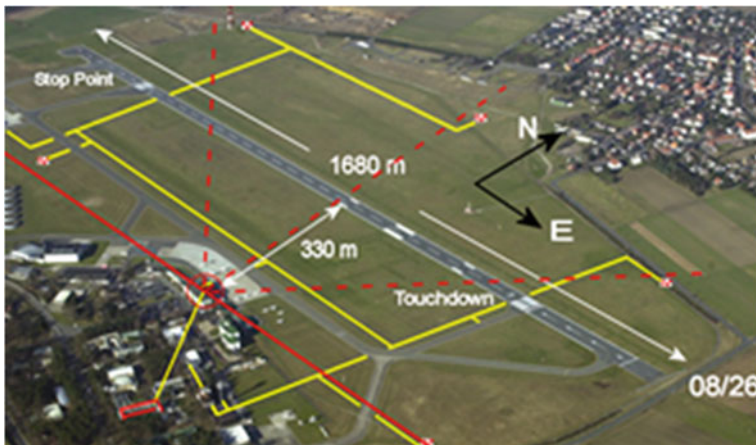
Fig. 1 Aerial view of Braunschweig airport showing the circled location of the simulated (and real) cameras, fields of view of the four cameras (radial sectors), and some dimensions and reference points (Ellis et al., 2011a; Fürstenau et al., 2012)

Fig. 2 Participant at a simulation console judging the outcome of a landing aircraft just after touchdown (2nd monitor from left). Approach on the rightmost monitor, touchdown is on the left side of second monitor from the right. Reconstructed panorama compressing the 180°-tower view to ca. 120° for subjects at the RTO-console (Fürstenau et al., 2012)

2.3 Experimental Design and Task

The three matched subject groups were used in an independent groups, randomized block design in which the three different landing deceleration profiles were used to produce 60 landings to the west on the Braunschweig airport's Runway 26. Each group was assigned to one of the three video frame rate conditions. The approaches were all equivalent nominal approaches for an A319 aircraft but varied in the amount of deceleration after touchdown.

The equation of motion used for the post-processing of logged simulation data assumed that the only braking force (deceleration) after touchdown is given by:

$$\ddot{x} = -b_{\min} - (b_0 - b_{\min})e^{-\frac{t}{\tau}} \tag{1}$$

with $d^2x/dt^2(t=0) = -b_0$, i.e. braking acceleration is assumed to consist of a constant and an exponentially decreasing part. Of course this is a strongly simplified model which neglects e.g. friction and different external forces like braking via reverse thrust. Parameter values as obtained from exponential fits to the logged simulation data are listed in Table 1. Also listed are the stop times $t_{\text{stop}} = t(v=0)$, $v(t=0) = v_0$ = 70 m/s and positions x_{Stop} as calculated from the solution to (1). The table verifies that only the highest nominal deceleration avoids runway excursion (stop for x < ca. 1500 m).

Braking acceleration profiles (decelerations) according to the equation of motion (1) with parameters in Table 1 are shown in Fig. 3. Calculations refer to runway coordinates with $x \parallel$ RWY, rotated by +4.1° with regard to (E, N, up)-coordinates; x

Table 1 Deceleration Profiles by fitting Eq. (1) to logged deceleration data [published in Fürstenau et al. (2012), with permission]

|	Landing braking parameters		
Nominal value (m/s²)	1.0	2.0	3.0
b_0 (m/s²)	1.33	1.76	2.39
$b_{\min}$ (m/s²)	0.45	1.01	1.64
τ (s)	41.3	22.0	12.0
t_{stop} (s)	85.1	54.4	37.4
x_{stop} (m)	2544	1748	1238
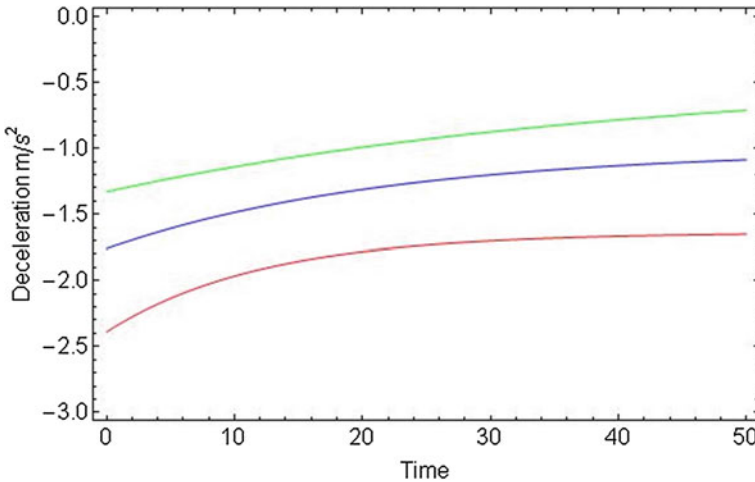
Fig. 3 Deceleration profiles (= decrease of braking acceleration) as obtained by fitting logged simulator data using Eq. (1) for the three nominal braking values 1, 2, 3 m/s² [published in Fürstenau et al. (2012), with permission]

= 0 at ARP. Touchdown is at x = +520 m. Closest distance from observation point to runway is $d_{\text{TWR}} = 330$ m at $x = +245$ m.

The participants' task was to report as soon as possible whether the landing aircraft would stop before the end of the runway (stop event S2 (high deceleration), no-stop event S1 (runway excursion due to low deceleration)), with response time measured by pressing the space bar. In all cases they were then allowed to watch the actual outcome and use a certainty level compatible with actual operations. The three different deceleration profiles were randomized to produce a sequence of 30 landings in 3 blocks of 10. The three blocks were repeated once to provide the 60 landings in the experimental phase used for each of the independent groups. The experimental phase was preceded by a training phase during which the subjects were given familiarity practice with 20 landings similar to those used experimentally. This approach gave participants a chance to learn the task and adapt to a head

Fig. 4 Error Rate as a function of repetition Block (Ellis et al., 2011a), copyright US-government: public domain)
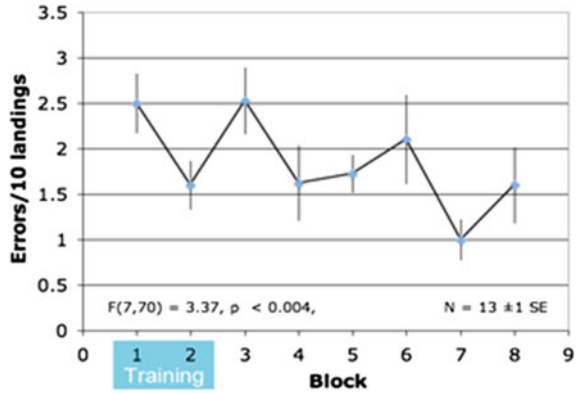
mounted video-based eye tracker that they wore during the experiment.[1] Including instructions, the experiment required 1.5–2 h per subject.

In addition to the objective data, we recorded participants' subjective certainty regarding each of their decisions on a 0–3 Likert-like scale presented after each landing (0-total guess, 3-total certainty).

3 Results

Errors, reaction times and estimates of judgment certainty were subjected to planned Two-Way independent groups ANOVA's based on a mixed design with Subjects nested within Update rate condition but crossed with Repetition which was quantized into 8 Experimental Blocks of 10 landings each, the period of randomization of the deceleration condition. Decision errors appeared to show a learning effect as can be seen in Fig. 4.

But once the training blocks were removed and the remaining blocks grouped into two categories First three (3,4,5) and last three (5,6,7) the statistically significant effect proved unreliable and disappeared ($F(1,10) = 1.52$, ns).

3.1 Response Times

Figure 5 shows the measured response times plotted into a graphic of the airport layout, as measured by participants pushing of the keyboard space-bar at the operator console (see Fig. 2). The space bar pressing with yes-answer (= stop predicted) or no-answer (= overshoot predicted) occurs typically at RT = 10–11 s after observed

---

[1] Eye movements will not be discussed in this Chapter. For analysis of eye movements see chapter "Remotely-Operated AFIS in Japan" and references therein.
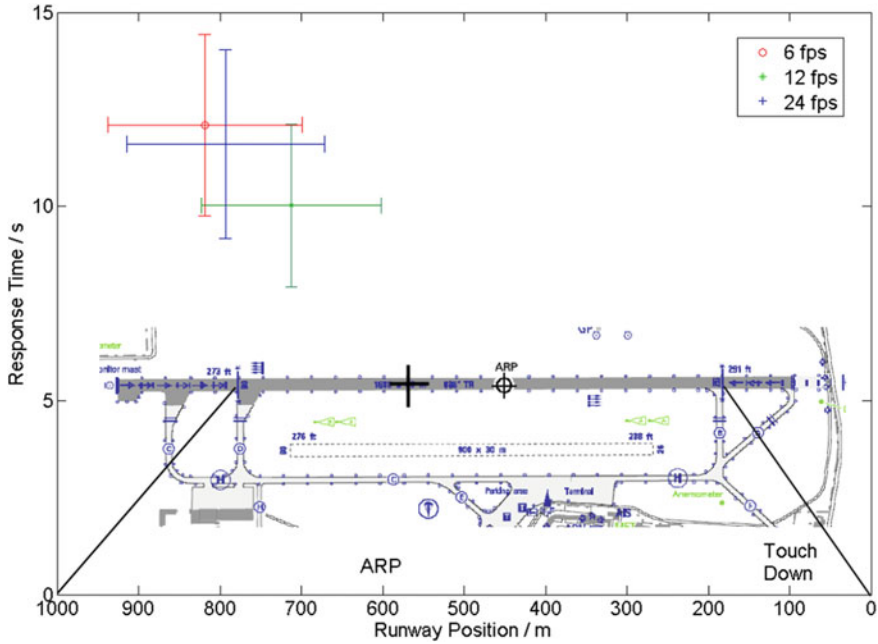
Fig. 5 Airport layout (inset projected to abszissa via solid black lines) with response times (ordinate) typically 10–11 s after touchdown, and with A/C typically around 800 m behind threshold (black cross), separated for the three framerates and averaged over all landings (decelerations) and participants. ARP = Airport reference point at 600 m

touchdown. RT corresponds to A/C positions between 700 and 900 m behind the threshold.

We achieved the goal of approximately equal response times in the different Frame Rate conditions ($F(2,8) = 0.864$, ns). Response times after training remained approximately constant across Blocks with a statistically significant variation ($F(5,40) = 3.91$, $p < 0.006$) of less than $\pm 2.5\%$ when the training blocks were excluded.

3.2 Decision Statistics: Response Matrix

The experimental results of this two-alternative decision experiment concerning decision errors as dependent on video framerate are summarized in the stimulus–response matrices of Table 2. It shows group averages of measured probability estimates, with standard errors of mean (), of correct rejection $C = P(no|S1)$, false alarm FA = $P(yes|S1)$, miss $M = P(no|S2)$, and hit $H = P(yes|S2)$. S1 = stimulus with runway excursion, S2 = stimulus with stop on the runway, yes = stop predicted (high deceleration perceived), no = no stop predicted (low deceleration perceived). Probabilities in horizontal rows (constant stimulus) sum up to 1.

Table 2 Response matrices (measured H, M; C, FA rates) for the three framerates (Fürstenau et al., 2012)

Alternative stimuli	Response for 3 video framerates: probability estimates				
	No-stop predicted			Stop predicted	
Low deceleration No-stop Stimulus S1 $n(S1) = 40$	$p(no|S1)$ $= C$	6	0.86 (0.02)	$p(yes|S1)$ $=$ FA	0.14 (0.02)
		12	0.89 (0.03)		0.11 (0.03)
		24	0.94 (0.01)		0.06 (0.01)
High deceleration Stop stimulus S2 $n(S2) = 20$	$p(no|S2)$ $= M$	6	0.55 (0.06)	$p(yes|S2)$ $= H$	0.45 (0.06)
		12	0.45 (0.05)		0.55 (0.05)
		24	0.22 (0.07)		0.78 (0.07)

These results may be presented in the form of Venn-diagrams as depicted in Fig. 6, that clarifies the character of the measured rates H, M, CR, FA as conditional probabilities and their base sets with regard to situations (world states) S1 = no stop and S2 = stop event.

The different areas (width) of the two columns representing situations (or alternatives) S1, S2 reflect different numbers of experimental no-stop ($n(S1)$) and stop rates ($n(S2)$) respectively to be observed by the subjects, and of corresponding a-priori probabilities $p(S1)$, $p(S2)$: $n(S1) + n(S2) = 60$ with $n(S2)/n(S1) = ½$ (see also Table 2).

As a preliminary analysis of the results Fig. 7 does show a significant effect of frame rate on the average error numbers per 10 landings and invites discussion. Extrapolation indicates a minimum framerate >30 Hz for minimizing decision errors.
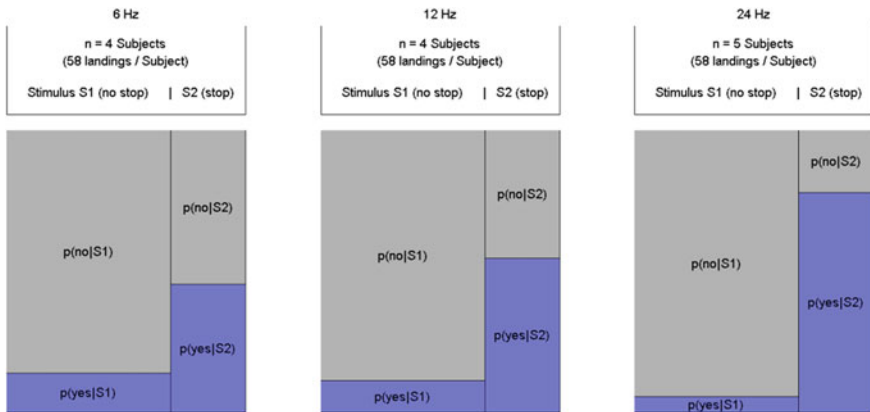
Fig. 6 Venn diagrams representing measured rates of correct ($H = p(y|S2)$, CR $= p(no|S1)$) and false decisions ($M = p(no|S2)$, FA $= p(yes|S1)$) for the two given world states (situations, events) S1 (= no stop on RWY, insufficient braking braking, alternative 1 or "noise", in terms of SDT, see below) and S2 (= stop on RWY, sufficient braking, alternative 2 or "signal + noise", in terms of SDT)

Fig. 7 Mean error Rate as a function of Frame Rate [published in Ellis et al. (2011a), with permission]
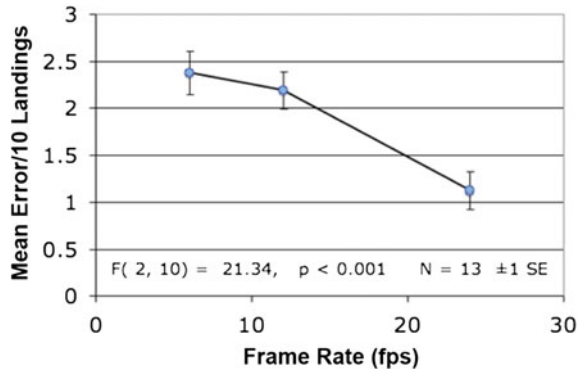
Also it can be seen in Table 2 that like in the averaged error plot of Fig. 7 the measured probability estimates indicate a trend dependent on framerate (FR): the hit rate $H = p(\text{yes}|S2)$ increases with framerate whereas the false alarm rate FA $= p(\text{yes}|S1)$ decreases. We will show in the following data analysis and discussion section how the measured probabilities in the response matrix can be used for deriving a (Bayes) inference on risk probabilities for safety critical decisions, dependent on the video framerate as system parameter (risk for a world state different from the predicted event, i.e. risk of surprise situation) by using the a priori knowledge on relative frequencies of the planned experimental situation alternatives S1, S2.

Besides the Bayes inference the conditional probabilities of the detailed response matrix (Table 2, Fig. 6) will be used to derive a theoretically grounded data analysis for narrowing down the quantitative framerate requirements. Specifically the measured estimates of response probabilities conditional on the priori knowledge of experimental conditions ($p(S1)$, $p(S2)$), suggests the use of signal detection theory (SDT) to derive a quantification of the detection sensitivity (discriminability) as the basis for estimating FR_{min}. This SDT-discriminability is free of a subjective criterion, i.e. free of a tendency towards more conservative (avoiding false alarms) or more liberal (avoiding misses) decision. For extrapolating towards a minimum required framerate we will provide an initial hypothesis of a perceptual model to be used for fitting our data. A model based data analysis would also provides guidelines for future experiments with the potential to generate further evidence supporting the conclusion.

Interestingly, during debriefings after the experiment subjects in the lower two frame rate groups reported that they felt the aircraft were moving "too fast" and that it was this extra apparent speed making discrimination hard. By "too fast" the controllers meant to refer to the apparent ground speed of a transport aircraft compared to what they would expect to see from a tower.

We examined this possibility by looking at a response bias that could arise from aircraft appearing to move "too fast." Such a bias would lead subjects to underestimate whether an aircraft actually coming to a stop would in fact stop, because it would seem to be going too fast. Aircraft in fact not stopping would not be subject

to a bias since they would merely seem to be overshooting the end of the runway in any case. Thus, we would expect subjects to be more likely to incorrectly identify a stopping aircraft (S2) as non-stopping versus one that is not stopping (S1) as stopping. Details of this analysis are also presented in the following data analysis and discussion (Sect. 4).

4 Data Analysis and Discussion

The present analysis will start with the simulation results of the movement/braking dynamics as obtained by integration of Eq. (1) using the parameter values of Table 1 with decelleration profiles of Fig. 3. It provides an impression of the requirements on perceptual discrimination during the experiments. The second subsection provides derivation of the Bayes-inference on risk of unexpected world states by using likelihood values and a priori knowlwdge based on the response matrix of Table 2. The Bayes risks in turn are used for estimating via linear regression the minimum frame rate requirement that minimizes the risk of predicting the false world state. This result will be compared to the frame rate extrapolations of maximum discriminability based on a hypothesized exponential discriminability decrease as obtained obtained from sensitivity index d' and nonparametric discriminability A (= average area under the ROC-curves). Also the associated response bias will be discussed in more detail.

4.1 Simulation of Movement After Touchdown

The integration of the simplified equation of motion (1) for the braking dynamics with accelerations shown in Fig. 3 yields the observed angular movement at the simulated control tower/camera position after transformation into the corresponding reference frame. The result for the velocity dependence on runway position before the transformation is shown in Fig. 8.

This phase- (ore state-) space diagram velocity v (position x) confirms that in fact only the highest deceleration value (red line) leads to a stop at 1200 m) before the runway end (at 1650 m). The medium braking results in a slight overshoot whereas the lowest deceleration leads to a dramatic runway excursion. The following Fig. 9 shows how this result translates into the viewing angle coordinates of an observer at the tower position.

The participants prediction about stop/no stop or sufficient/insufficient braking is done some time after passing the 0-angle point at ca. 44–48°, corresponding to the response time $R = 10$–11 s and 700–900 m distance from touchdown. In fact the decision seems to depend on subtle differences between trajectories in angular state space at decision time considering the fact that the real 180°-panorama view is compressed to ca 120° in the RTO-CWP panorama reconstruction. It was unclear during the preparation phase of the experiment if these small differences were large
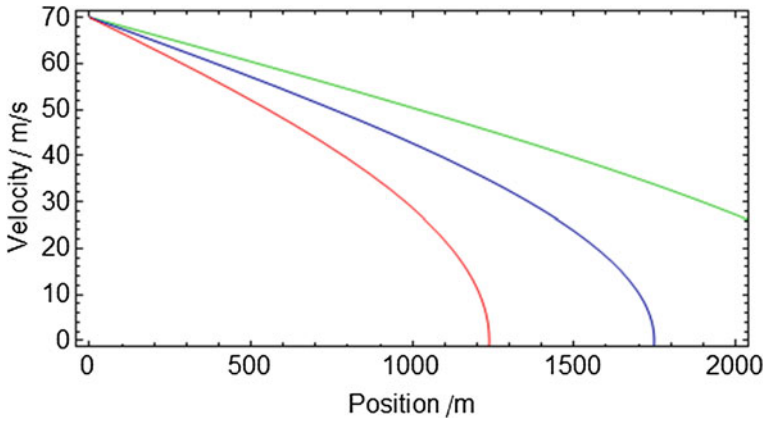
Fig. 8 Phase or state space diagram depicting simulated velocity (integration of equation of movement (1)) versus position

Fig. 9 Simulated angular velocity versus observation angle phase space after transformation of integrated equation of movement into observer coordinates at tower position. Highest angular speed near the normal from TWR to the RWY. $R =$ 10–11 s is at 44–48 deg
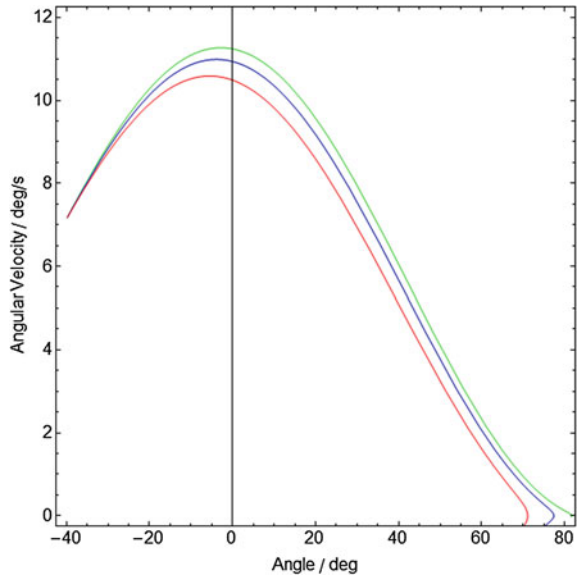
enough for discriminating at all between sufficient (stop event) and insuffient braking (no-stop event).

4.2   *Bayes Inference: Risk of Unexpected World State*

The Bayes inference probabilities, with standard errors of mean (), about unexpected event S1 (runway excursion with predicted stop) and unexpected situation S2 (stop occurring no stop predicted) as calculated via Bayes law using the measured likelihoods (yes, or no predictions conditional on situations S1 and S2 respectively) are summarized in Table 3. Here the probabilities (for the same FR) of the columns add to 1.

The runway overshoot probability conditional on stop predicted (Bayes inference on the probability of world state S1 different from prediction "stop" based on perceived evidence) is given by

$$p(S1|yes) = p(yes|S1)\, p(S1)/p(yes) \tag{2}$$

with a priori knowledge of no-stop stimulus probability $p(S1) = n(S1)\,/(n(S1) + n(S2)$, according to the ratio of the Venn diagram areas and $p(yes) = p(yes|S1)p(S1) + p(yes|S2)p(S2)$. Equation (2) quantifies the risk of an overshoot occurring when predicting a stop, i.e. a surprising unexpected world state. It is proportional to the likelihood of missing a planned overrun $p(yes|S1)/p(yes)$ (for a brief introduction on Bayes inference and references see Appendix B).

Figure 10 depicts the Bayes probability estimates for unexpected (surprise) world states dependent on framerate, i.e. (a) unexpected runway excursion (S1) conditional on erroneous perception of a high braking deceleration (answer "yes": stop predicted) and (b) the probability $p(S2|no\text{-}stop) = p(n|S2)\, p(S2)/p(n)$, that an unexpected stop occurs when predicting no-stop. Both surprise events suggest a linear fit to the three framerate data as most simple model. As expected the $p(S1|yes)$-graph (upper three data points) shows that for decreasing frame rates (FR $\rightarrow$ 0) the conditional probability for a runway excursion occurring when a stop is predicted rises to chance (0.48 $\pm$ 0.01).

Table 3 Bayes Inference matrix for probabilities of actual world states (situations) conditional on decisions based on perceived evidence (likelihood × a priori knowledge). Published in [Fürstenau et al. (2012), with permission]

| Event Alternatives	Bayes inference on event probabilities conditional on prediction				
	No stop predicted (no-response)			Stop predicted (yes-response)	
Low deceleration No stop event S1	$p(S1|no)$	6	0.78 (0.02)	$p(S1|yes)$	0.40 (0.03)
		12	0.81 (0.02)		0.30 (0.04)
		24	0.91 (0.02)		0.13 (0.02)
High deceleration Stop event S2	$p(S2|no)$	6	0.22 (0.02)	$p(S2|yes)$	0.60 (0.03)
		12	0.19 (0.02)		0.70 (0.04)
		24	0.09 (0.02)		0.87 (0.02)
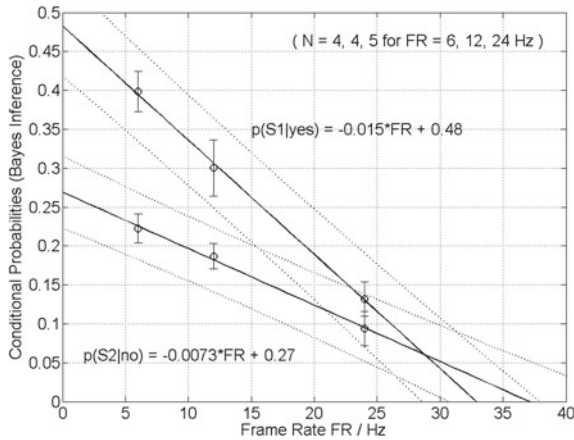
Fig. 10 Bayes inference for the three framerates (Abscissa) on probability of **a** (upper data points and fit) unexpected situation S1 "a/c will not stop before RWY-end" (braking acceleration < threshold), given the alternative (false) stop-prediction, as calculated from measured likelihoods of subjects predicting "stop on RWY" conditional on S1 (=FA); and **b** (lower data points and fit) of world state S2 "a/c will stop before RWY-end" (braking acceleration > threshold) as calculated from measured probabilities (likelihood) of subjects predicting "overshoot", conditional on S2 (a priori knowledge). Ordinate: mean (with stderr of mean) of probability for (unexpected) situation S_i conditional on prediction/decision d_i, averaged for all subjects within each FR-group. Straight line = linear fit with 95% confidence intervals (dotted)

Comparing both graphs one immediately recognizes a bias of the lower one, with $p(S2|no) \rightarrow 0.27$ for $FR \rightarrow 0$ Hz, indicating a significantly reduced number of unexpected stop events conditional on the false "no" response, as would be expected by chance for $\lim FR \rightarrow 0$. As mentioned above the S2/S1 imbalance of 1/3 stop events and 2/3 no-stop partly explains this bias: the extrapolation to $FR = 0$ (no movement information available), yields $p(S2|n) = 0.27$ and $p(S1|n) = 0.73$ for the complimentary case so that for low FR with large position jumping $p(S2|n)/p(S1|n) \approx 0.4$ reflects the S2/S1 imbalance of 1/2. The decrease of the $p(S2|n)$-bias and decision bias $p(n|S2)$ (tendency for false overshoot prediction under S2) with increasing FR goes in parallel with the decreasing overall decision error. So the Bayes analysis confirms the previously reported decision bias (Ellis et al., 2011a, 2011b) as quantified by M–FA = $p(n|S2)–p(y|S1)$ which also decreases with increasing framerate (see Fig. 11). Within the 95% confidence interval of the linear fit to the data also $p(S2|no)$ predicts zero bias and 100% correct response for frame rates >35 Hz, which is compatible with the FR-limit of zero-error prediction obtained with the "unexpected stop"—probability. The linear extrapolation of the Bayes analysis narrows the initial estimate of FR_{min} > 30 Hz as depicted in Fig. 7, to ca. 30– 45 Hz in Fig. 10.

The hypothetical visual memory effect mentioned above would suggest an exponential approach to a minimum error probability with increasing FR instead of a linear

Fig. 11 Error bias (M–FA, normalized for ten landings; $N = 13$, see Figs. 4 and 7) towards reporting a runway overrun increases the likelihood of missing a planned stop over missing a planned overrun. Effect decreases with FR [re-drawn from Ellis et al. (2011a), with permission]
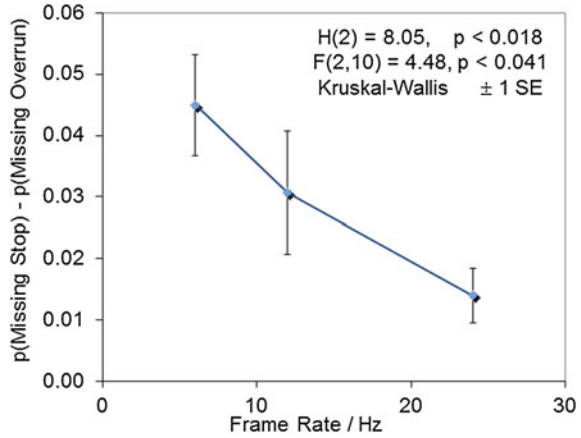
behavior. The exponential fit to our data, however yields a significantly reduced goodness ($F = 140$, $p = 0.054$) as compared to the linear case ($F = 645$, $p = 0.025$), which demonstrates the necessity of experimental data at higher framerates.

The Bayes analysis also confirms the observation reported before in Ellis et al. (2011a, 2011b) (see also below, Fig. 11) that the error bias appears exclusively connected with the preference of no-stop decisions, i.e. unexpected stop situations with a lower than chance error probability at FR = 0, because the false-stop prediction errors, as expected yield a chance Bayes probability $p(S1|yes) = 0.5$ for FR $\rightarrow$ 0 (see Fig. 10). The same is true for the complementary case $p(S2|yes)$. The observation of a significant bias of the unexpected-stop event inference ($p(S2|no)$) suggests the need for counter measures, perhaps temporal filtering to smooth out the discontinuities. Such an approach would undoubtedly benefit from a computational model of speed perception. One starting point for such analysis of the speed perception error could be the spatio-temporal aliasing artifacts that introduce higher temporal frequency information into the moving images.

The measured probabilities of Table 2 used for calculating the Bayes inference are based on error statistics composed of intrinsic discriminability and subjective criteria, i.e. it includes a decision bias or subjective preference for positive or negative decisions. In what follows parametric and nonparametric variants of signal detection theory (SDT) are used for quantitatively separating both contributions and comparing the resulting FR$_{min}$-estimates with those of the Bayes inference.

4.3 Response Bias

A response bias is a well known effect of low video frame rate on apparent speed of moving objects that is caused by *undercranking, a* movie camera technique of slowing

the image frame capture rate compared to the display rate, e.g. for visualizing the growth of plants at an apparently higher speed.

From the results described above we would expect subjects to be more likely to incorrectly identify a stopping aircraft versus one that is not stopping. Indeeed when we compared the likelihood of erroneously identifying an overshoot versus that of erroneously identifying a stop (Table 2) M–FA = $p(n|S2)$–$p(y|S1)$, all 13 subjects showed this bias. (sign-test, $p < 0.001$). This general bias towards identifying an aircraft as not stopping, however, is not surprising since approximately twice as many aircraft observed in fact do not stop versus those that do ($p(S1) = 2\,p(S2)$) and subjects quickly sense this bias during the experiment. What is interesting, however, is that the bias is a decreasing function of the frame rate as depicted in Fig. 11.

The significance of this result, however needs support based on theoretical considerations and on alternative analysis. The detection bias is clearly reflected by the Bayes analysis as performed above (Fig. 10). Like the error difference it exhibits a lower than chance probability for $p(S2|no)$ with lim FR $\rightarrow$ 0, yielding $p(S1|yes)/p(S2|no) \approx \frac{1}{2}$, that reflects the $p(S1)/p(S2)$-ratio and like the above error difference converges to zero with increasing FR.

Of particular practical interest is the inferred risk of missing a high speed turnoff or of a runway excursion occurring when a stop is predicted, i.e. the conditional probability of overshoot $p(S1|yes)$ (S1 = no stop event) due to low or abnormal braking when evidence suggests normal braking (stop prediction).

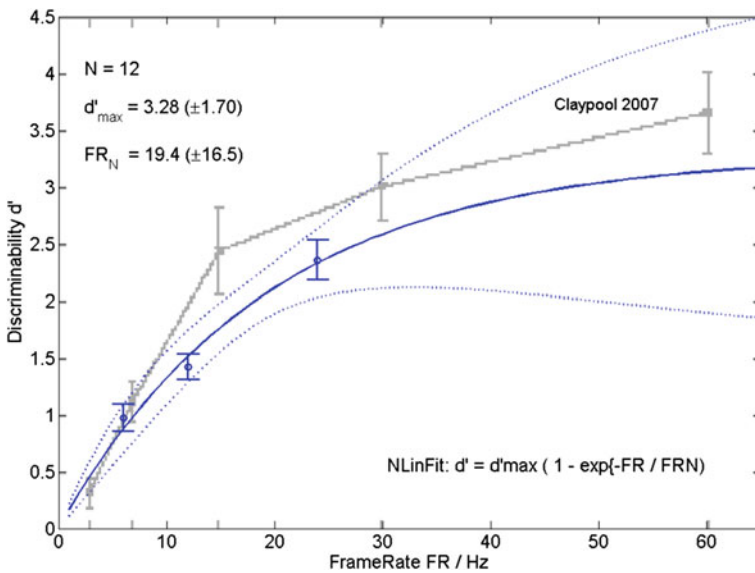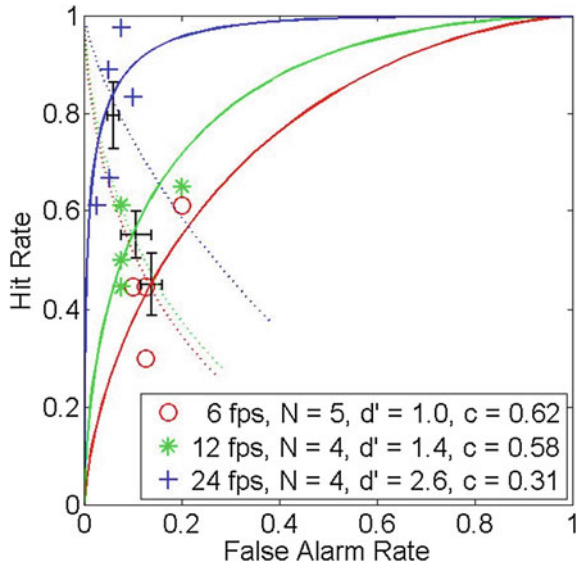## 4.4   SDT Discriminability d′ and Decision Bias c

The principal result of data analysis using signal detection theory (SDT) is shown in Figs. 12 and 13. It confirms the Bayes analysis and suggests that relatively high update rates $FR_{min} > 30$ Hz will be required for imagery in virtual or remote towers if controllers working in them are expected to perform the kinds of subtle visual motion discrimination currently made in physical towers. Figure 12 depicts the experimental results of Table 1 in ROC-space (receiver operating characteristics) H vs. FA. Plotted are the measured hit and false alarm rates for the 13 participants and the three framerates together with the respective averages (black crosses) and the ROC isosensitivity- and isobias-curves, parametrized by discriminability $d′$ and criterion value c respectively. $d′$ and c are calculated according to:

$$d\prime = 0.5(z(H) - z(FA)) \tag{3}$$

$$c = -(z(H) + z(FA)) \tag{4}$$

with z = z-score of cumulative Gaussian densities of the S1-, S2-familiarity distributions (see also Appendix A2).

Fig. 12 ROC curve pairs parametrized (d', solid curves, c, dotted curves) for each of the three frame rates based on Hit and False Alarm rates for each subject. Crosses are the averages for each framerate subgroup of participants. ROC-curves $d'(z(H), z(\text{FA}))$ and $c(z(H), z(\text{FA}))$ are calculated with the d' and c subgroup-averages of the 13 participants. Experimental data and d' parametrized ROC curves correspond to results initially presented in Ellis et al. (2011a)

Legend:
- ○ 6 fps, N = 5, d' = 1.0, c = 0.62
- * 12 fps, N = 4, d' = 1.4, c = 0.58
- + 24 fps, N = 4, d' = 2.6, c = 0.31

N = 12
d'_{max} = 3.28 (±1.70)
FR_N = 19.4 (±16.5)

Claypool 2007

NLinFit: $d' = d'_{max} (1 - \exp\{-FR / FRN\})$

Fig. 13 Group averages ($N = 12$ subjects) of experimental discriminability values d' and exponential regression model (blue solid trace) for the stop/no-stop discriminability of landing aircraft. The lighter grey trace plots comparative data from Claypool and Claypool (2007). Dotted lines shows the 95% regression confidence range. Comparable results for 13 subjects were initially presented in Ellis et al. (2011a)

The positive criterion values indicate the controllers tendency to make conservative decisions, i.e. avoiding false alarms, increasing misses and trying to be certain about their decisions, according to their work ethics and the written instructions of the experiment. The decrease of this effect is consistent with the decreasing error bias M–FA with increase of FR as reported above.

In Fig. 13 we have also replotted a result from Claypool and Claypool (2007) examining the effect of change in frame rate on video game shooting score. These overlaid data empirically support our theoretical supposition that the users performance at higher and higher frame rates may be modeled by an exponentially approached limit. It is certainly interesting that their report of the effect of frame rate on video game score in a first-person-shooter game resembles our results since their task and response measure was so different. In particular, their use of shooting score does not capture the interplay of shooting frequency and hits in a way analogous to that of correct detections and false alarms in our experiment.

Our analysis of d' is in contrast to their count of shots on target and it is particularly useful since it can be argued to be bias-free, independent of user criteria and primarily a function of the task requirements and perceptual estimation noise. It can additionally be cross checked with extrapolation of the error data shown in Fig. 4 and the Bayes inference in Fig. 10, but this extrapolation for errors is harder to justify theoretically without a computational error model. A linear extrapolation which likely underestimates the value, however, suggests a ~40 fps would be needed for a vanishingly small error rate. Based on our exponential memory (sample-and-hold) decay hypothesis the asymptote of the d'(FR)-analysis, like the Claypool (2007) data indicates a higher FR_{min} value, more towards 60 Hz.

4.5 *Nonparametric Discriminability* A *and Decision Bias* b

Detectability *A* and likelihood bias parameter *b* were suggested as improved "nonparametric" alternatives of the conventional discriminability d' and criterion c because it requires fewer statistical assumptions (in its final form it was presented by Zhang and Mueller (2005). In Ellis et al. (2011b) we compared *A* with d' to estimate user sensitivity of detection that an aircraft will stop. Discriminability *A* and *b* are independent of the distributional assumptions required for deriving the conventional d' and *c* parameters for detectability and bias (see Appendix A2). The Zhang and Mueller formulas yield the average area A under all possible proper ROC curves (i.e. all concave curves within the range (0,0)–(1,1)) with non-increasing slope, obtained from the measured hit (*H*) and false alarm rates (FA). The constant A-isopleths cut the constant *b*-isopleths at the group mean (<FA> , <H>) coordinates which are used for calculating the *A* and *b*-ROC-curves: $A := A_{mean}(H,FA)$ and $b := b_{mean}(H,FA)$ for the three different framerate conditions according to the Zhang and Mueller equations (see Appendix B).

Fig. 14 Measured hit versus false alarm rates (*H*, FA) for all 13 subjects and the three group averages with standard errors (crosses) and with ROC-curves for the three framerates. Straight lines = constant sensitivity *A*-isopleths; dotted lines = constant bias (likelihood ratio) *b*-isopleths [Results published in Fürstenau et al. (2012), with permission]
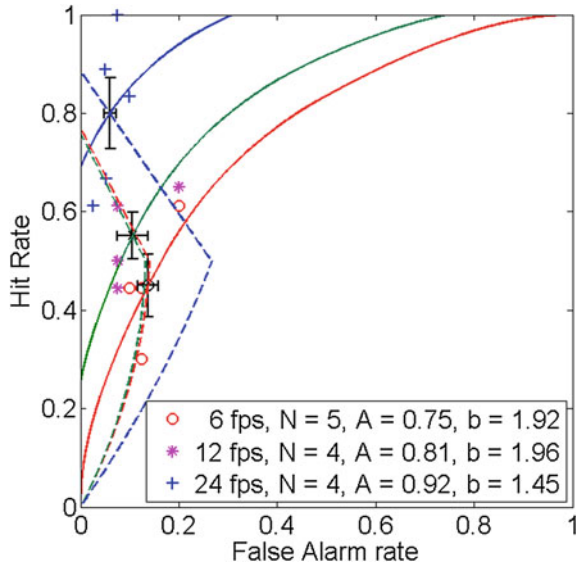

Figure 14 shows the measured hit rates versus false alarm rates for all subjects together with their means (black crosses, as given in Table 1) and isopleths parametrized by constant discriminability A(FR) and constant decision bias b(FR).

Individual hit rates (relative frequencies) are scattered between 0.3 and 1, whereas false alarms rates concentrate in the low probability range < 0.2, indicating conservative decisions, as would be expected for trained air traffic controllers. Circles, stars and crosses represent individual measurements (Hit, False Alarm) for FR = 6, 12, 24 Hz respectively, as obtained from the 13 subjects with repeated measurements (60 landings per subject). Black crosses with error bars show the group mean values of the individually measured (*F*,*H*)-values and the standard errors of means for the three different framerates. Solid curves represent the isopleths parametrized with the group mean *A*-values via Eqs. (15) in Appendix B. The three dotted curves represent the decision bias *b*, obtained from the parametric representation given in Appendix B. *b* apparently decreases with sufficiently high framerate FR towards the neutral criterion value $b = 1$ which confirms the Bayes inference result in Fig. 10 that the overestimation of speed (error bias in favor of misses, decreasing FA) decreases with framerate: the criterion shifts to more liberal values.

The three (group-average) discriminability parameters A(FR) are depicted in Fig. 15 together with an exponential fit and 95% confidence intervals (using Matlab "Nlinfit").

Again, like in the d'(FR)-analysis the exponential model fit to our three data points is based on the hypothesis that low framerates might disturb the visual short term memory so that with increasing visual discontinuity the speed estimate or sequential sampling of the speed information up to the decision time becomes biased. Since the A parameter unlike the classical d', does not require the usual assumptions of Signal
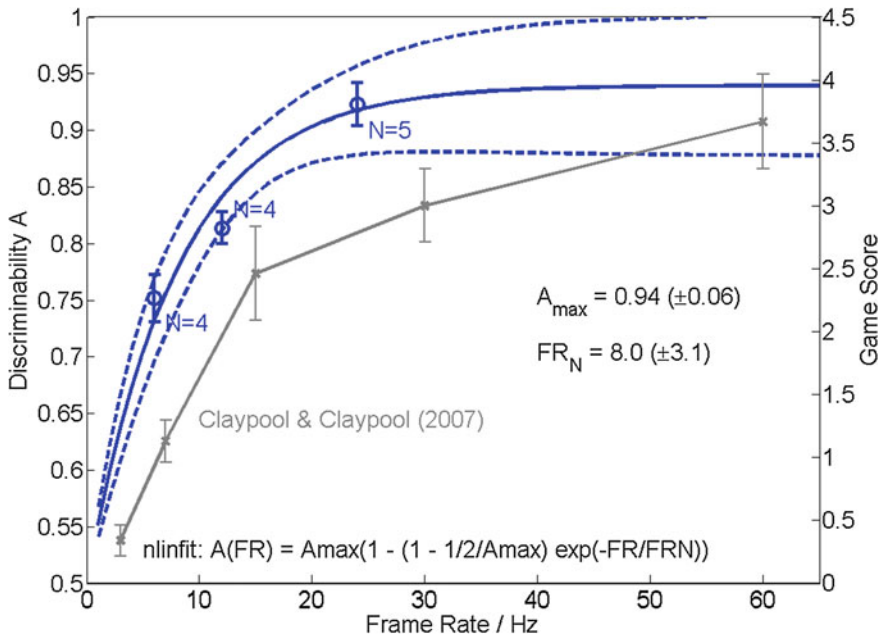
Fig. 15 Group averages (13 subjects) and exponential regression model for A (darkest solid trace) of the discriminability of landings with stopping vs non-stopping aircraft. 95% regression confidence intervals flanks the model fit. Lighter grey trace shows re-drawn comparative data from Claypool and Claypool (2007) [Result published in Ellis et al. (2011b), Fürstenau et al. (2012), with permission]

Detection Theory (SDT), e.g., normality of both the signal and noise distributions, it may be considered to provide a better estimate of the frame rate at which participants' performance asymptotes as provided in Ellis et al. (2011a) (see previous section). From Fig. 15 this value seems to be in the range 30–40 fps, a result close to the Bayes analysis with linear model extrapolation, (see above), whereas the parametric SDT analysis $d'(FR)$ appears to asymptote at a significantly larger value.

Alternatively and for the sake of parsimony our three data points, like with the Bayes analysis may be fitted with a straight line, yielding an extrapolation to ca. 31 Hz for $A = 1$ (maximum discriminability), which lies at the lower end of the Bayes fit confidence intervals.

Like in the $d'(FR)$ analysis our results are compared with the (re-drawn) published results of Claypool and Claypool (2007). The latter were obtained with subject scores in a shooter game under different framerates. As mentioned above they suggest a significantly higher asymptotic FR-value for maximizing shooter scores as compared to our extrapolation in Fig. 15, apparently more consistent with our $d'(FR)$-extrapolation.

Clearly, additional experiments with FR > 30 Hz are needed, if possible supported by a well founded theoretical model, in order to clarify this discrepancy between the different data analysis approaches. In the following chapter "Which Minimum Video

Frame Rate is Needed in a Remote Tower Optical System?" the perception of a broad spectrum of real-life traffic situations (typically dominated by lower angular velocity and acceleration) is investigated (replayed with selected videos using low frame rates 2, 5, 10, 15 Hz) which contrast our findings insofar as the perceived video quality and operability was subjectively judged "acceptable" for FR = 15 Hz.

5 Conclusion

It is clear from controller interviews that numerous out-the-windows visual features are used for control purposes (Ellis & Liston, 2010, 2011; Schaik et al., 2010) (see also chapters "Visual Features Used by Airport Tower Controllers: Some Implications for the Design of Remote or Virtual Towers", "Detection and Recognition for Remote Tower Operations"), which in fact go beyond those required for aircraft detection, recognition, and identification (Watson et al., 2009). In the present work, for analyzing frame rate effects on prediction errors we focused on the landing phase of aircraft because we expected any perceptual degradation to be most pronounced in this highly dynamic situation.

Our preliminary results on the minimum framerate for minimizing prediction errors ($FR_{min} > 30$ Hz) show that a definitive recommendation of a minimum video framerate and a confirmation of our initial hypothesis of visual short-term memory effects resulting in the proposed asymptotic characteristic requires a further experiment with FR > 30 Hz. This high-FR experiment was not possible with the video replays used in the described experiments for technical reasons. Obviously, the presented experimental data are not sufficient to decide in favor of the visual short term memory hypothesis versus a heuristic decision basis, e.g. sequential sampling or comparison of time dependent aircraft position with landmarks for thresholding. One alternative approach might be some variant of a relative judgement or diffusion model of two-alternative decision making [e.g., (Ashby, 1983)].

A formal model for predicting the hypothetical visual memory effects would also be of great help. Recent studies which might be of use for this purpose investigate neural models for image velocity estimation [e.g., (Perrone, 2004)] and quantify the temporal dynamics of visual working memory by measuring the recall precision under periodic display presentations between 20 ms and 1 s (Anderson et al., 2011; Bays et al., 2011).

Also more detailed tower controller work analysis would be useful to clarify the operational relevance of increased framerate for decision error reduction with dynamic events in the airport environment.

loop RTO-simulation trial, and we are indebted to them for valuable assistance in the conduct of the experiment.

References

Ashby, F. G. (1983). A biased random walk model for two choice reaction times. *Journal Mathematical Psychology, 27*, 277–297.

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *Journal of Neuroscience, 31*, 1128–1138.

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision, 11*(10), 1–15.

Claypool, K. T., & Claypool, M. (2007). On frame rate and player performance in first person shooter games. *Multimedia Systems, 13*, 3–17.

Ellis, S. R., & Liston, D. B. (2010). Visual features involving motion seen from airport control towers. In *Proceedings of 11th IFAC/-IFIP/IFORS/IEA Symposium on Analysis, De-sign, and Evaluation of Human-Machine Systems 31 Sep–3 Oct 2010*. Valenciennes, France.

Ellis, S. R., Liston, D. B. (2011). *Static and motion-based visual features used by airport tower controllers*. NASA TM-2011–216427. Ames Research Center, Moffett Field, CA.

Ellis, S. R., Fürstenau, N., & Mittendorf, M. (2011a). Frame rate effects on visual discrimination of landing aircraft deceleration: Implications for virtual tower design and speed perception. In *Proceedings Human Factors and Ergonomics Society, 55th Annual Meeting*, Sept. 19–23 (pp. 71–75). Las Vegas, NV USA.

Ellis, S. R., Fürstenau, N., & Mittendorf, M. (2011b). Determination of frame rate requirements of video-panorama-based virtual towers using visual discrimination of landing aircraft deceleration during simulated aircraft landing. *Fortschritt-Berichte VDI, 22*(33), 519–524.

Fürstenau, N., Mittendorf, M., Ellis, S. (2012). *Remote towers: Videopanorama frame rate requirements derived from visual discrimination of deceleration during simulated aircraft landing.* In D. Schäfer (Ed.), *Proceedings 2nd SESAR Innovation Days, Braunschweig.* http://www.sesarinnovationdays.eu/files/SIDs/2012/SID%202012-02.pdf.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Papenfuss, A., Kaltenhäuser, S. (2009). Steps towards the virtual tower: Remote airport traffic control center (RAiCe). In *Proceedings of EIWAC 2009, ENRI International Workshop on ATM & CNS* (pp. 67–76). Tokyo.

Grunwald, A. J., & Kohn, S. (1994). Visual field information in low-altitude visual flight by line-of-sight slaved head mounted displays. *IEEE Systems Man and Cybernetics, 24*, 120–134.

Hannon, D., Lee, J. T., Geyer, M., Sheridan, T., Francis, M., Woods, S., & Malonson, M. (2008). Feasibility evaluation of a staffed virtual tower. *The Journal of Air Traffic Control, Winter*, 27–39.

Kempster, K. A. (2000). *Frame rate effect on human spatial interpretation of visual intelligence.* Masters thesis, NPS, Monterrey CA.

Perrone, J. A. (2004). A visual motion sensor based on the properties of V1 and MT neurons. *Vision Research, 44*, 1733–1755.

Schmidt, M., Rudolph, M., Werther, B., & Fürstenau, N. (2007). Development of an augmented vision videopanorama human-machine interface for remote airport tower operation. In *Proceeddings of HCII2007 Beijing, Springer Lecture Notes Computer Science 4558* (pp. 1119–1128).

SESAR-JU Project. (2003). Remote Provision of ATS to a Single Aerodrome—Validation Report, edition 00.01.02, www.sesarju.eu.

Van Schaik, F. J., Lindqvist, G., & Roessingh, H. J. M. (2010). Assessment of visual cues by tower controllers. In *Proceedings 11th IFAC/IFIP/IFORS/IEA Symosium on Analysis, Design, and Evaluation of Human-Machine Systems.* Valenciennes, France.

van Scheijk, F. J., Roessingh, J. J. M., Bengtsson, J., Lindqvist, G., & Fält, K. (2010). Advanced remote tower project validation results. In *Proceedings 11th IFAC/IFIP/IFORS/IEA Symposium*

on Analysis, Design, and Evaluation of Human-Machine Systems. Valenciennes, France. https://doi.org/10.3182/20100831-4-FR-2021.00025.

Watson, A. B., Ramirez, C. V., & Salud, E. (2009). Predicting visibility of aircraft. *PLoS ONE, 4*(5), e5594. https://doi.org/10.1371/journal.pone.0005594.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrica, 70*(1), 203–212.

Which Minimum Video Frame Rate is Needed in a Remote Tower Optical System?

Jörn Jakobi and Maria Hagl

Abstract Bandwidth, often limited and costly, plays a crucial role in cost-efficient Remote Tower system. Reducing the Video Frame Rate (in the following referred to Frame Rate (FR), expressed in fps, also referred to video update rate (European Organisation for Civil Aviation Equipment. EUROCAE: Minimum aviation system performance specification for remote tower optical systems, ED-240A change 1, EUROCAE, Saint-Denis: 2021)) of the relayed video stream is one parameter to save bandwidth, but at the cost of video quality. Therefore, the present article evaluates how much FR can be reduced without compromising operational performance and human factor issues. In our study, seven Air Traffic Control Officers (ATCOs) watched real air traffic videos, recorded by the Remote Tower field test platform at the German Aerospace Center (DLR e.V.) at Braunschweig-Wolfsburg Airport (BWE). In a passive shadow mode, they executed ATS relevant tasks in four different FR conditions (2, 5, 10 and 15 fps) to objectively measure their visual detection performance and subjectively assess their current physiological state and their perceived video quality and system operability. Study results have shown that by reducing the FR, neither the visual detection performance nor physiological state is impaired. Only the perceived video quality and the perceived system operability dropped by reducing FR down to 2 fps. The findings of the study presented in this chapter will help to better adjust video parameters in bandwidth limited applications in general,

J. Jakobi (✉)
German Aerospace Center, Institute of Flight Guidance, Lilienthalplatz 7, 38108 Braunschweig, Germany
e-mail: joern.jakobi@dlr.de

M. Hagl
SHS - Sciences de L'Homme et de la Société Psychologie, Université Grenoble-Alpes, Grenoble, France
e-mail: maria.hagl@univ-grenoble-alpes.fr

405

and in particular to alleviate large scale deployment of Remote Towers in a safe and cost-efficient way.

Keywords Video update rate · Frame rate · Frames per second · Detection perfomance · Physiological stress

1   Introduction

In 2015 the first Remote Tower installation went in operation. Swedish ATCOs control air traffic of Sundsvall and Örnsköldvik airport from the Remote Tower Center (RTC) in Sundsvall (SAAB, 2018). Despite this first success, ambitions to improve the Remote Tower concept run high. Thus, new modalities for controlling a pan-tilt-zoom camera (Virtual & Remote Control Tower, 2016c) or to augment the video panorama vision (Fürstenau et al., 2016) are developed and adapted to various operational needs dependent on the operational context. For instance, an ATCO without any approach radar support would need a very high video resolution to detect traffic at farther distances. Instead, an ATCO who controls traffic movements on the aerodrome maneuvering area would probably need a sufficient Video Frame Rate (in the following referred to Frame Rate (FR), expressed in fps, also referred to video update rate (European Organisation for Civil Aviation Equipment, 2021)) to precisely judge about the velocity of the traffic. In fact, both, resolution and FR are important operational quality parameters but also bandwidth consuming and therefore cost-driving factors in Remote Towers systems. Thus, new Remote Tower implementations aim to optimize these parameters to a better benefit–cost ratio. With this in mind, effects of reduced FR in a Remote Tower context have been questioned in this chapter. Certainly, FR below the critical flicker frequency (CFF) could contribute to a perceived loss of movement fluidity, which might result in perceived loss of video quality. But does lower FR also evoke negative effects, such as reduced ability to detect traffic movements on the displayed video panorama or even cause physiological stress and lower system operability?

This study addresses therefore the following research question: What are the effects of lower FR in a Remote Tower environment on:

(1)   Visual detection performance,
(2)   Physiological stress,
(3)   Perceived video quality,
(4)   Perceived System operability?

This chapter is structured in the following parts: Sect. 2 aims at presenting theoretical background concerning the perception of movement, distortions that can appear during video transmissions and a review of scientific literature about the impact of low frame rates on the operator. Linking these three aspects together allows us to specify the research question and the hypotheses. Section 3 explains the chosen methods and the procedure of the study. Then, in Sect. 4 we will present the obtained results in descriptive and inferential statistics. In Sect. 5 these results will be explained and

discussed based on theoretical findings and the research question. Section 6 draws explicit conclusions by illustrating how the results of the conducted study contribute to research and future Remote Tower implementations.

2 Theorethical Background

A. *Motion perception*

In order to better understand potential impact of reduced FR on humans, we will firstly explain the importance for human beings to perceive motion and, secondly, give some information about fluid motion perception.

(1) *Importance of motion perception for human beings*

The perception of moving objects is a phenomenon that humans take for granted. In fact, since the earliest childhood, a baby's attention is guided towards moving objects (Kellman, 1995). First, motion perception allows humans to estimate the velocity of stimuli and to anticipate "*collision time*". Second, perceiving motion is useful to situate objects in a tridimensional environment (Nakayama, 1985). Other reasons that underline the importance of motion perception consist in distinguishing a stimulus from its background and understanding different textures of objects (Nakayama, 1985). For instance, if a gray airplane is in front of a gray cloud, it might be difficult to distinguish the flying object from its background. A light penetration from a different angle could be perceived when the plane moves. In conclusion, we can state that motion perception permits the observer to get to know more about the details of the environment s/he's in. In order to understand to which extent a movement appears to be fluid, some basics of psychophysics and cinematography are necessary and will be explained in the following section.

(2) *Fluid motion perception*

In psychophysics, psychologists refer to absolute threshold if the minimal intensity necessary to perceive a stimulus is perceived by 50% of the observers (Greene, 2015). Critical Flicker Frequency (CFF) is described as the frequency at which the flickering of a flash is not distinguishable from a constant light source (Wells et al., 2001). This threshold can vary by the luminosity of the discontinuous light (Greene, 2015). According to Kallonatis and Luu (2007a), the sensibility of CFF can also depend on the contrast between the stimulus and its environment. Therefore, the human eye is more sensible to temporal frequencies in high contrast situations between 15 and 20 Hz. The idea of a CFF is also used in cinematography. In cinematographic history, 13 presented images per second were identified as being critical for creating the sensation of fluid movement (Hess, 2015). Concerning the first movies ever produced, 16 frames per second (fps) were not sufficient for showing fluid movements because of the visually perceived intermittent time between each frame. Therefore, cinematographs found a solution by showing the same image two or three times in a

successive manner. In total, this means a presentation of 32 or 48 images per second (Hess, 2015) from which 16 are different. More precisely, movies were presented at 16 fps with a refresh rate of 32 Hz or 48 Hz. It's important not to confuse these two notions. Nowadays, the regular FR in cinemas and TV is either 24 fps or 30 fps (Hess, 2015). FR and refresh rate are two important notions to understand the meaning of human perception of fluid movements in virtual environments. However, perceived fluidity of movements is not the only factor that contributes to an almost perfect presentation of the outer world when it comes to cinematography. Therefore, the next section will treat distortions likely to appear during tele transmissions.

B.   *Reality distortions through tele-transmission*

The human visual system is complex and even though it is theoretically feasible to imitate an operator's "out of the window" visual performance, it is not a necessary condition to provide safe and efficient air traffic control (European Organisation for Civil Aviation Equipment, 2021). Despite the similarities between an optical sensor camera and the human eye, what most cameras represent and what we see with our proper eyes is slightly different. Perceiving the world around us in a stereoscopic manner is already a limit for most conventional cameras that render a monoscopic image. Further, image resolution plays an indispensable role (Bakka, 2017). It allows us to perceive objects from a far distance in a detailed manner. The higher the image resolution, the better we can discriminate stimuli at bigger distances. The human eye has a visual acuity of ca. 1 arc minute (Kallonatis & Luu, 2007b). In other words, from a distance of 1 km, the human eye can discriminate two points with a distance of 28 cm. However, conventional Remote Tower camera systems dispose of a medium image resolution (Schmidt et al., 2016) lower than 1 arc minute. With 2 arc minutes for instance, a camera could only discriminate two pixels with a distance of 56 cm from a distance of 1 km.

Latency or lag can be another distortion appearing in real-time tele-transmission systems caused by different sources (e.g., transmission problems, data conversion problems). They are expressed by a temporal delay between the input of information into a system and the output as a presentation of the information to the operator (Bryson, 1915). When we face different latency times in between of the presented frames, we talk about jitter.

Finally, the presented FR can result in a distortion of reality. By its reduction, the fluid perception of the movement drops as well. As we have seen it in the previous section, 13 fps are judged as being necessary in order to perceive fluid movement. This estimation is not absolutely correct, since the threshold can vary between several parameters, for instance the radial velocity of the perceived object. Another distortion related to FR refers to frequency interferences (Bakka, 2017), like the well-known wagon-wheel-effect. In a Remote Tower environment, this effect can appear wherever periodic movements are faced, e.g., rotor blades of an aircraft or blinking lights. Some blinking lights need time to light up and are only at their maximum of luminance for a few instants. Hence, by reducing FR, the probability to capture an image during the maximum of luminosity diminishes as well. This could be critical particularly at night or more general, in low visibility conditions.

C.   *Review of low FR effects on the operator*

Limited bandwidth made several concerned parties study impacts of low FR on operators (Chen et al., 2005; Schmidt et al., 2016; Stanney et al., 1997). Due to high data transmission costs, researchers have investigated several parameters in order to reduce bandwidth. In the next section, we will present studies that focus on the effects of low FR on performance, psycho-physiological health of operators, as well as on perceived video quality.

(1)   *Effects of low FR on operator visual detection performance*

In the context of a Remote Tower research study, DLR investigated effects of low FR in a real-time tower simulation scenario: 6, 12 and 24 fps were tested in order to evaluate the performance of ATCOs to visually discriminate and predict in real-time if an aircraft was in danger of a runway overrun (runway excursion) due to low braking after touchdown. By observing the corresponding angular speed profiles as visual stimulus participants had to decide as early as possible between in time hold on runway or a runway excursion (two-alternative decision with pushing of corresponding keys). The analysis of the response matrix by means of Bayes inference for conditional error probabilities (risk of unexpected event), and by means of the ROC-(hit(false alarms))—curves (separating discriminability from decision criterion) consistently provided minimum decision error and maximum discriminability for FR > 30 fps (Ellis et al., 2011; Fürstenau et al., 2016), which is also in agreement with (Claypool & Claypool, 2007).

Another study about unmanned ground vehicles and aircraft showed that the performance of detecting obstacles does not decrease by reducing the FR from 30 to 5 fps (Chen et al., 2005). Similar results were obtained in a study about target detection comparing a 2-fps to a 25-fps condition (Garaj et al., 2010). By reducing FR, the participants' performance did not change significantly. As we see, divergent results concerning the impact of lower frame rates on visual performance are observed. One possible explanation is that not only FR is a factor affecting performance. A systematic review about effects of low FR on performance (Chen & Thropp, 2007) concluded that the effect of low FR on performance is above all task-dependent. Moreover, an interaction of FR with image resolution was identified. The authors suggest that the right balance between FR and image resolution can help to perceive depth accurately and hence increase the perception of movement in the areas that are farther away from the observer. They also suggest that performance depends on the participants' characteristics, too.

Thus, experienced participants in virtual environments might be less affected by FR reduction. However, not only performance is an important factor to consider when reducing FR. The operator's well-being is an essential parameter to study before lower FR can be applied. Therefore, the next section will treat impacts of low FR on psycho-physiological health of operators.

(2)   *Effects of low FR on psycho-physiological health of operators*

Effects of low FR on psycho-physiological health have very rarely been evaluated. In a study about several unmanned ground vehicles and aircraft, physiological stress

was tested in terms of cyber sickness in different FR condition (Chen et al., 2005). The used Simulator Sickness Questionnaire (SSQ) has been validated within a sample of 4000 pilots who participated in trainings in different flight simulators (Kennedy et al., 1993). Nowadays, the SSQ is also used for evaluating cyber sickness in other virtual environments (Kolasinski, 1027). In the above-mentioned study (Chen et al., 2005), the effects of low FR were non-significant. Thus, participants did not feel sicker in a simulation at 5 fps than at 30 fps. Another study (Ellis et al., 1999) tested spatial stability in a virtual environment by varying the FR (6, 12 and 20 fps). As a result, more participants felt sick by reducing FR. Reference (Garaj et al., 2010) did not find significant results. Even though participants at 2 fps expressed higher workload and frustration, the expressed psycho-physiological stress was not significantly higher than at 25 fps. Furthermore, adverse health effects associated with low FR do not appear in the occupational disease lists (Bundesgesetzblatt, 2018). Considering past findings, no conclusive results can be found on psycho-physiological health issues to lower FR.

(3) *Effects of low FR on perceived video quality*

Reviewing past research, studies evaluate perceived video quality in different FR conditions. The perceived quality is often evaluated via acceptability and personal preference. In the context of a study concerning the performance in first person ego shooters, a study varied FR (3, 7, 15, 30 and 60 fps) and image resolution (320 × 240 pixels, 512 × 384 pixels & 640 × 480 pixels) (Claypool & Claypool, 2007). Results clearly indicate a significant preference for higher FR and even more for a higher image resolution. Surprising effects have been found in a study that aimed to evaluate the video quality under different FR conditions (6, 10, 12, 15, 18, 20 and 24 fps) and two image resolution conditions (low and high). A significant difference of video acceptability was not found among the different FR conditions. Moreover, participants preferred higher image resolution to higher FR (McCarthy et al., 2004). In a study testing the video acceptability in different FR conditions (5, 10 and 15 fps), it was stated that video acceptability decreases by reducing FR (Apteker et al., 1994). In reference (Masry & Hemami, 2001) the type of motion is stressed: "*The type of motion in a sequence was important when considering the effects of FR on subjective quality*".

To conclude, it is difficult to find an appropriate FR threshold to guarantee the spectator's satisfaction in terms of video quality. As the previous studies already have suggested, it is very likely that not only FR plays a determinant role for acceptance of video quality.

(4) *Effects of low FR on the perceived operability*

Until now, we have presented studies that examined effects of lower FR on performance, operator health and perceived video quality. However, these three parameters seem to be insufficient to evaluate if a Remote Tower system can be operated in a safe and efficient manner. If the user is not convinced of the system operability, errors can emerge by expressed mistrust in the system. In fact, confidence in a system and emerging risks can play a mediator role in the system reliability (Lee & See, 2004).

Therefore, a system can seem reliable to experts but is not if the user does not have a good feeling about it.

D. *Study context and research question*

The general aim shared by all Remote Tower actors is to develop a system that allows remote air traffic control in the best cost-efficient ratio. Regarding this aim, a known limit is bandwidth. Nowadays, data transmission still is expensive and can be a financial threat if resources are not used efficiently. According to Bakka (2017), crucial factors concerning bandwidth are field of view, image resolution, color depth, FR and data compression, which seem to be widely accepted, but opinions diverge largely when it comes to image resolution and FR. Some stakeholders believe that higher FR is preferable to higher image resolution. In fact, they believe that low FR can decrease performance and operator health. However, so far there is no scientific proof that justifies these presumptions. As it has already been expressed in the theoretical part, effects of low FR on performance are likely to be task dependent (Chen & Thropp, 2007) and do not give us clear information about operator health and satisfaction. Yet, impacts of low FR in Remote Tower environments have never been thoroughly tested in an experimental design. On the basis of context analysis and past scientific findings, we will now propose six hypotheses.

E. *Hypotheses*

$H_{1,1}$: By reducing the FR from 15 to 10, 5 fps or 2 fps, the adequate assessment of moving objects by ATCOs decreases.

$H_{0,2}$: By reducing the FR from 15 fps to 10, 5 and 2 fps, the operator's performance in visual detection tasks will not decrease.

$H_{0,3}$: ATCOs' performance in visual tracking tasks does not decrease significantly by reducing the FR from 15 to 10, 5 fps or 2 fps.

$H_{0,4}$: The physiological stress of operators does not increase significantly when FR is reduced from 15 to 10, 5 fps or 2 fps.

$H_{1,5}$: The ATCOs' perception of the video quality will decrease when the FR is reduced from 15 to 10, 5 fps or 2 fps.

$H_{1,6}$: The ATCOs' perceived system's operability will decrease when the FR is reduced from 15 to 10, 5 fps or 2 fps.

3 Methods

A. *Tested variables*

The independent categorical variable is the chosen FR that will be presented in a video at four modalities: 2, 5, 10, and 15 fps. Concerning the dependent variables (DV), the first measures the participants' visual performance in three different dimensions: "Adequate Assessment of Moving Objects" (AAMO), "Visual Detection Tasks" (VDT) and "Visual Tracking Tasks" (VTT). The second DV evaluates

the participants' "physiological stress", the third DV the ATCOs' "perceived video quality", and the fourth DV the ATCOs' "perceived operability of the low FR system".

B.   *Participants*

Seven male ATCOs between 31 and 58 years ($M = 41.7$, $SD = 12.0$) and five pseudo ATCOs (four men, one woman) between 26 and 52 years ($M = 44.0$, $SD = 10.42$) took part in the experiment. Their nationalities were English, German, Hungarian, Norwegian, Romanian, and Swedish. We chose pseudo ATCOs as a non-expert control group in order to control potential motivational bias concerning physiological stress. ATCOs were directly invited by an invitation letter. All ATCOs and pseudo-ATCOs were familiar with the Remote Tower concept.

C.   *Preparation of the study equipment*
(1)   *Video material collection, selection and edition*

For the experiment, the DLR Remote Tower field test platform at Braunschweig-Wolfsburg Airport (BWE) was used. Figure 1. shows the camera sensors on the roof of the DLR building surveying BWE aerodrome (left). On the right, the ATCO working position is shown. The research prototype is operated with 2 arc minute image resolution and a FR of 30 fps.

Several hours of audio and video material have been recorded via the platform, assessed and selected. Firstly, only records which complied with EUROCAE (European Organisation for Civil Aviation Equipment, 2021) standard test conditions were selected. Further, it was checked for broad traffic diversity and other relevant visual occurrences (e.g., flock of birds). This step was supported by four ATCOs. In a third step, the final 30-fps video stream was computed to four content-identical streams with 2, 5, 10 and 15 fps and a length of 80 min each. 2, 5, 10 and 15 fps were chosen as 30 fps is a multiple of them, which helps to avoid the maximum of jitter (European Organisation for Civil Aviation Equipment, 2021). 2 fps as the lowest FR


Fig. 1 DLR Remote Tower field test platform at Braunschweig-Wolfsburg Airport (BWE) (left: 360° Panorama & PTZ camera sensors; right: ATCO Participant working in passive shadow mode at the remote tower module observed and instructed by the experimenter (right))

was chosen since this FR corresponds to the minimum standard of FR tolerated in a Remote Tower environment (European Organisation for Civil Aviation Equipment, 2021). To complete the construction, the video material had to synchronize with the external sound and the radio transmissions. Finally, the jitter was measured in each experimental condition to ensure that it lies below the maximum tolerated value of 0.5 s (Fürstenau & Schmidt, 2016).

(2)　*Construction of the mid-run visual performance evaluation grid*

In a first step, we chronologically listed events that refer to ATC relevant visual tasks and associated them with the visual requirements stated by the interviewed ATCOs and those in the requirements for EUROCAE Remote Tower specifications (European Organisation for Civil Aviation Equipment, 2021). These events were divided into three different categories of questions: the AAMO, VDT, and VTT. As for the AAMO, we mainly took into account the ATCOs' fears of not being able to properly assess the velocity or the stimulus' movement direction. Thus, an exemplary task is to evaluate whether a flashing light can be perceived in a safe and efficient manner. Other tasks include the assessment of flying birds' direction, wind direction and movements of aircraft propellers or human beings at the aerodrome. An exemplary task for VDT consists in detecting an aircraft in the final approach area or in the traffic pattern as soon as possible. Perceiving an aircraft in those positions represents visual requirements according to the interviewed ATCOs. Regarding VTT, the instruction consists in following an aircraft during the take-off phase and hitting a buzzer when it was not noticeable anymore. After classifying all possible tasks, we created and selected a list of possible questions that follow the chronology of occurrences. Some example questions are referred to in Fig. 2.

(3)　*Construction of the post-run questionnaire*
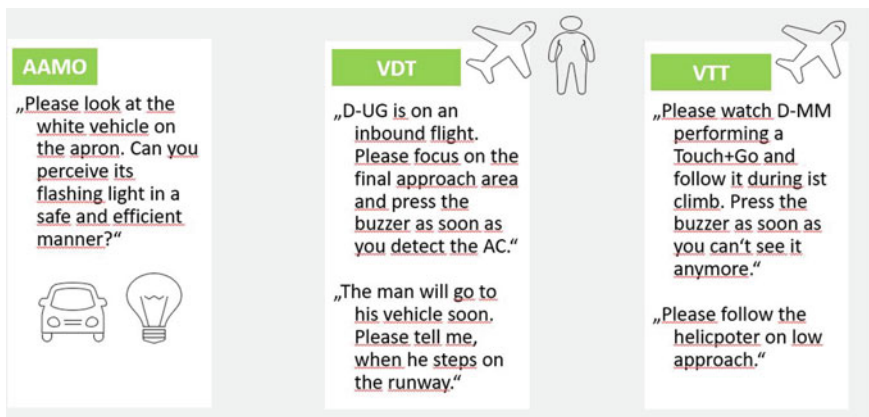
Fig. 2 Example questions referring different detections tasks: Adequate Assessment of Moving Objects (AAMO), Visual Detection Tasks (VDT) and Visual Tracking Tasks (VTT)

To measure the physiological stress of the participants, we concentrated on mentioned symptoms in the interviews of the pretests, such as fatigue, nausea, headache, eye strain or dizziness, which are consistent with the items in the SSQ questionnaire to evaluate cyber sickness (Kennedy et al., 1993). It contains 16 symptoms, whose severity is rated on a three-point Likert scale from 0 "none" to 3 "severe". The symptoms differently load on three overarching factors "Nausea" (*N*), "Oculomotor" (*O*) and "Disorientation" (*D*), which are, beside the total score, to be used for a deeper analysis of the symptoms' origin (see Table 1). The SSQ was instructed to be answered after each test run and in the beginning of the experiment serving as a baseline score to guarantee that the test subjects are free of basic physical stress symptoms. The instruction read: "The following 16 questions are designed to measure your current mental and physiological state. Please indicate how you feel right now by selecting your preferred answer amongst four possible choices. You can only select one answer per item. If you feel uncertain about the meaning of the items, just ask the experimenter."

Table 1 items of the Simulator Sickness Questionnaire (SSQ) and their load on three overarching factors (*N*, *O* & *D*) to be answered by each ATCO after each test run (post run)

Symptom	Nausea (*N*)	Oculomotor (*O*)	Disorientation (*D*)
1. General discomfort	x	x	
2. Fatigue		x	
3. Headache		x	
4. Eyestrain		x	
5. Difficulty focusing		x	x
6. Increased salivation	x		
7. Sweating	x		
8. Nausea	x		x
9. Difficulty concentrating	x	x	
10. Fullness of head			x
11. Blurred vision		x	x
12. Dizzy (eyes open)			x
13. Dizzy (eyes closed)			x
14. Vertigo			x
15. Stomach awareness	x		
16. Burping	x		

The second part of the post-run questionnaire consists in rating the perceived video quality, the perceived operability on a 7-point Likert scale, a request to estimate the FR of the video, which was just observed, and after all test runs to rank all video streams by their preferred frame rate.

(4) *Pretest*

A pseudo-ATCO and two ATCOs participated in the pretest to verify that the scenario did not contain inconsistencies, that the tasks relate to an ATCO's daily routine, and that the questionnaires are comprehensible. They accepted the setting and confirmed that the number of tasks was enough not to be bored and that the variety of tasks corresponds well to the different visual requirements that ATCOs have to face during their daily work.

D. *Experimental Procedure*

The study took place between May 15th 2017 and June 12th 2017. The procedure of the study was structured in two parts. The briefing phase represented the part in which ATCOs were informed and prepared for the actual experiment. The experimental phase corresponded to the video session and the completion of the post-run questionnaire. The written and spoken language was English. The participants were informed that they will see the same video four times at four different FR. They were left unaware of the FR to be tested to avoid potential effects of previously formed attitudes. They were explained that the order of the videos was randomized for methodological reasons and that they had to complete the SSQ questionnaire before the actual experiment to avoid methodological biases. The ATCO's eyes' position was 2.1 m distance from the 56″ HD screens in order to standardize experimental conditions and to guarantee the necessary visual acuity. The experimenter sat at the participant's right side. To randomize the observations, the different FR modalities were ordered in a Latin square. After the last session, the participants answered a supplementary questionnaire in which they gave demographic information about themselves and classified the watched videos in order of preference. Finally, they were asked to give their general opinion on Remote Tower in order to reduce potential motivational biases followed by a general debriefing session.

4 Results

Results are reported descriptively and inferentially. Because of the small sample size of seven, we used non-parametric measures for inferential statistics: A Friedman test represents a non-parametric version for a repeated measure ANOVA. The Wilcoxon test corresponds to a non-parametric alternative for post-hoc comparisons. IBM SPSS Statistics 22 was used for these analyses.

A. *Results of visual performance*

This section refers to the visual performance evaluated by tasks regarding the adequate assessment of moving objects, visual detection tasks (timed and non-timed) and visual tracking tasks.

(1) *Adequate assessment of moving objects*

In order to evaluate AAMO, the ATCOs' answers were coded as "1" when the movement is perceived "safe and efficient" and as "0" when the movement was perceived as "neither safe, nor efficient". It was observed that the movement of five objects was perceived as being "safe and efficient" by all ATCOs in each of the four FR modalities. These objects correspond to the propeller of three different aircraft on the apron, to a flag from which the ATCOs had to assess the wind direction and the direction of a flock of birds. Concerning the flock of birds, the ATCOs added that it was easy to identify the objects as birds and to deduce their direction.

    The other category of objects corresponds to the flashing lights of five vehicles: a fuel truck, a black airport vehicle, and three Follow-Me (FM) vehicles. Concerning the fuel truck and the black vehicle, we observed that most ATCOs judged the visibility of the flashing light as being perceivable safe and efficient in the 5, 10 and 15 fps conditions but not in the 2-fps condition. This tendency particularly appeared with the FM vehicles' flashing lights. In the 2-fps condition, a safe and efficient perception of the FM vehicles is only admitted one time out of 21 instances over all 7 ATCOs (see Fig. 3).

    By comparing all means of all flashing lights instances (35 in total per FR test condition), we observe that the flashing lights are perceived as being least visible in a safe and efficient manner in the 2-fps condition ($M = 0.17$, $N = 7$, $SD = 0.21$), followed by the 10-fps condition ($M = 0.77$, $N = 7$, $SD = 0.34$) and the 5-fps
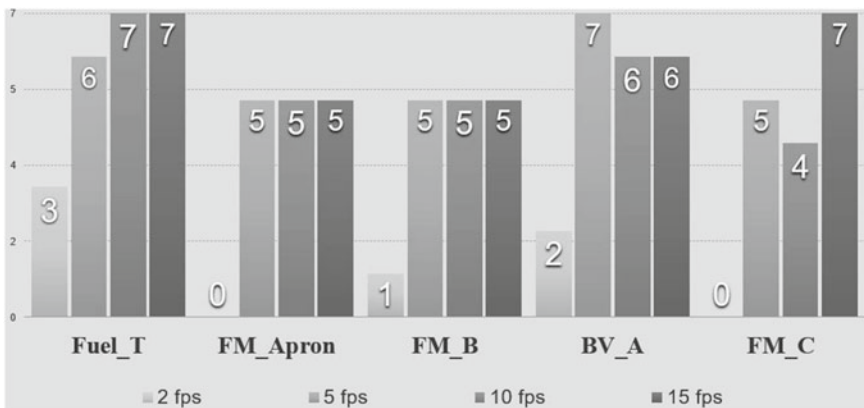
Fig. 3 Absolute frequency of ATCOs' perception of flashing light as safe and efficient of five vehicles: a fuel truck (Fuel_T), a black airport vehicle (BV_A), and three follow me (FM_Apron, FM_B & FM_C) vehicles ($N = 7$)

condition ($M = 0.8$, $N = 7$, $SD = 0.31$). In the 15-fps condition, flashing lights were perceived as the being most visible in a safe and efficient manner ($M = 0.85$, $N = 7$, $SD = 0.25$). A chi-square test supports this tendency: The perception of flashing lights decreases significantly when the FR drops from 15 to 10, 5 and 2 fps ($\chi^2_{(df = 3, N = 7)} = 1$, $p < 0.01$). But these results can only be found for flashing lights.

Thus $H_{1,1}$ is only partially assumed, i.e., by reducing the FR from 15 to 10 fps, 5 fps or 2 fps, only the adequate assessment of flashing lights decreases, all others remain unaffected.

(2) *Visual Detection Tasks*

The difference of each FR conditions mean detection time from the mean of all detection times, show that ATCOs take on average less time detecting an aircraft in the approach area at 10 fps ($M = -1.4$, $N = 7$, $SD = 6.33$) than at 15 fps ($M = -0.04$, $N = 7$, $SD = 5.85$), at 2 fps ($M = 0.21$, $N = 7$, $SD = 5.36$) or at 5 fps ($M = 1.5$, $N = 7$, $SD = 4.32$). However, a Friedman test did not show significant difference between ATCOs' reaction time at the four FR conditions ($\chi^2_{(df = 3, N = 7)} = 2.14$, $p = 0.54$). Thus, the reduction of FR does not appear to decrease the ATCOs' performance to detect an aircraft in the final approach area which supports our $H_{0,2}$ to retain the H_0.

Furthermore, all aircraft in different traffic pattern positions, as well as all human beings on the movement and maneuvering area were perceived by ATCOs in each FR condition. In addition, some ATCOs add that the jerky movements perceived in the 2 and 5 fps condition helped them to detect the aircraft in the very far distance quicker. According to them, the jerky movements of small pixel bunches cause a blinking effect and thus attract more attention than an aircraft that moves more smoothly at 10 fps or 15 fps.

(3) *Visual Tracking Tasks*

The measured times of VTT, again centered on the mean of each of the four FR conditions, indicate that on average, ATCOs could visually track departing aircraft longer at 15 fps ($M = 1.48$, $N = 7$, $SD = 3.95$) than at 2 fps ($M = -0.05$, $N = 7$, $SD = 4.98$) or at 5 fps ($M = -0.35$, $N = 7$, $SD = 8.44$), worst at 10 fps ($M = -1.06$, $N = 7$, $SD = 4.75$). Again, the Friedman test could not reveal significant difference of ATCOs' performance to visually track aircraft in all tested four different FR ($\chi^2_{(df = 3, N = 7)}$, $p = 0.62$), which supports our $H_{0,3}$ to retain the H_0: ATCOs' performance in visual tracking tasks does not decrease significantly by reducing the FR from 15 to 10 fps, 5 fps or 2 fps.

B. *Results concerning physiological stress*

Before the experiment and after each test run, participants answered the 16 SSQ-items on a Likert scale ranging from "0 = none", "1 = slight", "2 = moderate" to "3 = severe". The 16 items load on three subscales: "Nausea" (N), "Oculomotor" (O) and "Disorientation" (D) and on a total score (TS). The subscale's belonging items are summed up and weighted differently to make them comparable to the same standard deviation of 15: $N = \Sigma N \times 9.54$; $O = \Sigma O \times 7.58$; $D = \Sigma D \times 13.92$ and

Table 2 Categorisation of Symptoms based on SSQ Scores (Stanney et al., 1997)

SSQ score	Categorisation
0	No symptoms
<5	Negligible symptoms
5–10	Minimal symptoms
10–15	Significant symptoms
15–20	Symptoms are a concern
>20	A bad simulator

$TS = (N + O + D) \times 3.74$ (Kennedy et al., 1993). This weighting also results in different maxima: $N = 200.34$; $O = 159.18$; $D = 192.32$ and $TS = 253.62$. Stanney et al. (1993, 1997) suggests to use the median mean value with sample sizes smaller than 50, which is adopted in this analysis. They further proposed a categorization system based on their findings with large sample sizes of up to 6000 that allows deducing a normalized categorization of the median with respect to the criticality of the symptoms, which is referred to in Table 2.

Before the actual experiment phase, we checked if all test subjects were free of basic symptoms in order to avoid a manifestation of symptoms during the experimental runs. It turned out that 4 of 7 ATCOs showed symptoms, but with a maximum of 7,48 they were judged as "minimum symptoms" (cf. Table 2). Further we checked whether the results of the experimental ATCO group differ significantly from the pseudo-ATCO control group to exclude systematically effecting variance in terms of possible motivational biases on behalf of the ATCOs caused by their general attitude towards the Remote Tower concept. A Mann–Whitney-U-Test for independent samples did not show any significant differences on a significance level of 0.05. Thus, the expert group seems to be unaffected of systematic biases. Figure 4. shows the median values for the seven ATCOs over all four test conditions and for the "before" (base) measurement broken down to the three subscales "Nausea" (N), "Oculomotor" (O), "Disorientation" (D) and to the total score (TS). The highest scores with a maximum of 7.58 are reached with the scale "Oculomotor" and is equally observed in all four test conditions. The TS score shows its maximum values of 7.48 with the 2fps test condition and a minimum of 3.74 in all other test conditions and "before" (base) assessment. One test subject outlier has been observed with constantly high values in all 4 test conditions which causes relatively widely ranged 25–75‰ whiskers (cf. Fig. 4), particularly in the 2fps test condition. Summarizing, all median values are smaller than 10, respectively even smaller than 5. In accordance to Table 2 it is stated that all scores reveal only negligible or minimal symptoms.

To judge about hypothesis $H_{0,4}$ a nonparametric Friedman two-way analysis of variance for related samples was conducted but no significant difference between the four test conditions could been found, neither for the TS ($p = 0.22$), nor for the subscales "Nausea" ($p = 0.09$), "Oculomotor" ($p = 0.33$), and "Disorientation" ($p = 0.19$) As postulated, $H_{0,4}$ is to be retained: The psychological stress of operators did not increase significantly when FR is reduced from 15 to 10, 5 fps or 2 fps.
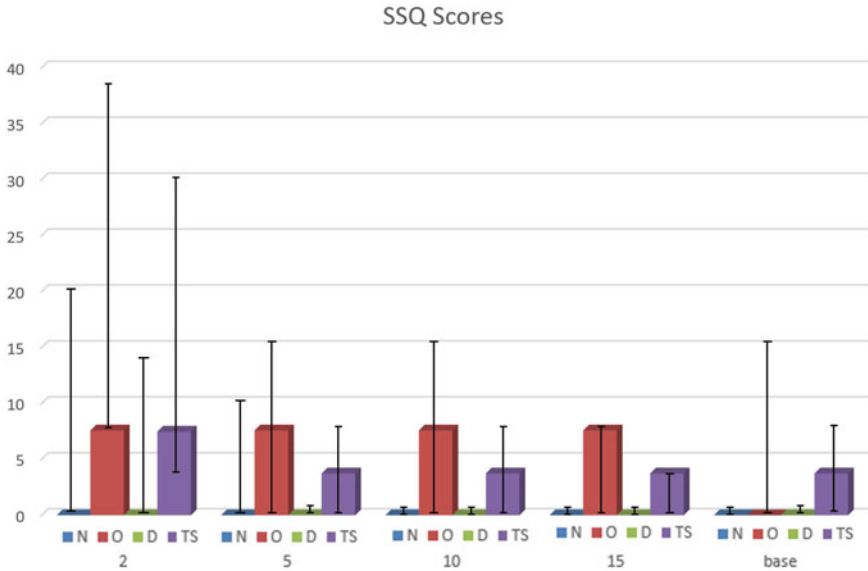
SSQ Scores

Fig. 4 SSQ Scores: Median for baseline and four test conditions (2, 5, 10 and 15 fps) (whisker range from 25 to 75‰, $n = 7$)

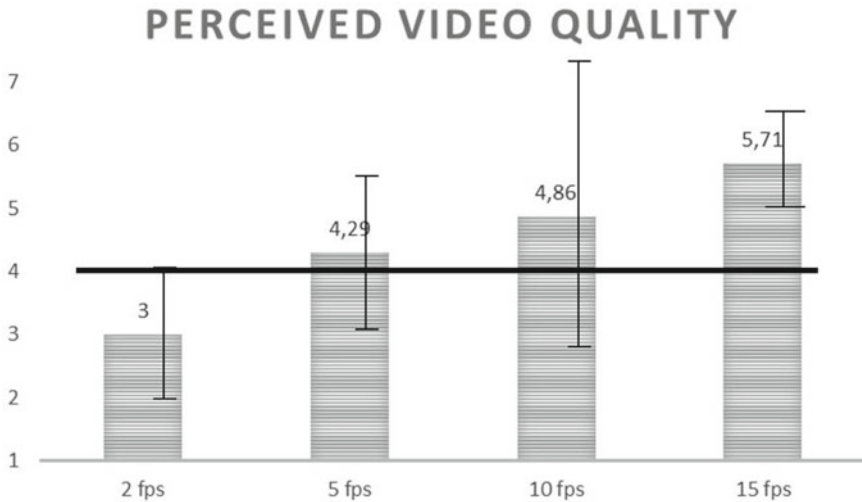### C. *Results concerning the perceived video quality*

Via a 7-point Likert scale ranging from "1 = totally unacceptable", "2 = unacceptable", "3 = slightly unacceptable", "4 = neutral", "5 = slightly acceptable", "6 = acceptable" to "7 = perfectly acceptable", ATCOs perceived the video quality as being more acceptable at 15 fps ($M = 5.71$, $N = 7$, $SD = 1.25$, Min $= 3$, Max $= 7$) than at 10 fps ($M = 4.86$, $N = 7$, $SD = 2.67$, Min $= 1$, Max $= 7$), 5 fps ($M = 4.29$, $N = 7$, $SD = 1.8$, Min $= 1$, Max $= 6$) or at 2 fps ($M = 3$, $N = 7$, $SD = 1.53$, Min $= 1$, Max $= 5$).

The more the FR is reduced, the less ATCOs judge the quality of the video as being acceptable. In the 2-fps condition, the quality was even rated below the average "neutral" (Fig. 5.) This tendency is supported by a Friedman test that revealed significant difference ($\chi^2_{(df = 3, N = 7)} = 12.05$, $p < 0.01$). As postulated in $H_{1,5}$, the perceived video quality in terms of FR decreased with the reduction of the FR.

D. *Results concerning the perceived operability of a low FR system*

On a 7-point Likert scale from "1 = totally disagree", "2 = disagree", "3 = somewhat disagree", "4 = neither agree nor disagree", "5 = somewhat agree", "6 = agree" to "7 = strongly agree", the ATCOs should answer the following statement: "I would be able to control the air traffic with the given FR.". The perceived operability increased with the increase of FR. Thus, the system operability was perceived least at 2 fps ($M = 2.86$, $N = 7$, $SD = 1.57$, Min $= 1$, Max $= 5$). It increases over-averaged with 5 fps ($M = 4.14$, $N = 7$, $SD = 1.87$, Min $= 1$, Max $= 6$), 10 fps ($M = 4.86$, $N =$
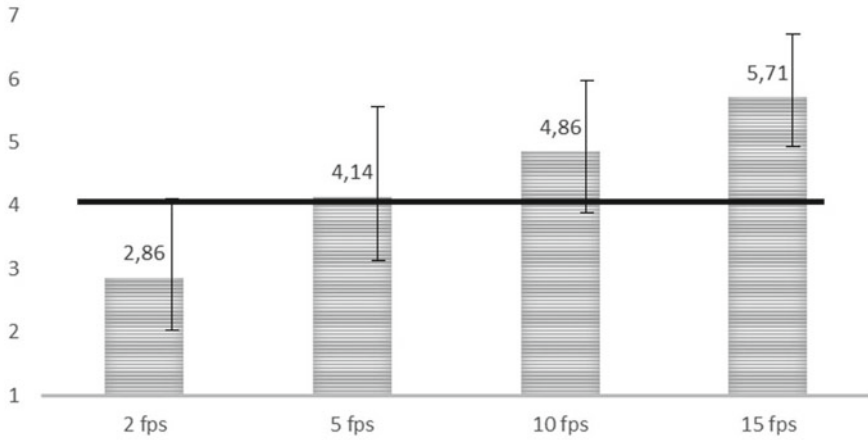
Fig. 5 Perceived video quality rated via a 7-point Likert scale ranging from "1 = totally unaccept-able", "2 = unacceptable", "3 = slightly unacceptable", "4 = neutral", "5 = slightly acceptable", "6 = acceptable" to "7 = perfectly acceptable" at four different frame rate test conditions (2, 5, 10, and 15 fps) ($N = 7$)

7, SD = 1.57, Min = 1, Max = 7) and finally with 15 fps ($M = 5.71$, $N = 7$, SD = 1.25, Min = 3, Max = 7). A Friedman test revealed this difference as significant ($\chi^2_{(df = 3, N = 7)} = 12.68$, $p < 0.01$). Even though only the 2-fps condition is judged below acceptable, $H_{1,6}$ is to be assumed: The lower the FR, the less ATCOs consider the system as being operable (see Fig. 6).

Furthermore, after each test run the ATCOs were asked to estimate the FR of the just watched video. Surprisingly, ATCOs always believed that the FR is superior to what it actually was: Answers after the 2 fps conditions referred to 3.14 fps by average, 5–6.29 fps, 10–15.29 fps, and 15–20.14 fps. By average they overjudged the FR by 53.9% (see Fig. 7).

Post exercise, after having seen all the four video streams, the ATCOs were requested to rank the video streams according to their preference regarding the perceived frame rate. Without surprise, the 15-fps video was ranked first (6 times rank 1 and 1 times rank 2), followed by 10-fps video (1 times rank 1; 3 × rank 2; 1 × rank 3 and 2 × rank 4), 5fps (2 × rank 2 and 5 times rank 2), and the 2-fps test condition (1 times rank 2; 1 times rank 3 and 5 times rank 4) (see Fig. 8).

PERCEIVED OPERABILITY

Fig. 6 Perceived System Operability rated via a 7-point Likert scale ranging from "1 = totally unacceptable", "2 = unacceptable", "3 = slightly unacceptable", "4 = neutral", "5 = slightly acceptable", "6 = acceptable" to "7 = perfectly acceptable" at four different frame rate test conditions (2, 5, 10, and 15 fps)
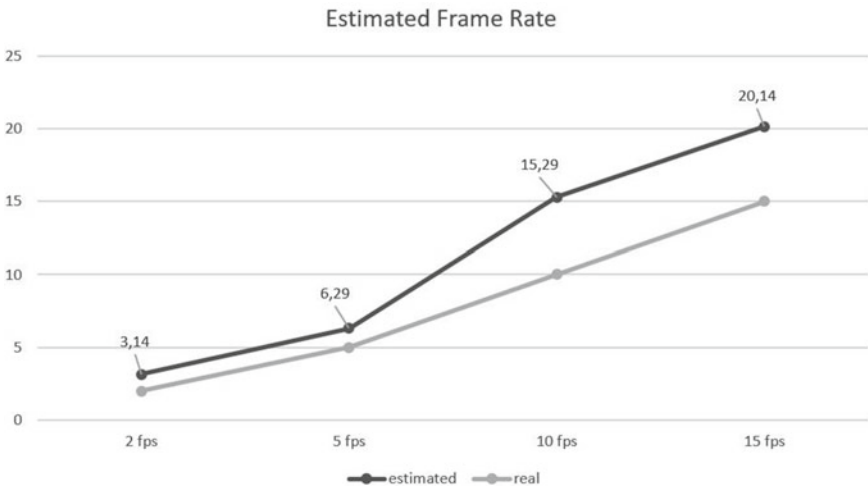
Fig. 7 Post run assessment of ATCOs' estimated fame rate by the, not knowing about the real presented frame rate
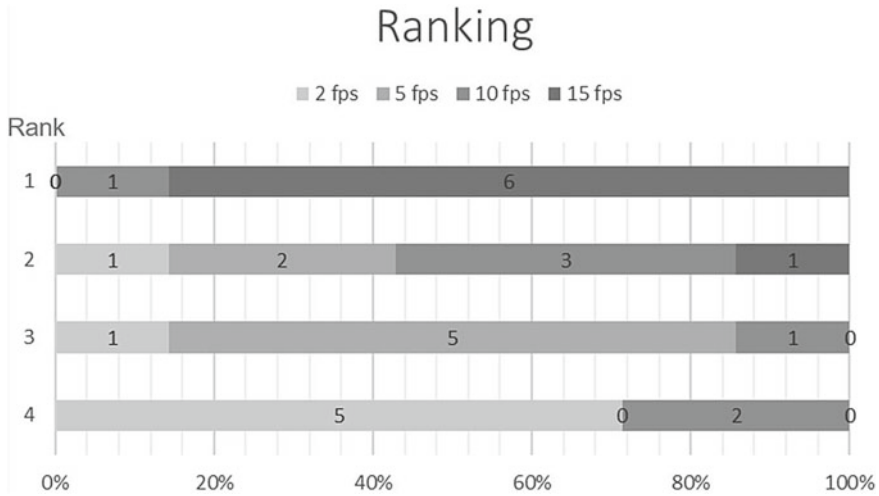
Fig. 8 Post exercise assessment of the preferred ranking of the four video streams referring ATCOs' different frame rate experience

5   Discussion

In line with past research, the results of this study suggest that the effects of lower FR on the performance of an adequate assessment of moving object tasks are multi-layered and cannot be judged generically. Surprisingly, all propeller movements, the wind flag and the flock of birds, as well as the movement of human beings on the aerodrome were perceived by all ATCOs in all four FR conditions (2, 5, 10, and 15 fps) in a safe and efficient manner. Most ATCOs commented that the rapid disappearance of the bird flock with an increasing distance made them worry much more than their jerky movement, which refers rather to an image resolution problem than to a lower FR. Concerning the flashing lights, most ATCOs judge flashing lights to be perceivable safely and efficiently down to 5 fps but when further reduced down to 2 fps the capturing of the rotating beacon at its full brightness decreased and the perception was no longer perceived as being safe and efficient by the majority of ATCOs. Those negative effects in 2 fps were expected and could be covered by using flashing lights with obscure/luminous phases that interfere less with the chosen FR.

As postulated, with respect to the performance in visual detection tasks, inferential statistics do not indicate a significant difference between the four FR conditions. Apart from the impression that aircraft seem to move jerkier at lower FR, especially when they have a high radial velocity, FR does not seem to play an essential role in detection tasks. In particular planes in the final approach or departure area do not have great radial velocity at all. ATCOs therefore perceive only a point that grows bigger when the plane approaches or shrinks at departure. The concern of not visually detecting an approaching or departing aircraft or an aircraft right downwind due to lower FR seems therefore be unjustified. In addition, it seems more logical

to detect an approaching aircraft earlier or to see an aircraft leaving the aerodrome longer by increasing the image resolution instead of higher FR. Moreover, even if the movement seemed jerky at times, several ATCOs noticed that the "jumpy" aircraft attracted their attention in certain situations.

The main purpose of this experimental study was to assess potential negative effects of lower FR on physiological stress. After each run, physiological stress was self-assessed via the SSQ. As presumed, the inferential results show that no negative effects of physiological stress could be measured. All SSQ scores did not show significant symptoms. Only in the 2 fps condition the severity of the symptoms increased slightly for some ATCOs, but far from any significance. One could argue that these findings are biased by very positive or negative beliefs or attitudes towards lower FR system. This potential side effect could be mitigated by using a pseudo-ATCO control group who performed the entire experiment but had a neutral attitude towards lower FR since they were not involved in the Remote Tower business: Both groups did not significantly differ in their SSQ scores. Thus, a systematical effect of bias for the experimental group could be excluded. For the correct interpretation of these results, it is also important to note that the study was dealing with a small but expert sample (Etikan & Bala, 2017). In other words, they share some personality characteristics and very specific professional skills, as well as specialized selection- and education criteria. Thus, it is very likely to transfer the results found in the inferential statistics to other ATCOs. For an implementation of a Remote Tower with a medium image resolution and low FR from 2 to 15 fps, it can be stated that effects expressed by physiological stress will most likely not appear.

As expected, the perceived video quality decreased significantly with the reduction of FR. These results are not that surprising since air traffic control requires high visual performance and reducing the FR is an obvious loss in terms of video quality. Therefore, this obvious loss of video quality could be compensated by an increase of image resolution in a real Remote Tower implementation setup. Since lower FR seem not to impair detection performance nor to induce physiological stress, this trade-off between FR and image resolution seems to be a valid approach to keep bandwidth consumptions low and better adapt the visual presentation to the air traffic service operators' task: For instance, for detecting small aircraft in a far-view distance, high image resolution is needed, and FR might not be that important. On the other hand, assessing the velocity of aircraft in a near-view camera sensor distance on the taxiways or apron causing a high radial velocity, higher FR would be essential and image resolution would not play such a significant role. As stated before, this compensation approach could not be realized in this experimental setting, but it can be assumed that the ATCOs' perceived video quality would have been more balanced over the different FR conditions if have done so.

Similar to the results of perceived video quality are the ones concerning the perceived system operability. By no surprise, also a significant difference between the four FR conditions was found. The average of ATCOs "disagree" or "somewhat disagree" about thinking to be able to handle air traffic at 2 fps. At 5 fps and 10 fps, ATCOs expressed to "slightly agree" being able to manage air traffic and at 15 fps, they expressed to "agree". Like already stated above, the experimental setting

neglected compensation in terms of image resolution which would probably have balanced the ATCOs' attitude as well.

To conclude the discussion on our findings, we can affirm that according to our results, a system at lower FR is justifiable at least starting from 5 fps. Thus, between 5 and 15 fps, the air controllers' visual performance is maintained at the same level. If one wants to set up a lower FR system, one should pay particularly attention to the used flashing lights at the aerodrome which do not interfere with the FR.

Concerning physiological stress, we did not find a significant increase of the scores when the FR is reduced from 15 to 2 fps. However, the comparison of the medians and percentiles in the descriptive statistics reveals more variance under 2 fps. To avoid physiological stress at a system similar to the one at BWE, we recommend rather 5, 10 fps or 15 fps.

With respect to the perceived video quality, the ATCOs preferred higher FR to lower FR. Their general attitude was mostly against 2 fps. In summary, if one wants to operate Remote Tower at a low FR, it is important to develop convincing strategies to increase the tolerance towards low FR. From a psychological point of view, it is not advisable to put ATCOs in front of a 2-fps system hoping that they will accept it. The user-centered approach teaches us how important it is for users to experience positive emotions to raise acceptability for a new product (Richter & Flückiger, 2016). Once the video quality of the low FR system is accepted by the ATCOs, the fear of getting sick could be taken away from them and self-efficiency for performance could be perceived more. By consequent, it is likely that the attitude towards the perceived system operability is expressed more positively.

Further operational simulation and field trials with the operator in the loop are recommended to increase confidence in low FR systems and to gain additional feedback from ATCOs to develop bests designed Remote Tower solutions for the given operational environment.

6   Conclusion

The optimal FR in Remote Tower environments is debated amongst many actors of the Remote Tower community: It must not be too low to endanger ATCOs acceptance, but also not be too high to increase the consumption of bandwidth or to compromise other parameters like image resolution. The results of this study can mitigate the concerns regarding lower FR settings. The major conclusion of this study is that the visual performance and physiological stress were not significantly affected by lower FR in between of 15 down to 5fps. In particular, these findings will allow more degrees of freedom in the design process of a Remote Tower implementation to best adapt a local solution to their operational environment. In future research, it remains to be studied how a trade-off between lower FR and compensation by higher image resolution would be judged by the ATCOs.

References

Apteker, R. T., Fisher, A. A., Kisimov, V. S., & Neishlos, H. (1994). *Distributed multimedia: User perception and dynamic QoS. In Conference of International Symposiom on Electronic Imageing: Science and Technology* (pp. 226–234). San José.

Bakka, B. O. (2017). Remote tower optical presentation. [Power Point Slides], Kjeller: Kongsberg (unpublished).

Bundesgesetzblatt. *BGBl*: *List of occupational diseases, 3rd proclamation for amendment of occupational diseases.* Retrieved March 13, 2018, from http://www.theaustralian.com.au/business/aviation/regional-airports-under-strain-acil-aallen-report-for-aaa/news-story/e95d864f60135d044b5212865ccabae4.

Bryson, S. (1993). Effects of lag and FR on various tracking tasks. In *Paper Presented at SPIE 1915, Stereoscopic Displays and Applications III* (pp. 155–156). San José, USA.

Chen, J. Y. C., & Thropp, J. E. (2007). Review of low FR effects on human performance. *IEEE Transactions on Systems Man and Cybernetics—Part A Systems and Humans ,37*(6), 1063–1070.

Chen, J. Y. C., Durlach, P. J., Sloan, J. A., & Bowens, L. D. (2005). *Robotic operator performance in simulated reconnaissance missions.* Technical report ARL-TR-3628, Army Research Lab Aberdeen Proving Ground Md Human Research And Engineering Directorate, Maryland.

Claypool, K. T., & Claypool, M. (2007). On FR XE "Frame rate" and player performance in first person shooter games. *Springer Multimedia Systems Journal (MMSJ), 13*(1), 3–17.

Ellis, S. R., Adelstein, B. D., Baumeler, S., Jense, G. J., & Jacoby, R. H. (1999). Sensor spatial distortion, visual latency, and update rate effects on 3D tracking in virtual environments. In *IEEE Conference on Virtual Reality* (pp. 239–265). Houston, Texas.

Ellis, S. R., Fürstenau, N., Mittendorf, M. (2011). Frame rate effects on visual discrimination of landing aircraft deceleration: implications for virtual tower design and speed perception. In *Proceedings Human Factors and Ergonomics Society, 55th Annual Meeting* (pp. 71–75). Las Vegas, NV USA.

Etikan, I., & Bala, K. (2017). Sampling and sampling methods. biometrics & biostatistics. *Biometrics Biostatistics International Journal, 5*(6), 00148.

European Organisation for Civil Aviation Equipment. (2021). *EUROCAE*: *Minimum aviation system performance specification for remote tower optical systems.* ED-240A change 1, EUROCAE, Saint-Denis.

Fürstenau, N., Rudolph, M., Schmidt, M., Werther, B., Hetzheim, H., Halle, W., & Tuchscheerer, W. (2008). Flugverkehr-Leiteinrichtung (Virtueller Tower). European Patent EP1791364, application date 2005.

Fürstenau, N. (2016a). Introduction and overview. In N. Fürstenau (Ed.), *Virtual and remote control tower* (p. 6). Cham (ZG): Springer International Publishing Switzerland.

Fürstenau, N. (2016b). Preface. In N. Fürstenau (Ed.), *Virtual and remote control tower* (p. xii). Cham (ZG): Springer International Publishing Switzerland.

Fürstenau, N. (2016c) Introduction and overview. In N. Fürstenau (Ed.), *Virtual and remote control tower* (pp. 14–15). Cham (ZG): Springer International Publishing Switzerland.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., & Halle, W. (2016). In N. Fürstenau (Ed.), Virtual and remote control tower (p. 17). Cham (ZG): Springer International Publishing Switzerland.

Fürstenau, N., Mittendorf, M., & Ellis, S. R. (2016). Videopanorama FR requirements derived from visual discrimination of decerleration during simulated aircraft landing. In N. Fürstenau (Ed.), *Virtual and remote control tower* (pp. 115–137), Cham (ZG): Springer International Publishing Switzerland.

Fürstenau, N., & Schmidt, M. (2016). Remote tower experimental system with augmented vision Videopanorma. In Fürstenau N (Ed.), *Virtual and remote control tower* (p. 180), Cham (ZG): Springer International Publishing Switzerland.

Garaj, V., Hunaiti, Z., & Balachandran, W. (2010). Using remote vision: The effects of video image FR XE "frame rate" on visual object recognition performance. *IEEE Transactions on Systems, Man and Cybernetics Society, 40*(4), 698–707.

Greene, E. (2015). Evaluating letter recognition, flicker fusion and the Talbot-Plateau law using mcrosecond-duration flashes. *PLoS One, 10*(4).

Hagendorf, H., Krummenacher, J., Müller, H.-J., & Schubert, T. (2011). Psychophysics. In H. Hagendorf, J. Krummenacher, H.-J. Müller & T. Schubert (Eds.), *Allgemeine Psychologie für Bachelor: Wahrnehmung und Aufmerksamkeit* (pp. 43–45). Berlin Heidelberg: Springer Verlag.

Hess, J. (2015). *The history of FR for film.* Retrieved March 03, 2013, from https://www.youtube.com/watch?v=mjYjFEp9Yx0.

Kallonatis, M., & Luu, C. (2007a). Temporal resolution. In H. Kolb & E. R. Nelson (Eds), *The organization of the retina and visual system.* Salt Lake City (UT): University of Utah Health Science Center.

Kallonatis, M., & Luu, C. (2007b). Visual acuity. In H. Kolb & E. R. Nelson (Eds), *The organization of the retina and visual system.* Salt Lake City (UT): University of Utah Health Science Center.

Kellman, P. J. (1995) Ontogenesis of space and motion perception. In W. Epstein & S. Rogers (Eds), *Handbook of perception and cognition* (Vol. 5 p. 394). New York: Academic Press.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220.

Kolasinski, E. M. (1995). *Simulator sickness in virtual environments* (Technical Report 1027: Army Project Number 2O262785A791—Education and Training Technology). Alexandria, Virginia: United States Army Research Institute.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80.

Masry, M., & Hemami, S. S. (2001). An analysis of subjective quality in low bit rate video. In *Paper presented at International Conference on Image Processing* (pp. 465–468). Thessaloniki, Greece.

McCarthy, J. D., Sasse, A. M., & Miras, D. (2004). Sharp or smooth? Comparing effects of quantization versus FR for streamed video. In *Paper presented at Conference on Human Factors in Computing Systems.* Vienna, Austria.

Nakayama, K. (1985). Biological image motion processing: A review. *Vision, 25*(5), 625–660.

Richter, M., & Flückiger, M. D. (2016). Usability and UX. In M. Richter & M. D. Flückinger (Eds.), *Usability and UX kompakt* (p. 12). Wiesbaden: Springer Vieweg.

SAAB. *Remote tower revolutionises air traffic management*. SAAB. Retrieved March 13, 2018, from http://saabgroup.com/Media/stories/stories-listing/2017-02/remote-tower-revolutionises-air-traffic-management/.

Schmidt, M., Rudolph, M., & Fürstenau, N. (2016). Remote tower prototype and automation perspective. In N. Fürstenau (Ed.), *Virtual and remote control tower* (p. 180). Cham (ZG): Springer International Publishing Switzerland.

Schwartz, S. (2004). Motion perception. In S. Schwartz (Ed.), *Visual perception: A clinical orientation* (pp. 219–220). New York City: McGraw Hill–Education.

Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997). Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 1138–1142).

Tripathi, A., & Claypool, M. (2002). Improving multimedia streaming with content-aware video scaling. In *Workshop on Intelligent Multimedia Computing and Networking (IMMCN).* Durham, NC, USA.

Wells, E. F., Bernstein, G. M., Scott, B. W., Bennett, P. J., & Mendelson, J. R. (2001). Critical flicker frequency responses in visual cortex. *Experimental Brain Research, 139*(1), 106–110.

Advanced and Multiple RTO: Development, Validation, and Implementation

The Advanced Remote Tower System and Its Validation

F. J. van Schaik, J. J. M. Roessingh, J. Bengtsson, G. Lindqvist, and K. Fält

Abstract The Advanced Remote Tower project (ART) studied enhancements to an existing LFV prototype facility (ROT) for a single airport remotely operated tower: projection on a 360 degrees panorama screen, adding synthesized geographic information and meteorological information, video tracking, fusion of video and radar tracks, labelling, visibility enhancement and surveillance operations with a remotely controlled Pan Tilt Zoom camera. The ART functions have been embedded in the existing Swedish test facility for single airport remote tower operations in Malmö airport Sturup observing Ängelholm traffic about 100 km to the North. These functions were tuned and validated during tests with 15 operational tower air traffic controllers. Emphasis was on the traffic and situation awareness of the tower controllers using remote cameras and a projection system for safe operational tower control, replacing direct view on the airport and its traffic. The validation results give valuable information for further development and operational application even outside the Remote Tower application area.

1 Introduction

The Advanced Remote Tower (ART, 2006) project studied from 2008 to early 2010 the concept of remotely operated Air Traffic Control (ATC) towers and supporting technologies in order to enhance regularity during low visibility operations and to substantially decrease the ATC related costs at airports. The ART enhancements are

F. J. van Schaik · J. J. M. Roessingh (✉)
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

J. Bengtsson · G. Lindqvist
Swedish Air Navigation Service Provider LFV, Stockholm and Malmö, Sweden

K. Fält
Saab Systems, Järfälla, Sweden

prototype functions with different level of maturity. They are supposed to be good candidates for application in remote tower control. This contribution is an extension of the paper presented on the IFAC HMS 2010 conference (2010)

ART was co-funded by the European Commission (Directorate General for Transport and Energy, TREN/07/FP6AE/S07.73580/037179). Partners in the ART project were: Saab (Project Coordination and system integrator), the Swedish Air navigation Service Provider LFV (Operational input and hosting the ART trials and ART facilities), the National Aerospace Laboratory of the Netherlands NLR (Validation and Safety Assessment), LYYN Sweden (Visibility Enhancement Technology VET) and Equipe Ltd. UK (projection facility).

The purpose of ART was to explore the concept of remotely operated towers and to prototype and validate additional sensors and the Human Machine Interface (HMI) that were supposed to enhance the air traffic controllers' situational awareness at reduced visibility conditions due to weather and darkness. ART evaluated promising new technologies, as well as technologies of today, applied and presented in an innovative and more efficient manner. The enhanced situational awareness was one of the main prerequisites for enhanced regularity at the aerodrome, which has proven to be one of the bottlenecks in today's Air Traffic Management system (ATM).

A cost benefit analysis (LFV-ROT, 2008) regarding remotely operated towers had been performed by the LFV Group. It showed substantial economic benefits compared to traditional ATC operations at airports. These benefits for the Air Navigation Service Provider (ANSP) will in turn reduce the cost for airline operators and travellers.

The concept and technology were tested in low-density areas in order to explore the applicability in medium and high-density traffic areas. The ART concept was in turn one of the bricks in the future concept of highly automated ATM at airports.

The concept of ART will also have spin-off effects in the area of training and investigation after incidents and accidents. ART opened the possibility to not only use recorded voice communication but to reproduce the course of events with audio and video of the controllers' situation.

Major deliverables were the ART concept of operations, system design, incorporation and adaptation of sensors and an ART demonstrator on a single low-density airport in Sweden with the possibility to explore the concept at any low to medium density airport. The associated reports can be found in TRIP-ART (2010).

The following steps have been made to achieve these objectives: Design and construction of a remote tower cab, evaluation by end-users of controller workload and situational awareness, evaluation of operational benefits with new possibilities to present information, identification of vital parameters for remote airport operations and evaluation of technical and operational safety issues.

Remote tower concepts were rather unexplored when the ART study was performed. LFV pioneered the hardware aspects for Remotely Operated Towers (ROT) in 2006. Brinton and Atkins (2006), provided a requirements analysis approach for remote airport traffic services. The German Aerospace Center DLR was performing remote airport tower operation research (Fürstenau, 2007) in a national program. US activity could be found in Ellis (2006). The ART project enabled NLR

Fig. 1 Part of the panorama screen and one of the video cameras

to perform a parallel study on detection and recognition by Tower Controllers see (IFAC, 2010) and chapter "Remotely-Operated AFIS in Japan" of this book.

Next sections explain the ART functions, the test and validation program, the results and the analysis and recommendations.

2 ART Functions

The ART project prototyped the following enhancements for Remote Tower Control:

2.1 360° Circular Panorama Display

Nine video cameras were mounted on top of the real tower to observe the total airport and Control Zone (CTZ). Images were projected on a circular projection screen (9 times 42° including overlap between projected images, 6 m diameter, 1360 × 1024 pixel resolution per projected camera image, 20–30 frames per second, Fig. 1).

2.2 Visibility Enhancement Technology

A sizeable part of a projected image could be improved by a digital real time Visibility Enhancement Technology (VET), see Fig. 2.

Fig. 2 Visibility enhancement for a part if the image

2.3 Presentation of Airport and Geographic Information

Synthetic contour lines could be activated enhancing the runway and taxiway edges
in low visibility conditions, see Fig. 3.
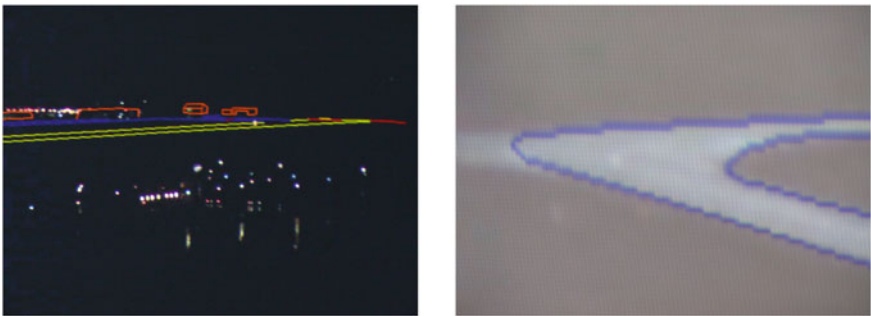
Fig. 3 Overlaid geographic information

Fig. 4 Meteorological overlay with actual wind speed, direction, 2 min average and minimum–maximum values

2.4 Presentation of Weather Information

Actual weather information was projected on the circular panoramic screen on places without covering traffic, see Fig. 4. Actual wind direction and speed are displayed including 2 min average and minimum and maximum values. Runway Visual Ranges were displayed in the lowest part of the panorama screen.

2.5 Sensor Data Fusion

Objects observed by the video cameras were tracked in the central tracking unit. Radar tracks from the Approach Radar were merged with the video tracks; see Fig. 5 *right part.*

Fig. 5 Left: Aircraft tracked only by the Terminal Approach Radar (labels with call sign or SSR code and altitude in hundreds of feet); Middle: aircraft tracked by the video camera only (label with track number); Right: aircraft tracked by both radar and video

2.6 Presentation of Aircraft and Vehicles

Aircraft and vehicles were automatically marked with a rectangle around their observed shape and were labelled with a track number when observed by the video tracker (Fig. 5 *middle* section). The track number (ID1234) could be changed into flight identity manually or by the automatic merge with the radar track. Radar information was added to the label if the track was detected by the radar (Fig. 5 *left* section) and when inside the airspace with specified range and altitude from the field. Aircraft both tracked by video and radar carried a rectangle-diamond contour and a radar label.

2.7 Pan-Tilt-Zoom (PTZ) Camera

The PTZ camera could be remotely controlled from its HMI, see Fig. 6 *left*. It had 768 × 576 pixel resolution and a zoom factor of 36 (1.7 degree minimum view angle). The PTZ camera would sweep 180 degrees in 2 s in order to catch an object quickly. The PTZ monitor (Fig. 6 *left*) provided presets for hot spots on the field (tiles

Fig. 6 Pan Tilt Zoom camera Human Machine Interface (left) and Picture in Picture (right)

around the PTZ image). Manual steering of the camera could be done by the mouse on either the PTZ monitor or the panorama screen. The actual heading direction and zoom factor of the camera was graphically indicated on a compass rose at the right top corner of the PTZ monitor (see Fig. 6 *left*). The PTZ camera could be slaved to a track and its image was also displayed on the panorama screen as a Picture-in-Picture (PIP) (Fig. 6).

3 Test and Validation Program

The requirements for the ART functions had been derived from 'problem driven' operational concept procedures for remote tower control, having in mind that solutions should be acceptable for remote tower controllers and cost beneficial. Emphasis was on safety and situational awareness. Both should be at least equal or better as compared to real tower operations. A preliminary safety assessment was part of the ART project. It was updated with the validation results and used in further research and development.

Early implementations of the ART functions were evaluated by air traffic controllers and further developed in at least two cycles before entering the evaluation and validation program. Fifteen air traffic controllers participated in the validation, each spending two days in the remote cabin in groups of two to three controllers. Seven controllers came from the Swedish field Ängelholm that was remotely displayed. Seven controllers came from other Swedish airfields and one controller came from a Dutch military airport. Their average age was 45 years, ranging from 28 to 58 years and they had an average job experience of 20 years as an air traffic controller, varying from 1 to 32 years.

Due to safety restrictions only passive shadow mode for single airport remote control was possible, meaning that actual control of traffic was done from the Ängelholm real tower, while controllers in the remote position in Malmö judged their function as if they were in full control.

The proper functioning of the ART functions was verified by testing against requirements. The validation was conducted by distant real time observation of traffic at the Swedish Ängelholm airport. Recordings were used to evaluate less frequently occurring visibility conditions. The European Operational Concept validation Methodology (E-OCVM, 2007) was applied. E-OCVM is a strict validation methodology leading to definition of objectives and hypotheses to be validated. For the ART functions about 70 had been defined and worked out in two questionnaires with about 138 statements ranking from '1' for complete disagreement to '6' for full compliance with the statement. Data were collected via debriefings, questionnaires for, during and after the test runs, and observations. Observations were carried out by Subject Matter Experts and Human Factors specialists.

The validation program consisted of a familiarisation and training phase during which the controllers could make themselves familiar with the proper operation

of the ART functions. The ART functions were validated incrementally and in combinations:

Part A—Validation of: Panorama Display, Weather Presentation and Geographic Information display;

Part B—Traffic Presentation (Labelling) and PTZ functions.

Part C—Pan Tilt Zoom Camera and Tracking functions;

Part D—Validation of the Visual Enhancement Technology;

Part E—Validation of the combination of all previous mentioned ART functions;

Part F—Expert Judgement Workshop.

The Expert Judgement Workshop, with an international audience of ANSP management and policy makers, covered all validation aspects that required involvement from such audience.

Ängelholm airport is an airport in southern Sweden with one runway, taxiways on both sides of the runway and an apron with passenger terminal on the opposite side of the runway about 1500 m from the tower. The tower has an elevation of 30 m above the field. The shortest distance from the tower to the runway is 700 m and the distance to the thresholds is about 1400 m.

4 Results

The prototype ART functions were validated during typical autumn conditions; rain, low visibility, dispersed showers and low cloud base conditions. Main emphasis was on the controller appreciation of working conditions and their situational and safety awareness. The program spent also several hours with each group of controllers during night time conditions. The traffic for Ängelholm consisted of about 20–30 aircraft per day. Aircraft movements consisted of a mix of scheduled flights, training flights and occasionally charters and business flights. Additionally, movements of vehicles on the taxiway and runway were surveyed. These movements could mainly be attributed to runway inspection cars, maintenance vehicles and towing trucks (either with or without rotating and flashing lights). The following results originate from the answers to the questionnaires and debriefings. In the context of this limited publication only the highlights are given. An extensive version of the ART prototyping results with more quantitative and descriptive details is being published as part of the project documentation (TRIP-ART, 2010).

4.1 Results for the Panorama Display

Visibility in the remote tower was found to be of less quality than in the real tower. Overall, the confidence in the projection system was anyhow high among the controllers. The controllers found the small distortions of the panorama image

due to the composition from nine cameras acceptable. The camera—display combination was not performing sufficiently in resolution and in detection capability to survey all objects and movements on and around the airfield, compared to real tower operations (for a quantification of the performance difference between DLR's RTO-video panorama and real tower out-of-windows view see Chapters "Model based Analysis of Two-Alternative Decision Errors in a Videopanoramabased Remote Tower Work Position" and "Model Based Analysis of Subjective Mental Workload during Multiple Remote Tower Human-in-the-Loop Simulations"). The controllers complained about missing depth of view. It was difficult for them to estimate distance and to judge which aircraft was closer. The controllers found the nine cameras in combination with the panorama display acceptable for ATC operations of single aircraft only. They expressed however to have problems to use this panorama set-up for handling multiple aircraft. The automatic camera adjustments for changing light conditions did not interfere much with the controllers' tasks, but a risk existed that controllers are not fully aware of the real daylight conditions, especially during twilight. During twilight, remote controllers might think that it is daylight condition. Overall the controllers' awareness of the meteorological conditions was less; they also expressed to have some difficulties to judge visual aspects of the clouds.

4.2 Geographic Information Display

There was no consistent opinion among the controllers on the use of geographic overlays. Controllers familiar with Ängelholm said that they didn't need extra synthetic reference information. This contrasted with non-Ängelholm controllers, who found the extra reference lines helpful. The participants slightly agreed that geographical information can be useful during darkness and low visibility though it has to improve. They judged it would not significantly benefit capacity. The overlay may obscure other important information and it is felt slightly cluttering the display.

4.3 Weather Presentation

The controllers slightly agreed that weather information on the display is useful. Controllers preferred to position the weather information at a location of the panorama display of their own choice, for instance close to the touch down zones. Overlaid weather information would be helpful to keep their eyes on the screen e.g. in gusty conditions. It would not cause more workload, but it could eventually cover other important information. The presentation of the Runway Visual Range was appropriate and controllers felt confident about it.

4.4 Results of Traffic Presentation (Labelling and Tracking)

Controllers preferred labels, irrespective of the sensor source from which these were derived. Target tracks and labels were considered useful, but mostly during night and low visibility. Their source (radar, video or both) should be indicated in the target symbol. Labels tended to increase controller's situational awareness, but controllers did not judge tracking performance good enough (so far) to increase capacity and to improve safety in low visibility conditions (visibility < 2000 m). Workload was judged slightly increased. Labels for aircraft and vehicles were expected to improve the capability of controllers to follow, monitor and control traffic. Controllers considered the risk to obscure important information with labels as slight. When labels overlapped, controllers were able to manually put them apart and make them legible. However, automatic label de-conflicting would be preferred. Label swops were considered a safety risk. Any mismatch between video and radar target should be removed. Adding a label, editing the label content and switching the label appearance was considered easy, which also applied to manual track termination. Display of different target symbols and labels for aircraft and vehicles was found intuitive with respect to the source of the track (video, radar or combined).

4.5 Results of Validation of the Visibility Enhancement Technology (VET)

VET increased the luminance of higher intensity areas with a factor 2 and lowered the lower intensity areas also with factor 2, providing more contrast between the high and the low brightness areas. The controller expectations were high (see through fog, make the invisible visible). Controllers wanted the whole picture to be enhanced in contrast and the effect should be larger. VET produced noisy pictures during night. In contrast, the PTZ turned out to be much more light-sensitive in the dark than the visibility enhanced panorama cameras. This effect was enlarged due to automatic exposure control of the cameras, which worked better for the PTZ (optimising a zoomed-in part) and less well for the panorama cameras (averaging the whole image).

VET did not convince the participating controllers to improve visibility and awareness in the way it was set-up in these validation exercises. This finding was irrespective of visibility and day/night conditions. VET did not allow operating at lower visibility thresholds as compared to standard Low Visibility Procedures. In low visibility, the additional visibility offered by VET didn't enable seeing all objects that controllers need to see at and around the airfield with sufficient detail. VET neither enabled earlier detection.

4.6   Results for the PTZ Camera and Object Tracking

The controllers found the PTZ rather useful for searching and detecting aircraft and vehicles, for manual and automatic runway inspection and for inspection of aircraft and vehicles, most of all during daylight and good visibility. The PTZ Picture in Picture should be moveable to any position on the panorama screen. The response of the PTZ camera was considered good enough and residual time delays were acceptable. The automatic tracking capability of the PTZ depended on the choice made for central video tracking and thus its performance. Controllers did not expect to handle more traffic with PTZ. The availability of the PTZ picture-in-picture camera favoured to keep a better focus on the panoramic display, but there was a risk to stay too long with the PTZ. Controllers found the PTZ operating procedures easy to use and felt confident using the PTZ camera.

4.7   Results of Validation of the Combination of All ART Functions

In comparison with real manned tower operations, the controllers could not stay ahead of traffic with the ART functions in the form that they were tested in these live trials. They had a slight tendency to focus too much and too long on the new ART functions. Controllers expressed a thought that in an ART environment (= more synthetic), there is a risk of forgetting something important since you don't have all "real" visual inputs in the same way. They also expressed a feeling of not being able to plan and organise tower control in the same manner as in the real tower. Despite the ART functions, controllers searched for information that is easier to find in the real tower. The ART functions therefore need more development and better integration before being accepted. Using just one mouse for all integrated ART TWR operations/systems, as tested in these trials, was somewhat complicated. The mouse had to be positioned on the appropriate screen before the desired effect was obtained. On the one hand, controllers expected to learn quickly how to use these tools. On the other hand they expressed the need for a lot of training. The ART facility was judged moderately realistic in reproducing the Ängelholm airport.

Some controllers experienced too much workload overall in the ART cabin. Fatigue was said to be caused by sitting in the cab with tempered light and noise from the cooling fans in the projectors.

4.8   Results from the Expert Judgement Workshop

About 25 subject matter experts participated in the Expert Judgement Workshop to share their opinion on matters not directly related to hands-on air traffic control.

They worked out their opinions in three ART related discussion blocks: (1) Implementation of remote tower functions, (2) Costs/benefits as expected for remote tower applications and (3) Opportunities as seen for ART.

The experts found that the implementation of ART functions can be broadened to non-remote applications at large airports (extra surveillance and contingency applications) and remote applications in areas with an extreme climate as there are for example airports in Polar Regions. Airports with Flight Information Service (AFIS) only can be enhanced in service with a selection of ART functions, giving better flight information remotely. The experts agreed on better resolution and detection capability in next maturity level of ART. ART procedures need to be further developed and special airspace for remote tower operations is given a thought. More elaborated safety and human factor cases were on the wish list, as were the development and implementation of ART regulations and licensing.

The experts expected a reduction in cost of tower operations on small and mediums size airports. Also more opening hours were expected in low visibility giving a better business case and probably attracting more customers. ART functions would benefit safety and thus save lives and avoid the costs of accidents and incidents. The ART technology would also be of benefit for airport and aviation security. The ART realisation could bring more uniformity in training and operations. However, working remotely increases the gap between the remote controllers and local personnel and decreases the 'on-the-spot' knowledge of the field.

Remotely operated airports might be specifically applicable for hosting of emergency openings at unmanned airfields and at airport with comparable geographical locations such as closely connected grouped airports (with similar weather and traffic conditions), airports with a similar infrastructure and airports at unfavourable locations.

Next steps for ART as suggested by the expert group are: better performance (resolution, depth of view, visibility enhancement, tracking, better positions for the cameras, better working conditions). Cooperation with other Air Navigation Service Providers was promoted. Study is needed to apply ART on more than one airport at a time and to introduce ART in active control. The PTZ was most preferred for application of ART functions on manned towers. This result is in agreement with findings of the DLR-DFS shadow mode validation experiment as reported in Chapters "Assessing Operational Validity of Remote Tower Control in High-fidelity Simulation" and "Model Based Analysis of Subjective Mental Workload during Multiple Remote Tower Human-in-the-Loop Simulations".

5 Analysis and Recommendations

5.1 Observations

The ART validation program was executed with live trials in 'passive shadow mode'. Live trials with a more active role for the air traffic controller were not possible because of time constrains and safety reasons.

The statistical analysis of the responses showed high standard deviations in the answers of the controllers on 100 of the 138 statements that were judged in the questionnaires. Possible explanations for the large standard deviation are: insufficient exposure to the scenario needed for testing the hypothesis, not sufficient familiarisation and training, system immaturity or misunderstanding of the questions. Further analysis showed a bias between controllers from Ängelholm and the other controllers. The local controllers from Ängelholm were on average less positive on the ART functions.

5.2 System Maturity

The ART project tested advanced functions with different maturity. The ART functions were not yet mature enough for operational integration. ART was just a step in the evolutionary process to develop optimal remote tower control facilities and procedures. Most of the ART functions needed further development and testing. ART participants were generally positive about the PTZ, and presentation of targets and labels. ART participants were somewhat negative about the current resolution of the panorama display, VET and the tracking performance.

5.3 Operational Aspects and Recommendations

The ART operational evaluation by 15 active controllers and 25 subject matter experts revealed valuable operational knowledge about the application of remote tower technology. The experiments showed that the ART level of maturity would, at this time, allow for single aircraft VFR and IFR operations only.

Resolution (1360×1024 pixels per camera) and detection capabilities with ART video cameras would need to be improved. Controllers suffered from lack of situation awareness when surveying traffic on the panorama screen. Higher resolution would require extra bandwidth of the data transmission channels. Smart data compression algorithms might be required to fit all data in existing and near future data communication means. This could be more expensive in application.

The optimal positioning of cameras is open for further investigation, mainly in order to keep camera costs low while optimizing the camera output.

With the ART functions as tested, performance of remote tower operations was perceived inferior to performance of real manned tower operations. This would be the main subject of investigation during the research and development phase for the next maturity level.

The automatic exposure of the surveillance cameras might lead to wrong controller perception of daylight conditions. A study could be undertaken to find the right automation in this context.

The overlaid geographic or synthetic information should be further explored. Controllers were happy with the option to switch it on or leave it off, but they asked for thinner and/ or dashed lines. This might be favourable in combination with higher picture resolutions.

Controllers liked to have weather information projected on the panorama screen but had no other preference for display of wind and runway visual range information on the panorama screen other than a copy of the existing instruments on their desk in the real tower.

Tracking of video objects and fusion of video with radar data are required to perform to high standards as this is giving the controllers confidence in automatic surveillance. Tracking is safety critical when controllers use it for decision making. High performance tracking is needed for reliable track stability and track identity. In this context the ART video tracking and data fusion should be improved. When it provides a better surveillance performance, controllers will make more use of it and they will get the benefit of improved detection capability as compared to visual surveillance. Installing cameras for video tracking of targets closer to the runways, taxiways and aprons should be investigated.

The track labels should be designed to automatically de-conflict with other labels or other objects. It will reduce the risk to cover important surveillance information.

To increase capacity in low visibility the ART Visibility Enhancement Technology (VET) was expected to look through fog. In the few validation occasions of low visibility controllers wanted more effect and to a greater extend, preferably on all images. The VET performed but not to controllers' expectations. The intrinsic noise of video cameras in low light conditions made VET in the current form less useful. Further enhanced trials need to be set up, and other sensors or combination with sensors, like infra-red would need to be tested.

The Pan Tilt Zoom (PTZ) camera was the best of class in the ART evaluation. Controllers wanted to have it for real manned towers also, already in its current set-up mode. If supplied with reliable automatic tracking control, it would even be more appreciated. Its feature to project a zoomed in enlarged picture on the panorama screen should get more flexibility in choice of position. The positive judgement regarding the PTZ camera and the requirement for automatic tracking agrees with the findings within the DLR-DFS shadow mode validation trials (with 8 participating active controllers, see Chapters "Which Metrics Provide the Insight Needed?... " and "Model Based Analysis of Two-Alternative Decision Errors…".

Working with the integrated ART tools could be improved by further research and development to improve the working conditions. The dimmed lighting conditions (in a dark room environment) and the 9 projectors with continuous noise seemed

to make controllers tired in comparison to the real environment. It is also possible that the picture frame update rate of 20 frames per second was visually tiresome. This is supported by the results of video framerate experiments at DLR, analyzing two-alternative decision errors as dependent on framerate. These experiments indicated a minimum framerate between 35–40 Hz for minimizing decision errors under observation of dynamic events (aircraft landing, see chapter "Videopanorama Frame Rate Requirements…"). The mouse operation as the central operation device for many ART functions should be further optimised. It should not be needed to drag the mouse over a large distance to activate a function on a specific screen.

The ART type of operations could be applied in other areas: in climate unfriendly areas, as contingency for large airports, possibilities to perform remote aerodrome control simultaneously for more than one small airport for a controller, and to improve the information provision on airports with only Flight Information Service (AFIS).

Additional to earlier detected cost benefits, ART could widen opening hours of airports and attract more users by providing punctuality in services. Also security can benefit from this technology.

It is recommended to continue to develop the remote tower procedures, to investigate multi-airport operations and to expand on safety and human factor cases, on regulations, training and licensing.

It is recommended to investigate the need of visual information quality in relation to sensor data information for control of aerodrome traffic.

References

ART. (2006). Advanced remote tower project. In *6th Frame Work Program Project Funded by the European Commission, DGTREN, Technical Annex to the Contract TREN/07/FP6AE/S07.73580/037179*.

Brinton, C., & Atkins, S. (2006). The remote airport traffic services concept: Opportunities and requirements. In *2006 I-CNS Conference*. Baltimore, MD.

Ellis, S. R. (2006). *Towards determination of visual requirements for augmented reality displays and virtual environments for the airport tower*, NATO HFM-121 RTG 042 HFM136.

E-OCVM. (2007). EUROPEAN operational concept validation methodology. *European Air Traffic Management Programme, Version 2*.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., & Werther, B. (2007). Augmented vision video panorama system for emote tower operation: Initial validation. *EUROCONTROL Innovative Research Workshop*.

IFAC2010. (2010). *11th IFAC/IFIP/IFORS/IEA Symposium on analysis, design, and evaluation of human-machine Systems*. Valenciennes, France, August 31–September 3, 2010 and NLR-TP-201-417.

LFV-ROT-BC, LFV. (2008). *Remotely operated tower business case*.

TRIP-ART. (2010). *Transport and innovation portal for the advanced remote tower project*. Webpage downloaded on May 9, 2015 from http://www.transport-research.info/web/projects/project_details.cfm?id=36007.

Designing and Evaluating a Fusion of Visible and Infrared Spectrum Video Streams for Remote Tower Operations

Fabian Reuschling, Anne Papenfuss, Jörn Jakobi, Tim Rambau, Eckart Michaelsen, and Norbert Scherer-Negenborn

Abstract The research project INVIDEON evaluated requirements, technical solutions, and the benefit of fusing visible (VIS) and infrared (IR) spectrum camera streams into a single panorama video stream. In this paper, the design process for developing a usable and accepted fusion is described. As both sensors have strengthens and weaknesses, INVIDEON proposes a fused panorama optimized out of both sensors to be presented to the air traffic service officer (referred to as 'operator' in this context). This chapter gives an overview of the project and reports results of the operators' perception performance with the visual, infrared, and fused panorama as well as results of the acceptance and usability of the INVIDEON solution. Main findings of requirements for fusing VIS and IR camera data for remote tower operations are highlighted and set into context with the air traffic control officer's (ATCO's) tasks. A specific fusion approach was developed within the project and evaluated by means of recorded IR and VIS data. For evaluation, a testbed was set up at a regional airport and data representing different visibility conditions were selected out of 70 days data recordings. Six operators participated in the final evaluation. The objectively possible detection performance and the recognition performance of the

F. Reuschling · A. Papenfuss (✉) · J. Jakobi · T. Rambau
German Aerospace Center (DLR), Institute of Flight Guidance, Lilienthalplatz 7, 38108
Braunschweig, Germany
e-mail: anne.papenfuss@dlr.de

F. Reuschling
e-mail: fabian.reuschling@dlr.de

J. Jakobi
e-mail: joern.jakobi@dlr.de

T. Rambau
e-mail: tim.rambau@dlr.de

E. Michaelsen · N. Scherer-Negenborn
Department Object Recognition, Fraunhofer-IOSB, Gutleuthausstrasse 1, 76275 Ettlingen,
Germany
e-mail: eckart.michaelsen@iosb.fraunhofer.de

N. Scherer-Negenborn
e-mail: norbert.scherer-negenborn@iosb.fraunhofer.de

operator was determined, evaluated, and compared to the theoretical performance derived from the Johnson Criteria. Subjective data on perceived usability, situational awareness, and trust in automation was assessed. Furthermore, qualitative data on Human Machine Interface (HMI) design and optimization potential from debriefings and comments was collected and clustered.

1   Introduction

Remote Tower Operations base on introducing cameras as a substitute for the out-of-the-window view (OTW) to enable the provision of air traffic services (ATS) from a working position remote of the OTW and the need of physical presence of ATCOs at the airport's location. This allows the ATS to be centralized and the flexible allocation of personnel. Furthermore, controller assistance systems, based on sophisticated optical sensors and image processing, can be implemented. For instance, it is of high interest to support the ATCOs situation awareness, particularly in impaired visibility conditions, because candidate airports for remote tower operations are likely to be equipped only with a minimum of sensors and heavily rely on visual observation.

The basic remote tower sensors refer to visible spectrum (VIS) video cameras. Nevertheless, in some of today's operational remote tower sensor suites, infrared (IR) cameras are included providing an infrared panorama in addition to a VIS panorama. Those IR enhanced implementations of the remote tower concept run into a display dilemma. They may either permanently display both panoramas, VIS and IR, on top of each other, or, when just one panorama is displayed at a time, controllers have to switch between VIS or IR optical sensor presentations in order to check which view currently is the better one. From a human–machine interface design point of view, each solution has disadvantages in terms of switching gazes between different panorama displays or maintain gazes but switching between VIS and IR presentation. Both, switching gazes and switching modes, consume mental resources (switching costs). Furthermore, important information could be missed, because it might only be visible in one of the representations, which is just not monitored.

Therefore, within the research project INVIDEON, a solution for fusing infrared and visible camera streams into a sole panorama display was developed. It bases on a series of four workshops in which operators' requirements (all of them were air traffic controllers) with regards to usability and safety were defined and evaluated. With the evaluated requirements, a technical solution for the fusion process was developed and systematically evaluated by operators in a controlled setting.

2 Theoretical Background

2.1 Characteristics of VIS and IR Data

Images and videos taken in the visual domain are omnipresent. Most often, they come in the three bands known as red, green, and blue. The atmosphere permits high transparency in these bands. The corresponding imagery is taken in analogy to the human visual perception system. The goal is realism, i.e., a video recorded with this method is meant to give the observer the same impression as if they were watching the scene directly. In the ideal world, there should be no difference between the OTW view and a VIS-camera recording displayed on screen. This makes these frequency bands perfectly suited to remote tower applications.

However, it is well known that the dynamic range, the dynamic resolution, and the geometric resolution of the human visual perception can be very good. Camera and screen equipment meeting or surpassing these specifications is very expensive, in particular if panoramic projection is needed. Here compromises must be made. E.g., the human visual fovea is capable of resolving between $0.4' = 0.12$ mRad and $2.0' = 0.58$ mRad, where operators are supposed to be on the better side. In the INVIDEON experimental setup the visual cameras had a resolution of 0.48 mRad (1.65 arc minutes) per pixel. The number of pixels raises quadratically with the resolution and with it the required bandwidth for transmission. Accordingly, the costs for camera-, transmission-, processing-, and monitor-equipment corresponding to human visual resolution are orders of magnitude from making sense.

Otherwise, the human visual perception has no sensitivity to the thermal spectral bands at 3–5 µm and 8–12 µm, respectively. In these bands the atmosphere also allows large detection ranges, sometimes even larger than in the visual domain. The property that is measured in these bands is mostly apparent temperature (instead of reflected sunlight in the visual case). The situation is similar to that of a blacksmith perceiving a hot iron workpiece and estimating its temperature from its brightness. In these bands our whole environment is glowing. To the untrained observer such imagery is puzzling and blurred. This would make such perception-mode unsuited to the task at hand. It is, however, in some respect complementary to the VIS-mode, e.g., also working at night without actively lighted targets. The IR-mode shows properties, such as the temperature of an engine, that are not seen in the VIS-mode.

With the IR-equipment, even more radical compromises must be made than with the VIS-equipment. Thermal cameras are generally much more expensive than VIS-cameras, they provide monochrome images and they have much lower geometrical resolutions. The dynamic resolution can be very high, but only for cooled systems, which are even more expensive.

2.2   Influence of the Atmosphere and Aspect

ATCOs are well aware of the limits given by the transmission of the atmosphere in the visual spectral domain. They know that the inspectable range is highly dependent on the weather. Likewise to the OTW view, fog, dust, precipitation, etc. reduce the detection ranges dramatically. In the thermal spectral bands, the atmosphere is also not completely transparent, and the transmission range is depending on weather parameters such as moisture, dust and aerosol densities (Hanafy, 2014), precipitation, turbulence (Stein et al., 2014), etc., as well. However, there is one order of magnitude between the wavelengths as compared to the visual spectral domain, with the scattering and attenuating effects of droplets depending on the ratio of droplet size and wavelength (Hanafy, 2014). In particular the transition between Rayleigh and Mie scattering shifts considerably to larger droplets and particles. Thus, in temperate climate zones the transmission range of thermal bands tends to be longer than the transmission range of the visual band. On the other hand, serious signal-blocking, e.g., due to dense snow fall, will affect both visual and thermal sensing in a similar manner.

Moreover, the appearance of the objects of interest must be considered: Military aircraft are often painted in a camouflage tone deliberately chosen to reduce their visibility in the visual spectrum. On the contrary, civil aircraft often feature mainly bright tones, such as silver or white, contrasted with salient high hue colors also in large patches. This makes their detection much easier in the visual domain. For the thermal domain, such differences in paint are irrelevant. Instead, the thermal contrast between the aircraft and its background is perceived. In particular in dim lighting conditions, the visual detection of aircraft depends on their active emission using their various lights. In the thermal domain, such lights may also appear a little warmer than the surrounding, albeit much less salient. Thus, an approaching aircraft at night will be visible in the visual domain before it is seen in the thermal domain— due to its landing lights. Of course, the detectability also depends on the aspect of the aircraft. In the visual domain an aircraft covers more area when seen from the side, as compared to viewed frontally or from behind. In the thermal domain, the aspect dependency is much stronger, and the saliency also depends on operational status: When viewed from aft, the thermal signature is dominated by the hot nozzles or exhausts. It is very bright as long as the engines are active, and even brighter on full throttle. Thus, a departing and climbing aircraft can be observed much longer in the thermal domain, as compared to the visual domain. However, viewed from other aspects, such as frontal directions, these most salient parts may well be occluded.

2.3   Image Fusion

Image fusion is a special branch of data fusion, which aims at improving visual quality. The goal of image fusion approaches can be described as merging contrasting,

but complementary features from multiple sensors (Omar & Stathaki, 2014). The relevant features are defined by the users of such a system. In general, relevant objects should be highlighted by preserving information of the context, like weather information.

A prerequisite for image fusion is accurate pixel-to-pixel co-registration. In general, this can be difficult, because the entrance aperture of the visual and the thermal cameras cannot be at the same locations, causing parallax. Here, airport scenarios are rather benign, because all objects of interest are in large enough distances from the tower. In particular flying aircraft are observed over distances of several kilometers, so that about a meter of aperture displacement can be neglected. Birds and objects on the ground, such as vehicles or taxying aircraft, as well as contours of the infrastructure can be so close, that the parallax accounts for a few pixels displacement. This can be a bit disturbing as the example in Fig. 17 shows. Also, calibrating the co-registration causes extra efforts that need to be accounted for. However, these need not be repeated ever so often. Moreover, for the investigations presented here, the videos were recorded with the different spectral channels running asynchronously. Thus, a temporal registration of all used material became necessary, which also caused extra efforts, and may also be the source of occasional double-appearance of moving objects. In an operational scenario, for every time instance the latest frames from each source will be fused in real time. Hence such artefacts should disappear automatically.

In the literature on the topic there is often an implicit assumption of symmetry between the two sources, i.e., they are of roughly equal importance, and the resulting fused output image should contain equal content from both sources (Bavachan & Krishnan, 2014; Canga, 2002; Firooz, 2005; Jagalingam & Hegde, 2014; Khidse & Nagori, 2014; Naidu & Raol, 2008; Toet, 1989; Toet et al., 1989). Moreover, in most of these approaches the scale of the objects of interest is assumed unknown. Therefore, approaches based on image pyramids prevail. Both these assumptions are violated in the case at hand, and thus the relevance of the image-fusion state-of-the-art to the task at hand is in question.

2.4 Detection and Recognition of Objects

Operators are responsible for the preventing collisions, organizing and expediting the flow of air traffic. So, monitoring the aerodrome to detect expected and unexpected objects is one of the main tasks. Of special importance is the detection and recognition of aircraft arriving and departing from the airport, in order to monitor landing and take-off, as these are the most safety–critical flight phases.

In the literature, the detection of an object is described as one of three basic discrimination tasks of image evaluation, being (1) detection, (2) recognition, and (3) identification of objects. Object detection refers to it being "sufficiently different from the background in which it is embedded so that the observer becomes aware of its presence" (Taylor, 1969). This difference against the surrounding may stem from

a contrast in luminance or color, the size, shape, texture, or movement of the object or its temporal characteristics (Taylor, 1969).

The higher-order image evaluation tasks recognition and identification rely on the preliminary detection of the object. Generally, recognition is defined as the classification of the detected object (i.e., if it is a car or an aircraft), while naming the corresponding member of the class is referred to as identification. In some cases, an object may implicitly be recognized as it is detected, because it is the only type that would be plausible in the specific context. For example, an object moving in the sky at high altitude will most probably be an aircraft (Taylor, 1969).

The minimal object size still allowing for a successful completion of each of the three discrimination tasks described as well as for the correct judgement of orientation has been quantified first by John Johnson (1985). It is determined by placing an object in the distance in which a detection, orientation judgement, recognition, or identification is barely possible. Subsequently, a pattern of alternating black and white lines with the same contrast to the background as the object is moved to the same distance. The number of lines is increased until an observer can just distinguish them.

The required resolution for each of the discrimination tasks is then expressed as the number of line pairs, consisting of one black and one white line, over the critical dimension of the object. The critical dimension is the one that is the most difficult to perceive and depends on the specific object. It commonly is the smallest dimension, as exemplified in Fig. 1 for the front and side view of an aircraft.

Johnson found the necessary resolution to be roughly equal for the nine military objects he tested. He therefore concluded, that it is constant and independent from the object's size, distance to the observer, contrast to the background, and signal to noise ratio. The respective average resolution values, known as the Johnson Criteria, are given in Table 1.

The criteria greatly simplify the process of object perception and do not account for various factors that affect the detection and recognition performance of human observers. A non-exhaustive summary of these factors is given by Hershel C. Self (1969) and includes the image's size and overall brightness and the target's shape, size, color, and isolation from the background and other objects. It also lists human

Fig. 1 Front and side view of aircraft with critical dimensions and associated line pairs, illustrated in analogy to Johnson (1985)
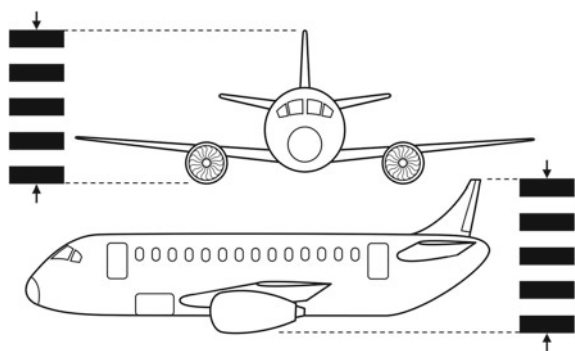
Table 1 Summary of the Johnson criteria (Sjaardema et al., 2015)

Discrimination level	Description	Necessary resolution [Line Pairs per Critical Dimension]
Detection	Awareness of significant object	1.0 ± 0.25
Orientation	Object's aspect perceived	1.4 ± 0.35
Recognition	Classification of object (car, aircraft, …)	4.0 ± 0.8
Identification	Specific member of class discernable	6.4 ± 1.5

factors like training, experience, motivation, instructions and task briefing, assumptions, and the compromise of speed versus accuracy expressed in individually acceptable false detection rates (Valeton & Bijl, 1995). Other influences not listed by Self are the targets' aspect ratio, orientation, and contrast against the background, the clutter of the scene, and the weather condition (Sjaardema et al., 2015).

The influence of image fusion on detection and recognition has been examined in multiple studies, some of which found increased performance while others found the opposite (Krebs & Sinai, 2002). The benefit of fusion is therefore considered to be task dependent, with improvements possible in spatial orientation and scene recognition tasks, but not for target detection. In general, the utility of an image fusion for detection and recognition tasks depends on the circumstances and the specific fusion algorithm used.

For Remote Tower Optical Systems, the discrimination levels formulated by Johnson are adapted in the Minimum Aviation Systems Performance Standard (MASPS) for Remote Tower Optical Systems (ED-240A change 1) (Minimum aviation system performance standard for remote tower optical systems ED-240A change 1). In this context, recognition is defined differently than in Table 1 as the perception of the "attributes of an object relevant to ATS [Air Traffic Service] provision" (Minimum aviation system performance standard for remote tower optical systems ED-240A change 1), such as an aircraft's size, configuration, and painting or the vehicle type (e.g., fire truck, fuel truck, baggage trailer) instead of the sole classification of an object, e.g., as aircraft. Mapping the recognized aircraft to a specific flight plan (or aircraft registration) is then referred to as identification, also differing from Johnson's definition. However, air traffic control officers (operators) are usually not able to perform this discrimination task without additional information, like position data, besides the visual observation (Minimum aviation system performance standard for remote tower optical systems ED-240A change 1). In the context of Remote Tower Optical Systems, only detection (becoming aware of a significant object) and recognition (perception the object's attributes) are therefore considered.

Specific requirements for a Remote Tower set-up are developed for each aerodrome individually by defining the range in which a certain object shall be detectable and in which it shall be recognizable. From this Detection and Recognition Range Performance (DRRP) (Minimum aviation system performance standard for remote

tower optical systems ED-240A change 1), the necessary system properties, such as the cameras' resolution, are then determined. This new user requirements driven approach is widely accepted now. It better matches operational needs and equipment performance and goes beyond the former attempts to define strict performance requirements with respect to resolution. ED-240A change 1 (Minimum aviation system performance standard for remote tower optical systems ED-240A change 1) includes a detailed explanation of how to apply the Johnson Criteria in the context of Remote Tower Optical Systems.

3 Designing Image Fusion

The goal for fusing VIS and IR data at the remote tower controller working position (CWP, in the remote tower control context rather referred to as Remote Tower Module [RTM], which then per definition includes the visual presentation (OSED for remote provision of ATS to aerodromes)) is to provide the operator with one single information source that contains the best information possible about the air traffic and relevant infrastructure of the outside world. Even in a conventional tower working environment, operators have to scan several information sources spread over different layers and combine them into a mental representation, which is called the mental traffic picture (Moehlenbrink et al., 2011; Mogford, 1997). The times, when operators do not look outside but scan other information sources, are called "Head-Down-Times" (Hilburn, 2004; Papenfuss et al., 2016). Those gaze switches between head-up and head-down, respectively far and close view should be reduced to a minimum, because it induces stress for the human eyes and even worse, relevant information can be overseen. Following this argumentation, the visual and infrared presentation of the outside world should be fused to a sole presentation in order to reduce the number of information sources to be scanned by the operator.

3.1 Key Requirements

The requirements for such a fusion are motivated by two hypotheses: First, image fusion should compensate shortcomings of Remote Tower technology, mainly lower resolution of video data compared to the OTW view. Second, image fusion should provide more information than available today in a conventional tower, especially in low visibility conditions. Hence, the operators' performance with the fused panorama should not be worse than with the best single panorama in any relevant situation.

In general, operators have to be aware of every aircrafts' position and movement around the aerodrome to conduct safe and efficient air traffic control. While this can be achieved by for example using radar surveillance, detection in the OTW view has benefits regarding safety and efficiency. Low visibility conditions limit the operators' ability to detect aircraft in the OTW view, requiring them to use technical aids,

such as radar surveillance or even an A-SMGCS (Advanced—Surface Movement Guidance and Control System). Depending on the visibility range and the available systems, this leads to more work for the operators and may require them to limit the number of aircraft allowed to move at a time—in some cases even to one movement at a time, as outlined in the ICAO's European Guidance Material on All Weather Operations at Aerodromes (European guidance material on all weather operations at aerodromes—EUR Doc013.2016).

An optimal fused panorama would allow operators to detect aircraft and vehicles more reliably in impaired visibility conditions, reducing the need of costly radar surveillance aids or to restrict the number of aircraft movements allowed. Furthermore, recognition of aircraft type—for instance by aircraft contour—is beneficial for the efficiency of aerodrome ATS operations and hence the second requirement for the fused panorama.

In a Remote Tower environment, the visual panorama presentation is important for the operator's monitoring task as it provides relevant information of the context. This includes the observation of weather phenomena like clouds and visibility conditions and the perception of an aircraft's or object's color. Thus, the VIS video data should minimally be disturbed in the fusion process and information from the infrared panorama only added in locations where they are necessary to highlight aircraft in the sky and on ground.

3.2 Hot-Spot Only Fusion

The above described requirements lead to a strong asymmetry between the VIS and IR data. This contradicts the typical assumptions for fusion algorithms found in the literature (compare Sect. 2.3) requiring a custom fusion method. The thus designed fusion scheme filters out the brightest areas of the infrared image—which originate from warm objects like aircraft—and overlays them on the visual panorama. It was therefore named "Hot-Spot Only Fusion" and consists of the following six steps:

1. High pass filter on the IR image:
   Only objects of a certain fairly small size are of interest. Large homogenous image segments, that typically constitute the vast majority of the IR-image pixels, should not be used in the fusion. This is achieved by subtracting a low-pass version of the IR-image from the IR-image itself. Written as arithmetic operation this step takes the form

$$I_{IR,1} \leftarrow I_{IR,0} - G\big(I_{IR,0}, 10\big) \tag{1}$$

   Here $I_{IR,0}$ is the IR image as given by the sensor. G is a Gaussian low-pass filter implemented by Fast Fourier Transform (FFT) with parameter $\sigma = 10\ pixel$. $I_{IR,1}$ is the IR image after this first filter step. The implementation

via FFT is preferred to a straight-forward convolution, because it is more efficient concerning computational resources. Recall, very large panoramas are to be processed in real time here. The choice for $\sigma = 10 \; pixel$ was based on the example material provided. This is roughly the order of magnitude of the expected objects of interest.

2. Remove artefacts lower than 0, and at the margins of $I_{IR,1}$. Recall, operation (1) will give negative values occasionally. After that the image $I_{IR,1}$ is mostly black, with hot spots of the desired size preserved, as well as some lines where salient image edges where in $I_{IR,0}$.

3. Narrow low-pass filter on the IR image $I_{IR,1}$:
   This improves appearance, and removes fuzzy artefacts. The effect is actually rather small. There is no need for another FFT. Such small filters can be efficiently executed by straight-forward convolution:

$$I_{IR,2} \leftarrow G(I_{IR,1}, 1) \tag{2}$$

   $I_{IR,2}$ is the IR image after this second filter step.

4. Gamma correction emphasizing contrasts at low temperatures:
   γ curves are non-linear functions applied to the intensities of an image. If the first parameter of the function is larger than 1, the curve will be bending upward—enhancing contrasts in the bright image regions. This is not desired here, because the hot parts of aircraft, such as the engines, already have very high contrasts in the IR-frequency domain at hand. If the first parameter is smaller than 1, the curve will be bending downward—enhancing contrasts in the darker image regions. This is desired here, because those parts of the aircraft, that are moderately warmer than the background—such as the fuselage—are of high interest here. Also faint small dots in the sky, resulting from distant aircraft, can thus be enhanced. We choose 0.5 for that parameter resulting in a rather strong contrast enhancement. The second parameter of the γ curve sets the fix-point. Here, grey-value 30 remains fixed. In the image material at hand, this intensity was frequently corresponding to the fuselage of nearby aircraft manoeuvring on ground.

$$I_{IR,3} \leftarrow \gamma(I_{IR,2}, 0.5, \; 30) \tag{3}$$

   $I_{IR,3}$ is the IR image after this γ correction step.

5. Linear contrast amplification:
   Here, a piecewise linear transformation is applied:

$$I_{IR,4} \leftarrow \min\left(1, \frac{1}{30} \cdot I_{IR,3}\right) \tag{4}$$

   It yields values bounded between 0 and 1. $I_{IR,3}$ intensities higher than 30—e.g., at hot engines—end up in saturation at value 1. However, recall: most of the resulting image is 0 in $I_{IR,3}$ as well as in the new image $I_{IR,4}$, usually over

99% of the pixels. The hot saturated parts are often surrounded and completed by small regions with mediocre values. $I_{IR,4}$ is the IR image after this fourth enhancement step.

6. Fuse using the α-channel:

In image processing, the α-channel controls the transparency of an overlay. The filtered and enhanced IR-image $I_{IR,4}$ is utilized in this control role. The overly itself is the original IR-image $I_{IR,0}$. The underlying image is the visual colour image $I_{VIS,0}$:

$$I_{fused} \leftarrow I_{VIS,0} + \alpha_+(I_{IR,4}) \cdot I_{IR,0} \tag{5}$$

All zero-parts of $I_{IR,4}$ act as transparent. There, the visual image remains untouched. Only the few bright spots and lines in $I_{IR,4}$ are opaque. Wherever $I_{IR,4} = 1$ (in saturation), the visual image will be completely replaced by the original IR-image. Where $0 < I_{IR,4} < 1$, there is a corresponding semi-transparent mixture. I_{fused} is the final fused image displayed to the ATCO.

Figure 2 gives two examples of this fusion. The upper one shows an aircraft over the runway after takeoff. The aircraft is warmer than the surroundings resulting in a hot spot. It appears with high contrast in the fused image. Most of the rest of the fused image is taken from the VIS camera, while some of the runway contours are enhanced as well, and hot spots also appear in the background resulting, e.g., from persons or vehicles. The lower example given in the figure was taken under poorer visibility. Here, the aircraft has just landed. Again, the fuselage and engines are warmer and thus appear much enhanced in the fused image. Also, contours of the runway and taxiways appear in the fused image. These were invisible in the VIS channel.
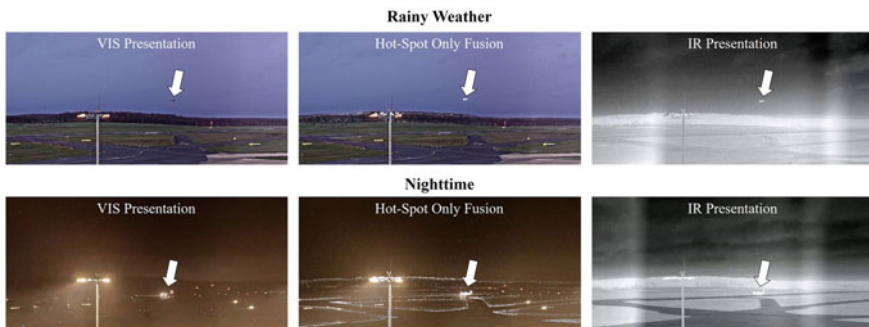

Fig. 2 Two examples of "Hot-Spot Only" fusion applied to VIS and IR images under two light situations, the arrows mark the aircraft's position

4 Evaluating Hot-Spot Only Fusion

The aforementioned Hot-Spot Only fusion is evaluated by presenting recorded videos of aircraft movements at the three different visual representations (VIS, IR, and Hot-Spot Only Fusion) to an expert group of participants. As opposed to using live data, this allows the selection of specific weather conditions and movements. In the following section, the preparation of the videos and the experimental set-up and procedure are described.

4.1 Research Questions

Within the study, the necessary data is collected to answer four fundamental research questions:

1. Which visual presentation objectively offers the best detection performance?
2. In which visual presentation is the relevant information best visible to the operators?
3. What is the benefit of fusing VIS and IR images?
4. How do the operators subjectively evaluate the fusion of VIS and IR images?

4.2 Sample

The experiment took place in Braunschweig, Institute of Flight Guidance, from March 18th to 22nd 2019. Five ATCOs from the German Air Navigation Service Provider (DFS-AS) took part. Additionally, one former ATCO, who still held a valid license, participated, but did not answer all subjective questionnaires.

4.3 Remote Tower Camera Set-Up

The Remote Tower camera set-up at the Braunschweig-Wolfsburg aerodrome (BWE) comprises two circles of 10 visual spectrum cameras and 16 infrared cameras, capturing a full 360-degree horizontal field of view (HFOV). The set-up is located on the roof of a building (see Fig. 3) close to the aerodrome's apron and at approximately the same distance from the runway as the aerodrome's tower is.

As only the arrival and departure sectors and the northern traffic pattern are of interest to the project, the panorama recordings and presentation was limited to 220°. This area is covered by the HFOV of six visual and 10 infrared cameras, as depicted in Fig. 3.

The visual spectrum cameras are Q1635s from AXIS, employing a CMOS sensor with a resolution of 1920×1080 pixels, covering a field of view of $53.8°$ by $30.2°$. The
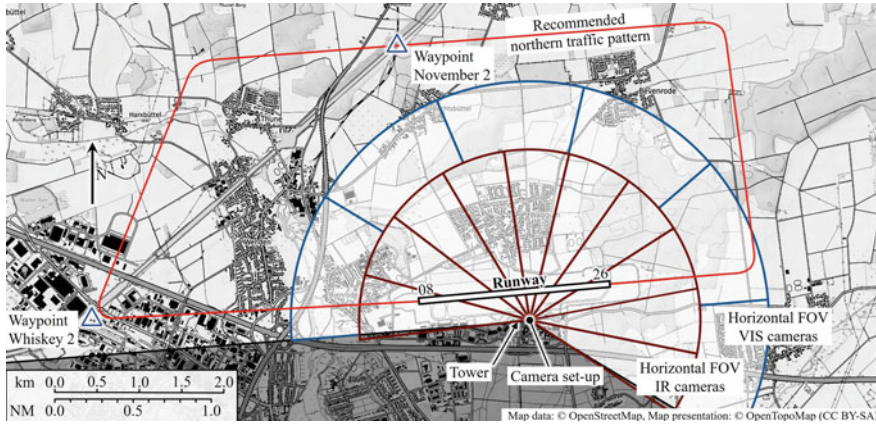
Fig. 3 Camera set-up at BWE airport; blue circle segment indicating FOV of the six VIS cameras and red segment indicating HFOV of IR cameras used for evaluation

infrared cameras are from DRS Technologies WatchMaster® IP Elite 6000 series. They use an uncooled VOx microbolometer detector and register infrared radiation in the 8–14 μm (LWIR) spectral band. They feature a resolution of 640×480 pixels with a field of view of $37.5°$ by $28.0°$. Both camera types operate at 25 frames per second (FPS).
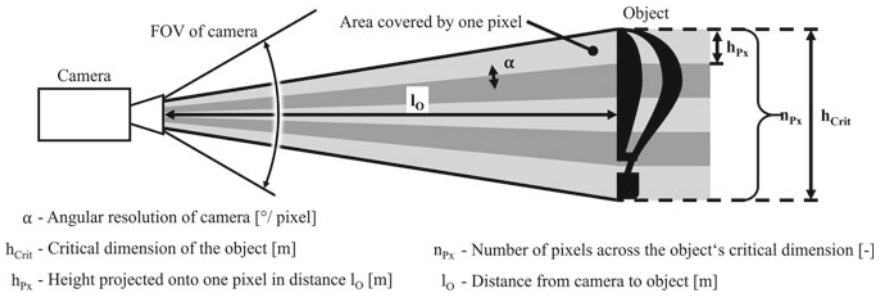
By applying the Johnson Criteria introduced in Sect. 2.4, a rough estimate of the theoretically possible detection and recognition distances with the camera set-up used is calculated. The first step in this process is to translate the numbers of line pairs (i.e., the actual Johnson Criteria) into numbers of pixels, which is a one-to-one conversion according to Johnson (Sjaardema et al., 2015). In literature, however, a second calculation scheme described by Bailey, Wilz, and Arthur III (2012) can be found. For this approach, each line—black or white—equals one pixel, so that the number of pixels across the critical dimension of an object is

$$n_{Px} = 2 * n_{linepairs} + 1$$

with one pixel being added as the pattern starts and ends with a black line (see Fig. 1). In this paper, both approaches are considered and will be compared to one another and to the actual performance of the operators in Sect. 6.5.

The geometric relations necessary for the calculation are pictured in Fig. 4. The relevant variables are the critical dimension of an object h_{Crit}, the distance from the camera l_O, the camera's angular resolution α, the number of pixels the object's critical dimension is projected on n_{Px}, and the projected object height per pixel h_{Px}.

For the described camera set-up, the angular resolution of the visual spectrum cameras—calculated from the resolution and the field of view—is $0.028°$/Pixel. The critical dimension is determined for a Dornier 328, which is representative of the traffic at BWE during the period of the study. It generally refers to an object's

α - Angular resolution of camera [°/ pixel]

h_{Crit} - Critical dimension of the object [m] n_{Px} - Number of pixels across the object's critical dimension [-]

h_{Px} - Height projected onto one pixel in distance l_O [m] l_O - Distance from camera to object [m]

Fig. 4 Geometry of the Johnson criteria

smallest dimension—in this case the aircraft's height—so that $h_{Crit} = 7.24$ m is used. Lastly, the necessary number of pixels is derived from the Johnson Criteria (compare Table 1).

The object's height that is projected onto one pixel and the resulting distance to the camera are then calculated as

$$h_{Px} = \frac{h_{Crit}}{n_{Px}} \text{and} l_O \approx \frac{h_{Px}}{\tan(\alpha)}$$

The detection and recognition ranges determined with the two calculation schemes are listed in Table 2.
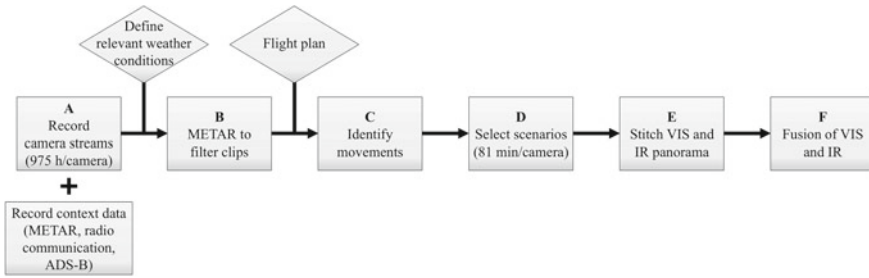
4.4 Preparation of Video Streams

The workflow for video stream preparation is visualized in Fig. 5. In addition to the video data, context data was recorded (A). They include official published weather reports (METARs), radio communication data, and aircraft position data (ADS-B). They were recorded time synchronized and in chunks of 15 min.

Starting in early December 2018 until mid-February 2019, video data sums up to 975 recorded hours. Based on preliminary test recordings with a reduced set-up, five weather/visibility conditions for arriving and departing aircraft were agreed on with the DFS-AS dealing as "visibility" treatment factors in the validation design: good visibility, reduced visibility down to 3400 m, rain, bright light shining into the camera, and night. Additionally, two conditions for traffic patterns—good visibility and cloudy sky—and the startup of an aircraft's engines were selected. The properties of the weather and traffic conditions are summarized in Table 3.

In order to extract appropriate video sequences, the recorded videos were analyzed in a multi-step process. Firstly, the METARs were scanned for the sought weather conditions (B). Then, the scheduled traffic at the aerodrome was identified using flight plan data to determine timeframes of interest (C). For these timeframes, the weather visible in the recordings was compared to the corresponding METAR, giving

Table 2 Theoretical detection and recognition distances with the visual panorama according to the Johnson Criteria for a critical dimension of 7.24 m

| | | n_{Px} as by Bailey et al. (2012) | | | | n_{Px} as by Johnson (1985) | | | |
| | | Detection | | Recognition | | Detection | | Recognition | |
		Min	Max	Min	Max	Min	Max	Min	Max
Necessary line pairs per critical dimension (from Table 1)		1.25	0.75	4.80	3.20	1.25	0.75	4.80	3.20
Necessary number of pixels per critical dimension	n_{Px} [−]	3.50	2.50	10.60	7.40	1.25	0.75	4.80	3.20
Projected height per pixel	h_{Px} [m]	2.069	2.896	0.683	0.978	5.792	9.653	1.508	2.263
Distance to object	l_O [m]	4233	5926	1398	2002	11,852	19,753	3086	4630

Fig. 5 Workflow of video stream preparation

Table 3 Properties of the agreed-on weather and traffic conditions; the defining characteristics are marked in bold

Weather condition	Properties			
	Time of day	Visibility	Downfall	Other
Arrivals/Departures				
CAVOK	Daylight	Above 9999 m	No downfall	**CAVOK reported**
Bright light	Sunrise or daylight	Above 9999 m	No downfall	**Sun shining directly into camera**
Low visibility	Daylight	**Below 3400 m**	No downfall	Significant fog
Rain	Daylight	Above 9999 m	**Significant downfall**	
Night	**Night**	Above 9999 m	No downfall	Landing lights clearly visible
Traffic patterns				
CAVOK	Daylight	Above 9999 m	No downfall	**CAVOK reported**
Cloudy	Daylight	Above 9999 m	No downfall	**sky covered in clouds**
Engine startup	Not specified	Not specified	Not specified	Not specified

an overview of the weather conditions covered. At the same time, the actual traffic, including aircraft type and operation mode, was noted. This enables the selection of scenarios with identical aircraft throughout all weather conditions and a maximum of additional traffic. Subsequently, the video sequences for the experiment were selected (D).

The scenarios chosen were then cut from the recordings. As to accommodate the whole departure or arrival and a maximum of traffic, the length of each scenario has been set to six minutes. Differing from that, the aerodrome circuits are nine minutes and the engine startup three minutes long. While extracting the scenarios, the videos of each camera were synchronized up to one frame (0.04 s), enabling the

best possible overlay in stitching and fusion. Afterwards, VIS and IR video images of each scenario were stitched into a single video panorama of 8192 × 1024 pixels (E). Finally, both video streams were fused following the above described "Hot-Spot Only" approach (F).

4.5 Independent Variables

In total, four independent variables are manipulated during the study. These are the visual presentation of the scenarios, the weather condition, and the traffic and aircraft types visible in the video sequence.

Visual Presentation—The same traffic scenarios were presented to the operators in three different modes—IR, VIS and fusion (compare Fig. 2). For VIS, the video streams of six VIS cameras were stitched. For IR, the video streams of 10 IR cameras were stitched and enlarged to the pixel size of the VIS panorama. For fusion, the two stitched video streams were fused according to the procedure described in Sect. 3.2.

Weather—The selected scenarios cover the weather conditions listed in Table 3. The actual weather at the time of arrival or departure is shown in Fig. 6 as frame from the visual panorama.

Traffic—For each weather condition a scenario with a departing aircraft and one with an arriving aircraft was selected. To be able to compare the scenarios, the same type of aircraft, a Dornier 328, was chosen. Equally, runway 26 is used throughout all scenarios. In addition, the startup of a Dornier 328's two turboprop engines and traffic patterns of an Aquila A 210 were selected. These represent situations typical to the Braunschweig-Wolfsburg aerodrome and allow for a broader discussion of the fusions' benefits. The engines are numbered as pictured in Fig. 6, with engine one being the first to start and engine two being the second.

**Fig. 6** Overview of selected weather conditions

Table 4 Number of evaluated aircraft and with position data available in brackets per weather condition

Aircraft	Wingspan [m]		Weather condition				
			CAVOK	Bright light	Night	Rain	Low visibility
Robin R-1180 Aiglon	9.08	Departures				1 (−)	
Cessna 172 N Skyhawk II	10.90	Departures	1 (−)				
Dornier 328	20.98	Departures	2 (2)	1 (1)	1 (1)	3 (3)	1 (1)
		Arrivals	1 (1)	1 (1)	1 (1)	1 (1)	1 (−)
Dassault Falcon 7X	26.21	Arrivals		1 (1)			1 (1)
ATR 72-500	27.05	Departures					1 (−)
Total number		Departures	3 (2)	1 (1)	1 (1)	4 (3)	2 (1)
		Arrivals	1 (1)	2 (2)	1 (1)	1 (1)	**(1)**

Aircraft Type—Besides the aforementioned reference aircraft, four more aircraft types are visible in the video sequences. The types, wingspan, and respective number of departures and arrivals per weather condition are listed in Table 4. Furthermore, in brackets the number of aircraft for which ADS-B position data are available is stated.

4.6   Dependent Variables

Objective Detection Time (ODT)—The objective detection performance reached with each visual representation is evaluated by determining the first appearance (for arrivals) or last appearance (for departures) of an aircraft as single pixel discernable from the background noise. The corresponding timestamp is independent from the properties of the video sequence, like start time and length, and precisely defined for each aircraft in all three visual presentations (VIS, IR, and Fused). It is determined by a frame by frame analysis in which every single aircraft is followed from the point where it is clearly visible until it has fully disappeared, as exemplified for an arriving aircraft in Fig. 7. The precise frame the first pixel—for arrivals—or last pixel—for departures—of an aircraft is visible in is determined by skipping the video sequence and enlarging areas when needed. The timestamp associated with that frame is the Objective Detection Time (ODT) of an aircraft. For arrivals, this is the earliest possible time an aircraft can be recognized by the operators and for departures, it is the latest possible time the operators can still recognize the aircraft.
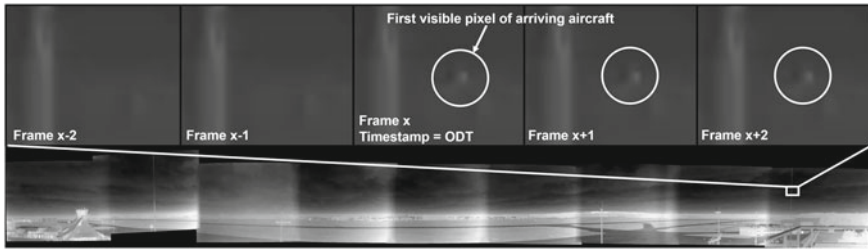
Fig. 7 Illustration of a frame by frame analysis of an infrared panorama to determine the Objective Detection Time (ODT) of an arriving aircraft

Recognition Time (RT)—The visual performance of the operators using the fused panorama is assessed by calculating the time needed to recognize an aircraft in each of the three visual representations VIS, IR and Fused compared to the ODT. The time of recognition was marked by the participants by pressing the space key on a standard computer keyboard while watching the sequences. The Recognition Time for each aircraft is then derived from comparing these timestamps with the ODT. The rules for calculating Recognition Times are explained in detail in Sect. 4.9.

Detection and Recognition Distances—With the position information available for some aircraft (see Table 4), the distances of the aircraft to the camera location at the time of detection and recognition in the three visual representations are calculated. The ADS-B data used therefor include the callsign, a timestamp, the position information as a latitude–longitude pair in degrees (accurate to four decimals), and the barometric and geometric altitudes in meters (accurate to two decimals). The time interval between two data points, as calculated from the timestamps, varies between 1 and 36 s with a mean of 11.8 s. In general, ADS-B data were found to be up to 33 m accurate, which is sufficient in the context of this study (Zhang Jun & Zhu, 2011).

The ADS-B data are processed to extract the horizontal distance between the aircraft and the camera set-up—this can be displayed on a 2D-map—and the direct distance which factors in the height difference between aircraft and set-up. If necessary, a linear interpolation is used between two data points, which was deemed appropriate for the general estimate sought herein.

Usability—Usability of the system was assessed with the System Usability Scale (SUS) (Brooke, 1996). The scale consists of ten items, with five of the items formulating positive, and five items formulating negative behaviors of the system. The items were rated on a 5-point Likert scale. The scale was administered after participants had the option to use the system. The SUS score is further normed to a range from 0 to 100. According to literature, a system rated 70 or higher is "rather usable" (Brooke, 2013).

Situation Awareness—Situation Awareness (SA) is an important human factor contributing to overall human performance. To achieve adequate SA, operators need to perceive all relevant information (level 1), understand their meaning (level 2) and project the current status into the future (level 3) (Endsley, 1995). Typically, in human factors evaluations level 1 SA is assessed. The following four specific questions for

assessing the impact of image fusion on the perception of relevant information were formulated. Operators rated four items on a 7-point Likert scale.

1. "The fused panorama gives me a sufficient overview of the overall situation at any time, allowing me to easily locate relevant vehicle and aircraft movements."
2. "I was able to see arriving aircraft just in time."
3. "I was able to see departing aircraft long enough."
4. "I would have been able to safely control the traffic with the fused panorama."

Trust—Fusing VIS and IR video streams increases the level of automation at the CWP, as some information processing tasks are conducted automatically. It is essential, that operators can trust the results of automated tasks. On the one hand, in case of insufficient trust, operators will double check the results and as such, the fusion will generate no benefit but additional taskload. On the other hand, degraded overall performance can be associated with an "overtrust" in an automated system, that is also referred to as complacency or reduced monitoring of the automated system. Empirical studies have proofed that in cases, where automation is not perfectly reliable—which is the case for the image fusion—operators should be aware of these limits and monitor raw information sources (Parasuraman et al., 2008). Consequently, operators should not be complacent and be aware of potential artefacts generated by image fusion. A commonly used questionnaire to assess trust is the SHAPE Automation Trust Index (SATI) from EUROCONTROL (Dehn, 2008; SESAR HPrepository: SATI—SHAPE automation trust index). It consists of six items which are rated on a 7-point Likert scale.

4.7 Apparatus and Testbed

The experiment has been conducted in the Tower Lab at the DLR Braunschweig. There, the final panorama video sequences were displayed to the operators on a six-monitor set-up that is capable of showing every single pixel of the 8192×1024 pixel panoramas, as depicted in Fig. 8. Additionally, the corresponding radio communication was played in synchronization with the video. The controllers were given the complete documentation of the Braunschweig-Wolfsburg aerodrome including all necessary aerodrome charts to accommodate themselves with the surroundings and operational procedures. For all 13 scenarios, a short half-page briefing was handed to the operators including the METAR, runway in use, traffic situation (departure, arrival, traffic pattern, engine startup), and expected aircraft movements.

4.8 Procedure

Before the start of the experiment, the participants were briefed on the overall scope of the INVIDEON project. They were introduced to the technical set-up and their

Fig. 8 Testbed in DLR's tower lab

experimental task, especially on taking the timestamps. For the evaluation, the 39 different video sequences—13 scenarios times three visual representations—were presented to the participants with a randomized order between participants to control for learning effects.

The operators were asked to take timestamps, if they were certain that—depending on the scenario—one of five cases applied:

1. Arrival: They can see an arriving aircraft for the first time.
2. Departure: They cannot see a departing aircraft any more.
3. Traffic Pattern: They lose an aircraft in the traffic pattern.
4. Traffic Pattern: They find an aircraft in the traffic pattern again.
5. Engine Startup: Engine one or engine two is starting up.

They were furthermore encouraged to also take a timestamp when they made an observation which was specific to the respective part of the video sequence.

For each timestamp raised by the participant, the observer took a short note of the case and the corresponding aircraft. Overall, the experiment took nearly four hours. At the end, participants were asked to answer questionnaires to assess their subjective evaluation. They subsequently had time to make any comments or remarks that have not been covered so far.

4.9   Processing of Timestamps

The timestamps captured during the study are processed to extract the indicators needed to answer the research questions one to three of those outlined in Sect. 4.1 for departing and arriving aircraft.

Filtering of Data

Regardless of the research question to answer, all recorded timestamps are filtered before further processing them. From the total of 900 timestamps, firstly those are eliminated that concern the ability to recognize secondary objects, like birds or vehicles, or technical limitations of the system, such as aircraft leaving and reentering the field of view to the top. Those timestamps and the corresponding observer's notes are analyzed separately together with all further remarks of the participants by clustering them into areas of interest which are reported in Sect. 6.4.

The remaining timestamps are searched for wrong recognitions of objects as aircraft. These are either defined by the operators themselves—for example by commenting "I see some white thing over there, but I don't think it's an aircraft"[1]— or by comparison with the Objective Detection Times. Any recognition occurring before the aircraft is objectively detectable cannot be valid. For departures, a slightly later Recognition Time than objectively possible can be valid due to the reaction time of the operators which leads to all timestamps being taken after the actual recognition. Via this comparison of the Recognition Time with the ODT, three outliers were identified and eliminated. In one case, the first engine's startup was recognized three seconds before objectively possible. In another case, an arrival was recognized 30 s before the Objective Detection Time. In the last case, the startup of the second engine was recognized twice as late as by the other participants.

In 21 cases, an operator took more than one timestamp for the same aircraft in the same visual representation. For these, the timestamp closest to the Objective Detection Time is used. Overall, 514 timestamps for 18 departures and arrivals, two traffic patterns and the engine startups remain that are further analyzed. There is a total of seven timestamps for two departures, one arrival, and the engine startup that were not recorded by the operators.

Research Question 1—Objective Detection Performance

For the objective evaluation of the three visual representations, the Objective Detection Time introduced in Sect. 4.6 is used. To be able to compare the ODTs of different aircraft, the best Objective Detection Time across the three visual representations is determined for each aircraft. For departures, this is the highest timestamp, corresponding to a longer visibility. For arrivals, it is the lowest timestamp. The ODTs are then referenced to the best Objective Detection Time, resulting in relativized Objective Detection Times. The relativized ODTs are zero for the best visual representation and higher the lower the performance with the corresponding visual representation is.

---

[1] Original: "Da sehe ich etwas Weißes, aber ich glaube es ist kein Flugzeug.".
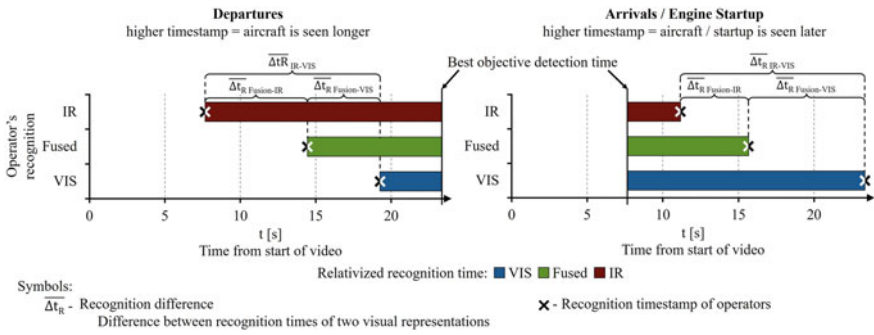
Fig. 9 Scheme for calculation of relativized Recognition Times

Research Question 2—Visibility of relevant Information

The second research question is assessed using relativized Recognition Times (relativized RTs), which are calculated by referencing the recognition timestamps of the participants to the best Objective Detection Time. The corresponding scheme is visualized in Fig. 9, where an "x" marks the time of recognition by the participants and the vertical line the best Objective Detection Time—both measured from the start of the video sequence. The difference between these values—pictured as colored bar—is the relativized Recognition Time. It is defined to always be positive and is higher the worse the operator's performance is.

Research Question 3—Benefit of Fusion

Lastly, the benefit of the Hot-Spot Only Fusion is evaluated by determining the differences of the Recognition Times for all pairs of visual representations (IR-VIS, Fused-VIS, and Fused-IR; see Fig. 9). The resulting recognition difference $\overline{\Delta t_R}$ is defined to be positive if the performance with the first representation is better—equaling a lower relativized Recognition Time—and negative if it is with the second representation.

Special Traffic Scenarios—Engine Startup and Traffic Patterns

For the engine startup scenario, higher timestamps indicate a later recognition and therefore a worse performance. This is analogous to how the timestamps for arrivals are interpreted. Hence, the timestamps for the engine startup scenario are processed the same way those for arrivals are.

The timestamps for the traffic patterns need to be treated differently, however. There is no singular time of recognition by the operator, but alternating timestamps of an aircraft being recognizable and it not being recognizable anymore. They enclose timeframes in which the aircraft is seen by the individual participant. Equally, the Objective Detection Times enclose timeframes of objective visibility. For both, the

proportion relative to the length of the video sequence—nine minutes—is calculated for the three visual representations. The objective and subjective timeframes of visibility are then visualized over the length of the video sequence.

5 Evaluation Results

In this section, the relativized Objective Detection Times and participants' Recognition Times for departures and arrivals as well as the results of the subjective questionnaires are presented. They are grouped according to their relevance for research question 1 (Sect. 5.1.1), research question 2 (Sect. 5.1.2), research question 3 (Sect. 5.1.3), and research question 4 (Sect. 5.3). Furthermore, the special traffic scenarios—engine startup and traffic patterns—are presented in Sect. 5.2.

5.1 Departures and Arrivals

5.1.1 Research Question 1: Which Visual Presentation Objectively Offers the Best Detection Performance?

Objective Detection Times—For the first research question, the relativized Objective Detection Times are of relevance. In Fig. 10, these are presented for departures and arrivals in the five weather conditions studied (CAVOK, Bright Light, Night, Rain,
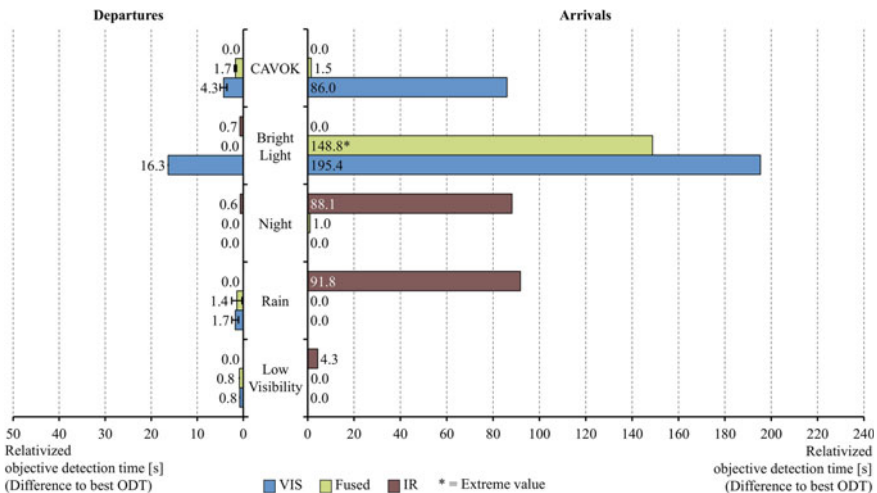
Fig. 10 Relativized objective detection times for departures and arrivals; average values for departures in CAVOK and rain weather conditions

and Low Visibility). For departures in the CAVOK and Rain weather conditions, relativized ODTs of two and three aircraft (see Table 4) are averaged. In these cases, the standard deviations are indicated. They range from 0.2 for the fused representation in the CAVOK condition to 1.1 for the fused representation in the Rain condition.

For departures, the visual panorama allows for the lowest Objective Detection Time, i.e., a relativized ODT of 0.0, solely during night. The maximum difference to the best ODT is 16.3 s in the Bright Light weather condition, which is about four times as high as the second highest value—4.3 s in the CAVOK condition. In the infrared representation, departures can be seen the longest in the CAVOK, Rain, and Low Visibility conditions. For the fused panorama, this is the case in bright light and during night.

Arrivals in night, rain, and low visibility can be seen first in the visual panorama. In the other two conditions, the objective detection occurs 86.0 and 195.4 s after the best Objective Detection Time. In contrast, the infrared panorama allows for the lowest Objective Detection Time in the CAVOK and Bright Light conditions. During night and in rainy weather, the ODT is 88.1 and 91.8 s higher than the best Objective Detection Time. With the fused panorama, the earliest objective detection is possible in the rain and low visibility conditions. While the difference to the best ODT is a maximum of 1.5 s in the CAVOK and Night conditions, it is 148.8 s in bright light. In the overall context, this is a very high value, as will be discussed in Sect. 6.1.

Overall, the differences to the best Objective Detection Times are larger for arrivals than for departures. The five highest values of arrivals—lying between 88.1 and 195.4 s—are at least 5.4-times greater than the highest value of departures—16.3 s.

5.1.2 Research Question 2: In Which Visual Presentation is the Relevant Information Best Visible to the Operators?

Recognition Times of operators—The second research question is answered using the processed (see Sect. 4.9) operators' recognition timestamps. The resulting relativized Recognition Times (RTs)—averaged over corresponding aircraft and all six participants—are visualized in Fig. 11 together with the standard deviations. Overall, the standard deviations are much larger than for the relativized ODTs, ranging from 0.2 for departures in the Low Visibility condition observed with the visual panorama up to 39.4 for arrivals the Bright Light condition observed with the infrared panorama. Unique to the CAVOK condition is that the standard deviation is very high—at least as high as the relativized RT itself—for all three visual representations.

Comparing the visual panorama to the other two visual representations (Fused and IR) in the same weather condition, the participants can follow a departing aircraft the longest during night where they lose it on average 2.2 s earlier than objectively possible. In the CAVOK and Bright Light conditions, the difference to the best ODT is much higher being 33.5 and 25.7 s. For the infrared panorama, the lowest relativized Recognition Time is observed in bright light, rain, and low visibility. The relativized RT for the CAVOK condition—22.7 s—is again much higher that the second highest value—3.7 s in bright light. Lastly, the fused panorama allows the operators to follow
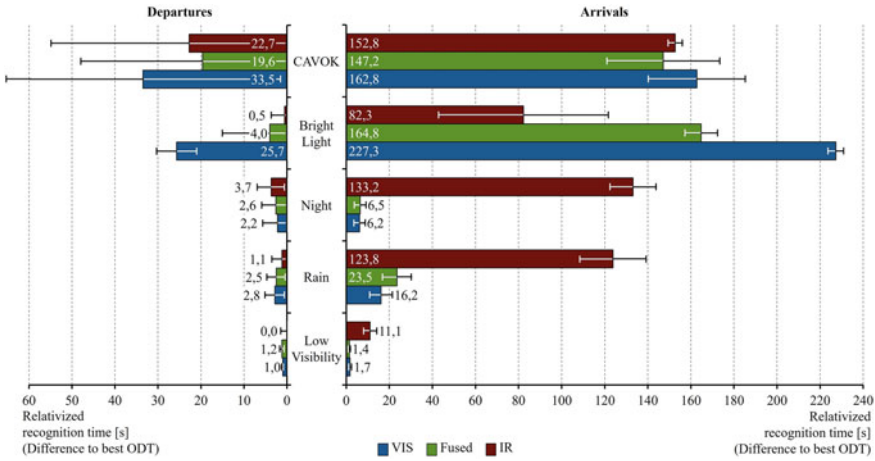
Fig. 11 Relativized recognition times of operators for departures and arrivals; average over n = 6 operators and corresponding aircraft

a departure in CAVOK the longest. On average, they lose an aircraft 19.6 s (SD = 28.3) earlier than objectively possible which is 4.9-times higher than the second highest relativized RT.

Arrivals during night and in rain are first seen in the visual panorama, i.e., the relativized RT is the lowest among the three visual representations. While the average difference to the lowest relativized Recognition Time is just 0.3 s in low visibility, it is much larger—15.6 and 145.0 s—in the other two conditions. The infrared panorama allows the participants to recognize arriving aircraft in bright light first compared to the Fused and IR representations. In this condition, the difference to the best ODT is 82.3 s (SD = 39.4). During night and in rain, arrivals are seen much later—126.7 s and 100.3 s—than with the second best representation. Using the fused panorama, the participants recognize arrivals in CAVOK and low visibility sooner than with the other two representations. In general, the differences to the best Objective Detection Times in the CAVOK and Bright Light conditions—147.2 and 164.8 s—are much larger than in the other three conditions—between 1.4 s and 23.5 s.

Detection and Recognition Distances—For operational procedures, the distances in which aircraft are recognized are much more important than the Recognition Times. Hence, the horizontal distances from the camera set-up are calculated for aircraft with available position information (see Table 4) from the Objective Detection Times and Recognition Times using the method described in Sect. 4.6. The results are illustrated in the upper part of Fig. 12. In the lower part, the geographical context is given together with the flightpaths approximated from the ADS-B data, the border of the aerodromes' control zone, and the area that is covered by the combined field of view (FOV) of the cameras installed in the set-up (see Fig. 3).

It is to be expected that the detection and recognition distances correlate with the relativized Objective Detection Times and relativized Recognition Times, meaning
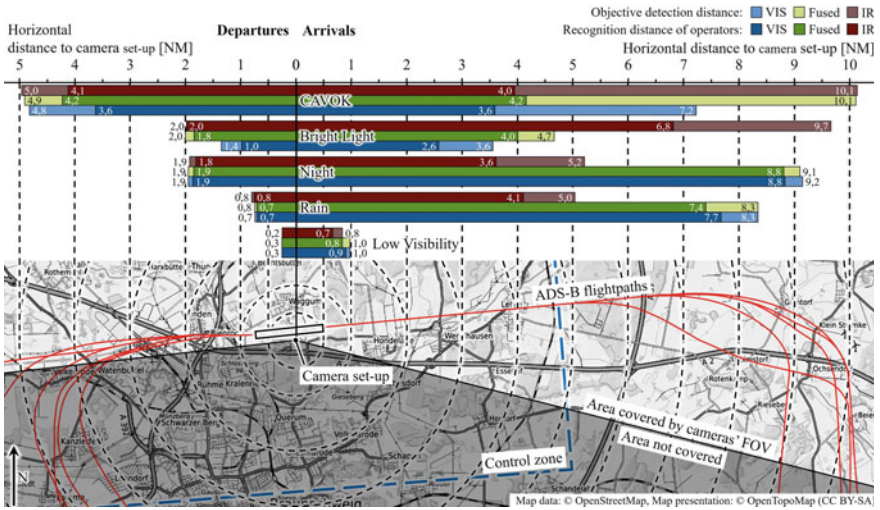
Fig. 12 Horizontal detection and recognition distances for departures and arrivals; average over n = 6 operators and medium and large aircraft

that the visual representations with the lowest relativized ODT or RT enable the largest detection distance or recognition distance. This is confirmed by the comparison of the distances presented in the above Fig. 12 with the relativized ODTs visible in Fig. 10 and relativized RTs depicted in Fig. 11. While the observations described for the Objective Detection Times (Sect. 5.1.1) and operators' Recognition Times (beginning of this section) can be equally made for the detection and recognition distances, the geographical context allows for three further findings:

First, six recognition distances for arriving aircraft throughout four weather conditions lie in a narrow range between 3.6 and 4.1 nautical miles (between 6.7 and 7.6 km) from the camera set-up. These distances roughly coincide with the time at which, after entering the airport's control zone, the aircraft's pilots contact the tower. Second, eight objective detection distances are greater than 7 NM (13.0 km). This point can be seen as the start of the final approach segment, clearly visible from the different flightpaths merging into one line. Finally, in low visibility, arrivals are first seen—objectively and by the validation participants—shortly before they reach the start of the runway at 0.5 nautical miles (0.9 km) from the cameras' location. The distance from the runway ranges between 0.5 NM and 0.3 NM (between 0.9 and 0.6 km) for the objective detection and between 0.4 NM and 0.2 NM (between 0.7 and 0.4 km) for recognition by the participants. Departures in this condition disappear about half way to the end of the runway.

5.1.3 Research Question 3: What is the Benefit of Fusing VIS and IR Images?

Difference between the Recognition Times of visual representations—The recognition differences calculated as described in Sect. 4.9 are presented in Table 5. In addition to the three pairs VIS-IR, Fusion-VIS, and Fusion-IR, the difference between the relativized Recognition Times of the fused representation and the best single representation (VIS or IR) is listed in the rightmost column of the table.

In the second column of Table 5 the observations of the two previous sections are mirrored: First, the differences for arrivals are higher—ranging from 9.38 to 144.98 s—than for departures, which range from 0.94 to 25.14 s. Second, the visual panorama allows for an earlier recognition compared to the infrared panorama for departures during night and arrivals in the night, rain, and low visibility weather conditions.

Table 5 Recognition differences of operators for departures and arrivals; positive values indicate that first representation is better

Traffic and weather condition	VIS vs. IR [s]	Fusion vs. VIS [s]	Fusion vs. IR [s]	Fused vs. best single representation [s]
Departures				
CAVOK	−10.72	13.82	3.09	3.09
Bright light	−25.14	21.70	−3.44	−3.44
Night	1.50	−0.40	1.09	−0.40
Rain	−1.67	0.31	−1.36	−1.36
Low visibility	−0.94	−0.20	−1.14	−1.14
Average	−7.39 (SD = 9.79)	7.05 (SD = 9.10)	−0.35 (SD = 2.24)	−0.65 (SD = 2.13)
Arrivals				
CAVOK	−10.04	15.62	5.58	5.58
Bright light	−144.98	62.51[a]	−82.47[a]	−82.47[a]
Night	127.00	−0.33	126.68	−0.33
Rain	107.64	−7.37	100.27	−7.37
Low visibility	9.38	0.29	9.68	0.29
Average (excluding extreme value)	17.80 (SD = 97.30)	2.05 (SD = 8.39)	60.55 (SD = 53.76)	−0.46 (SD = 4.60)
Average (excluding extreme value)	5.20 (SD = 70.29)	4.83 (SD = 9.13)	26.72 (SD = 46.94)	−0.56 (SD = 3.46)

[a]Extreme value (see Sect. 5.1.1)

The comparisons of the fused panorama with the visual panorama (column three) and infrared panorama (column four) result in an average recognition difference of 4.83 s (SD = 9.13) for the Fusion-VIS comparison and a larger 26.72 s (SD = 46.94) difference for the Fusion-IR comparison. In both cases, the extreme value for arrivals in the bright light weather condition identified in Sect. 5.1.1 was excluded.

When compared to the best single representation, the fused panorama provides the best recognition performance for departures in the CAVOK weather condition and arrivals in the CAVOK and low visibility conditions (see last column). On average, however, the best single representation enables 0.56 s (SD = 3.46) better recognition performance when excluding the extreme value.

5.2 Specific Traffic Scenarios

5.2.1 Engine Startup

The relativized Objective Detection Times and Recognition Times of the operators for the engine startup scenario are determined as described in Sect. 4.9 and are visualized in Fig. 13. They are averaged over all six participants.

For both engines, the best Objective Detection Time is achieved with the infrared panorama. The objective detections of the first engine's startup in the other two panoramas both occur 0.7 s after the best ODT. For the second engine, a detection of the startup is possible 0.3 s (VIS) and 0.7 s (Fused) after the best ODT.

For the startup of the first engine, the observers' recognition performance with the visual panorama is the best of the three visual representations with a relativized Recognition Time of 7.5 s and just slightly better than with the fused panorama (also 7.5 s). With the infrared panorama, the participants recognize the engine startup 9.8 s after the best ODT. In comparison, the startup of second engine is recognized later in the visual panorama and infrared panorama with the relativized RTs being 9.6 s
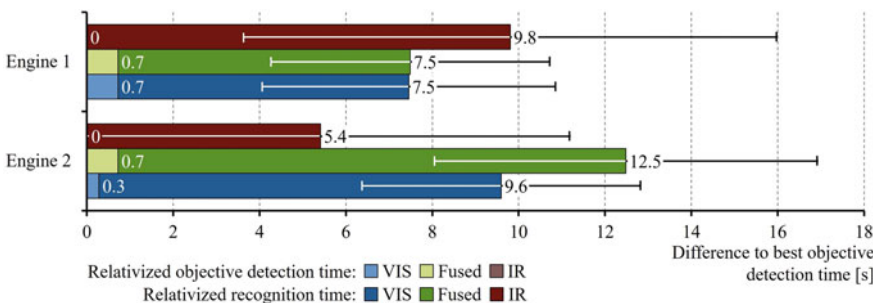
Fig. 13 Relativized objective detection and recognition times for engine startup scenario, average over n = 6 operators

and 12.5 s, respectively. The infrared panorama enables a faster recognition of the second engine's startup, 5.4 s after the best ODT.
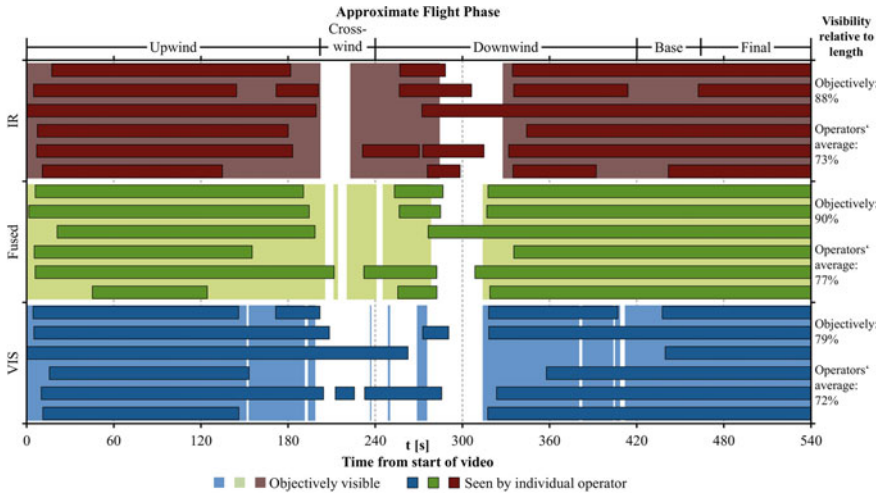
5.2.2 Traffic Patterns

The timeframes of objective visibility and visibility to the individual operator (compare Sect. 4.9) for the VIS, Fused, and IR representations and the CAVOK and cloudy weather conditions over the length of the video sequences are illustrated in Figs. 14 and 15. The approximate phases of the pattern (upwind, crosswind, downwind, base, and final) are marked.

In the CAVOK weather condition, the objective visibility with the visual panorama displays wide gaps during the downwind, base, and final flight phases. In total, aircraft are seen for 54% of the time. In the infrared representation, they are visible for the entire length of the video sequence. The overall objective visibility in the fused panorama amounts to 85% of the video length, with gaps during the downwind and base flight phases. For the cloudy condition, gaps are visible during the crosswind and downwind flight phases in all three visual representations. In the visual, infrared, and fused panorama, aircraft are objectively visible for 79, 88, and 90% of the video sequence length.

The individual operator sees aircraft in both weather conditions for a smaller proportion of the video sequence's length than objectively possible. The gaps in the visibility generally match the ones previously described. In the CAVOK weather condition, the operators on average recognize an aircraft for 46% of the video length in the visual panorama and for 96% of the length in the infrared panorama. For the fused
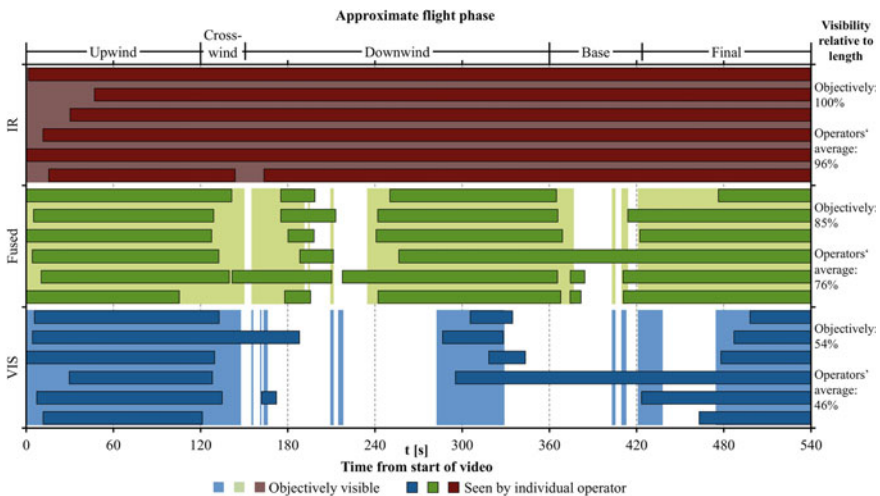
Fig. 14 Timeframes of objective visibility and visibility to operators for traffic pattern in CAVOK weather condition

Fig. 15 Timeframes of objective visibility and visibility to operators for traffic pattern in cloudy weather condition

panorama, the proportion is 76%. In cloudy weather, the proportions are 72% for the visual panorama, 73% for the infrared panorama, and 77% for the fused panorama. Compared to the CAVOK condition, the proportions are much closer together: While the difference from lowest to highest is 50% for the CAVOK condition, it is 5% for cloudy weather.

5.3 Research Question 4: How Do the Operators Subjectively Evaluate the Fusion of VIS and IR Images?

5.3.1 Perceived Usability

After observing the 13 different scenarios in all three visual representations, participants rated their perceived usability with a mean of 72.92 ($SD = 9.41$). The result is visualized in Fig. 16. This value is above the rating of 70, which is regarded as an threshold towards good usability (Brooke, 2013). However, it is lower than the reference usability score of 85.0 ($SD = 0$), assessed with n = 3 operators in the last INVIDEON workshop. In the workshop, a simulated workplace was set up where operators controlled simulated air traffic at Braunschweig airport by using a fused panorama that was based on simulated visual and infrared images. The details of the workshop are presented in (Hagl et al., 2018).
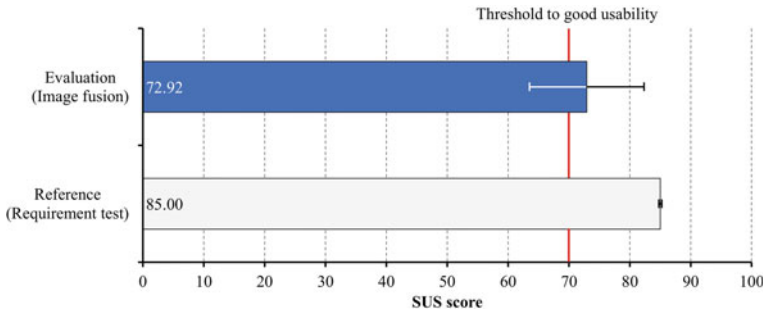
Fig. 16 Perceived usability of image fusion, mean of n = 6 participants

Table 6 Ratings for situation awareness items

Item	Mean	SD
"The fused panorama gives me a sufficient overview of the overall situation at any time, allowing me to easily locate relevant vehicle and aircraft movements."	4.00	1.22
"I was able to see arriving aircraft just in time."	4.20	1.30
"I was able to see departing aircraft long enough."	5.00	1.22
"I would have been able to safely control the traffic with the fused panorama."	4.00	1.87
Average	4.30	1.19

5.3.2 Perceived Situation Awareness

Four statements with relevant aspects of operators Situational Awareness where rated on a scale ranging from 0 to 6, the numbers are shown in Table 6. All statements were rated above the neutral point of three with standard deviations ranging from 1.22 to 1.87. The highest rating is 5.00 and was given for the statement "I was able to see departing aircraft long enough.". On average, situation awareness was rated with 4.30 ($SD = 1.19$).

5.3.3 Perceived Trust

With regards to the perceived trust, the rating on the standardized SATI scale, ranging from 0 "never" to 6 "always", was 3.69 ($SD = 1.01$). This rating is above the neutral point of 3 "often". In Table 7, the individual items of the SATI questionnaire and the rating of operators are shown. The term "system" refers to the technical set-up described Sect. 4.7. The best rating is 4.50 (SD = 1.05) and was given for the item, that the system was very often understandable. The lowest rating was given for robustness of the system, which scored 2.83 (SD = 0.75). This is also the only item that was rated below the neutral point of three.

Table 7 Ratings for trust-items

I felt that …	Mean	SD
… the system was useful	3.67	1.51
… the system was reliable	3.67	1.51
… the system worked accurately	3.83	0.75
… the system was understandable	4.50	1.05
… the system worked robustly	2.83	0.75
… I was confident when working with the system	3.67	1.21
Overall SATI score	3.69	1.01

6 Discussion

Based on the previously presented validation results, the four research questions are discussed in the following sections. At first, only the times measured for the visual and infrared panorama are considered to answer the first two research questions (Sects. 6.1 and 6.2). Then, the benefit of the Hot-Spot Only fusion is discussed, followed by an analysis of the subjective questionnaires in Sect. 6.4. Additionally, the applicability of the Johnson Criteria described in Sect. 2.4 is analyzed with reference to the calculated detection and recognition distances.

6.1 Research Question 1: Objective Detection Performance

The ODT sets the baseline the operators' performance is then compared to, as recognition can only occur after an object's detection. Furthermore, it is the basis for an objective assessment of what is the best visual representation in the weather conditions studied. In this section, only the visual and infrared representations are considered.

Different observations are made for arrivals in the CAVOK and bright light weather conditions and arrivals in the night, rain, and low visibility conditions. In good weather, the best detection performance is achieved with the infrared panorama. This can be explained by the low contrast of the usually light-colored aircraft against the bright sky (CAVOK condition) or the direct sunlight (bright light condition) in the visual panorama. In the infrared panorama, the warm aircraft appears white while the less radiant sky appears in dark colors (see also Sect. 2.2). Hence, there is a higher contrast between aircraft and background in the infrared panorama compared to the visual panorama leading to an earlier detection.

For arrivals in the night, rain, and low visibility weather conditions, the opposite effect is observed: In these conditions, aircraft are better detected in the visual panorama. Common to all three weather conditions is that the bright landing lights of the aircraft are the first element that is visible. They also have a high contrast against

the dim surrounding making them widely detectable, also outlined in Sect. 2.2. The landing lights cannot be discerned in the infrared panorama, however, as they do not have a distinctive heat signature. Furthermore, the uncooled cameras used in the set-up produce more noise the higher the humidity is leading to fuzzy images in the rain and low visibility weather conditions. A "see-through effect" often associated with infrared cameras could not be observed. Consequently, aircraft are less detectable in the infrared panorama than in the visual panorama.

For departing aircraft, the described effects are less visible or not visible at all. The reasons for this are threefold: First, departures in the CAVOK weather condition are seen until they leave the field of view, effectively restricting the relativized ODT (cf. Fig. 12). Second, in the bright light condition, the sky is less bright, leading to a higher contrast in the visual panorama. Third, in the night, rain, and low visibility conditions, the landing lights are not visible to the observer as the aircraft are seen from behind.

6.2 Research Question 2: Visibility of Relevant Information to the Operators

Analyzing the participants' Recognition Times, it is noticed that aircraft are best recognized in the infrared panorama in good weather, i.e. CAVOK and bright light, and in the visual panorama during night and in rain and low-visibility. It can be seen that this is the same distinction described for the Objective Detection Times in the previous section and leads to the conclusion that the panorama with the lowest relativized ODT is also the one with which an observer can follow departures the longest and sees arrivals the earliest. In other words, the panorama enabling the best *objective detection performance* is also the one enabling the best *subjective recognition performance*.

However, in the CAVOK condition the participants best perceive the information in the fused panorama—this is the representation with the lowest relativized Recognition Time—although they are objectively best seen in the infrared panorama. This may be because the fused panorama is more pleasant to watch than the grayscale infrared panorama as has been repeatedly remarked by the operators and will be explained further in Sect. 6.4.1.

Furthermore, it is seen for arrivals in the CAVOK and bright light weather conditions that the Recognition Times of the participants differ for all three representations to a great extend from the corresponding Objective Detection Times. In the night and rain conditions, this is visible for the infrared panorama, but not for the other two panoramas. When working with the visual or the fused panorama, the operators recognized aircraft close to the ODT. For these scenarios, it means that the operators quickly recognize a small number of moving pixels—that are perceivable as of the ODT—as aircraft. This is caused by the prominent landing lights during night and

in rain (see Sect. 6.1) leading to an "implicit recognition" (compare Sect. 2.4) of the bright pixels only visible in the visual panorama and the fused panorama.

It the other cases, the recognition occurs just after the aircraft enters the control zone at which time the pilots contact the aerodrome's tower and state their distance to the runway (see Fig. 12). As the participants were explicitly asked to only take a timestamp when they are *certain* that they recognized an aircraft, the position information is apparently necessary for the participants to be confident enough that the perceived pixels are an approaching aircraft.

The recognition distance is also affected by the individually required certainty (see Sect. 2.4). High standard deviations—for example for departures in the CAVOK weather condition—therefore result from the participants being unsure when to recognize a detected object as approaching aircraft or—in case of departures—at what time they do not recognize but rather only detect the aircraft. This observation was also pointed out by the operators themselves, who commented that there is a "Wide bandwidth of 'do I really recognize it [the aircraft]'".[2]

6.3 Research Question 3: Benefit of Hot-Spot Only Fusion

In the preceding discussion, it was deduced that neither the visual nor the infrared panorama enables the best detection or recognition performance in all weather conditions. The visual panorama is superior in low-light and low-visibility conditions—specifically during night and in rain and low visibility—while the infrared panorama is superior in good weather, like the CAVOK and bright light conditions. Consequently, operators presented with the two single visual and infrared panoramas need to switch between them to achieve the best visibility in all weather and traffic conditions. A fused panorama is beneficial to the operators' work if it allows for a detection and recognition performance equal or very close to the performance of the best single panorama, as this would obviate the necessity to switch between two panoramas. It therefore frees mental resources that can be redirected to other tasks.

The objective benefit of the Hot-Spot Only fusion is visible in the relativized ODTs with the fused panorama. For nine of the 10 situations, the values are low with an average of 0.71 s (SD = 0.68). Hence, the Hot-Spot Only fusion is capable of combining the benefits of the single visual and infrared panoramas, effectively enabling the earliest possible detection regardless of the weather condition. The much higher relativized ODT in the remaining situation—arrivals in bright light—is explained by the fusion process used, which preserves the bright background of the visual panorama (compare Fig. 6) and fuses in the equally bright pixels from the infrared panorama. Hence, the fused-in pixels have no contrast against their surrounding and cannot be detected. This can be accounted for by an improved fusion algorithm that inverts the pixels from the infrared panorama if necessary.

---

[2] Original: "Große Bandbreite von 'erkenne ich ihn wirklich'".

An equal conclusion is drawn from the recognition difference between the fused panorama and the best single panorama (see Table 5). In this case, the operators' recognition performance with the visual or infrared panorama is on average just 0.56 s (SD = 3.46) better than with the fused panorama. Moreover, departures and arrivals in the CAVOK weather conditions are better recognized by the participants in the fused panorama, as already noted in Sect. 6.2. In summary, the panorama generated by the Hot-Spot Only fusion enables the operators to follow departures as long and to recognize arrivals as early as with the best single panorama. This is a clear benefit of the fusion approach as it obviates the necessity to switch between the visual and infrared panorama to perceive all relevant information.

The benefit of the Hot-Spot Only fusion is also evident in the illustration of the traffic patterns in Figs. 14 and 15: Information from the infrared panorama is used to fill the timeframes in which an aircraft is not visible in the visual panorama reducing the total timeframe in which the aircraft cannot be perceived. The resulting fused panorama consequently improves the detection and recognition performance compared to using only the visual panorama without the need to switch between two panoramas.

Within the study, only the startup of the second engine was identified as scenario in which a detection or recognition occurs later in the fused panorama than in both the visual panorama and the infrared panorama. In this particular case, a considerable parallax effect is visible which originates from the proximity to the camera set-up (see Fig. 17). The control points for image registration where chosen on very distant objects in order to avoid perturbation of the registration by parallax. Thus, no control points were available at such nearby locations. Accordingly, the registration errors tend to be larger on the field close to the tower. The information from the visual panorama consequently is obstructed by those from the infrared panorama, hindering the perception of the engine's startup. In general, the objective detection and subjective recognition of an engine's startup is dependent on additional factors, such as the engine type and viewing angle, compared to an arrival or departure. Hence, the turboprop engine's startup discussed in this work is an example pointing to potential problems (i.e. considerable parallax), but cannot be generalized.

6.4 Research Question 4: Subjective Evaluation

6.4.1 Perceived Usability

The participants rated the system's usability above the threshold value of 70, therefore indicating a good usability. The score is nevertheless lower than the very good usability reference score of the requirements test. However, while watching the video sequences, no human–computer interaction to control the fusion and the share of VIS and IR data was provided. The missing control options were mentioned by operators and their necessity was highlighted. So, the lower SUS rating in the evaluation compared to the reference can be explained by this missing control option. As one

aim of providing a fusion of VIS and IR panoramas was to reduce mental workload caused by switching costs, controlling the fusion and the human–machine interface therefore should not cause mental workload similar or exceeding these switching costs. Compared to switching between the VIS and IR representations, the control of the fusion is expected to be used less often and mainly for adjustments to the weather condition. Hence, it can be assumed that mental workload caused by adjusting the fusion will be lower than switching costs for current Remote Towers providing separate VIS and IR panoramas. Because the evaluation set-up was a demonstration prototype and not a fully functional working position, the operators' usability rating is a very satisfying result.

During the evaluation and in the debriefing, operators commented on perceived usability. The result from the debriefing are reported here: One general remark was that stitching of the camera pictures needs to be very accurate. The same applies to the fusion of both video streams. For the evaluation, the overlay was optimized for the horizon and runway. Because of the parallax effect, the apron area was slightly shifted and the images did not fit perfectly, as can be seen in Fig. 17.

A further result from the debriefing is, that compared to the VIS images, operators could detect aircraft better in the IR and consequently also in the fusion. The advantage of the fusion against the IR data is that operators found it stressful to observe IR images for longer time periods. Operators mentioned the high contrasts and the image noise in the IR data as stressors. They found it more difficult to judge the weather situation in the IR images. The fusion was thus perceived as usable; as it gives the advantage that monitoring feels less stressful but at the same time provides the relevant information from the IR data. Nevertheless, image noise from the IR in the fusion, as well as marginal time lags between IR and VIS data were perceived as distracting. In general, operators demanded a very high quality of the fusion in terms of matching data sources in time and place.

6.4.2 Perceived Situation Awareness

The Situation Awareness is an important human factor contributing to overall human performance. In the context of this study, it was assessed with four specific questions

regarding the impact of image fusion on the perception of relevant information. The operators rated their SA as above the neutral point of three, indicating that they somewhat agree with the statements. As SA is crucial, this rating should be improved.

Operators reported during the experiment that IR video images felt unnatural for them and that they had a better connection to the real-world situation with VIS pictures. This effect can be explained with inexperience of operators in interpreting IR video images. Operators are used to the view outside of the tower, which is captured by VIS-images and have a high experience in interpreting this view. The effect also applies to the fusion of both data sources, which is new to operators. This novelty can explain the suboptimal ratings for SA. While it is to be expected that the operators familiarize themselves with the fused images over time, improvements in certain areas are required, as outlined below.

The areas for improvements are derived from analysis of the single items, all of which are rated on average above the neutral point of the scale, meaning a tendency of operators to agree with the items. Controllers agree that the fusion enabled them to monitor departing aircraft long enough; the value for arriving aircraft is lower.

The runway 26 in Braunschweig is surrounded by a forest, which captures heat and has a signature in the IR images. This led to the fact that during the landing—shortly before touch-down—the warm aircraft could not be separated in the IR images from the forest. For operators the aircraft disappeared, reducing their Situation Awareness.

Nevertheless, during the experiment and the debriefing operators stated that in some weather conditions detection of aircraft was much better with information from IR video images. Furthermore, during night conditions, contours of runway, taxiways and aircraft were visible and perceived as very helpful. The lower ratings for items 1[3] and 4[4] may be due to the general effect of the remote tower concept. Detection of all relevant information is a critical point in the controllers' monitoring task and is highly influenced by replacing the OTW view by a technical system.

6.4.3 Perceived Trust

As fusing VIS and IR video streams increases the level of automation at the CWP, it is important to assess whether the operators have sufficient trust in the system. If trust is low, they will double check the results and as such the fusion will generate no benefit but additional taskload. However, ATCOs should not be complacent when using the fusion and be aware of potential artefacts generated by image fusion.

The overall score of the SATI questionnaire indicates that the operators often trusted the system. Furthermore, there is no indicator of complacency. The participants also rated it as very often understandable. While the robustness of the system received the lowest rating, it still indicates that operators often, but not always, felt that the system works robustly.

---

[3] "The fused panorama gives me a sufficient overview of the overall situation at any time, allowing me to easily locate relevant vehicle and aircraft movements.".

[4] "I would have been able to safely control the traffic with the fused panorama.".

During the evaluation and the debriefing, explanations for perceived trust were collected. Camera noise, especially from the IR video images, caused false targets which sometimes confused operators. Furthermore, within the IR and consequently the fusion, birds were highly visible and sometimes operators mixed them up with aircraft.

One drawback of fused videos is a small parallax between IR and VIS cameras, particularly with objects very close to the camera position, which cannot be prevented due to the fact that IR und VIS camera cannot be physically at the same position and thus results in different angles to the object. Furthermore, the cameras were not externally triggered, so the individual IR and VIS frames were not perfectly synchronized. Especially during take-off or landing, where aircraft move with high relative speed parallel to the image plane of the cameras, the fused pictures showed the aircraft not perfectly aligned.

Operators recognized these artefacts and found them problematic. Further experience with the fused images might mitigate this rating. More sophisticated technical set-ups with externally triggered framerates can also prevent some time delays. Nevertheless, imperfect synchronization is an issue and operators need to be informed about the sources and possible impacts on the image they use for their monitoring task.

6.5 Comparison with Theoretical Distances According to the Johnson Criteria

In addition to the four research questions, the applicability of the Johnson Criteria to the data collected in the study is briefly discussed. The detection and recognition distances derived from the ADS-B data are therefor compared to the theoretical distances calculated from the Johnson Criteria by using the two approaches described in Sect. 4.3. To ensure similar conditions to those in which the criteria are established, the distances of arrivals—whose visibility is not limited by the camera set-up's FOV—in the CAVOK weather condition and with the visual representation are used. The results are presented in Fig. 18 for a Dornier 328 with a critical dimension of 7.24 m.

The bars in the upper part of the figure represent the calculated ranges in which an object is first detected and first recognized according to Johnson (1985) and Bailey, Wilz and J. Arthur III (2012). In the lower part, the objective detection distance and the range of the participants' recognition distances are visualized. The objective detection distance is pictured as a tilted square while the recognition distances are displayed as a bar ranging from the minimum recognition distance to the maximum recognition distance.

As can be seen in the figure, the objective detection distance lies within the detection range determined with the method of Johnson, but is more than six kilometers higher than the theoretical range calculated with the method of Bailey et al. The
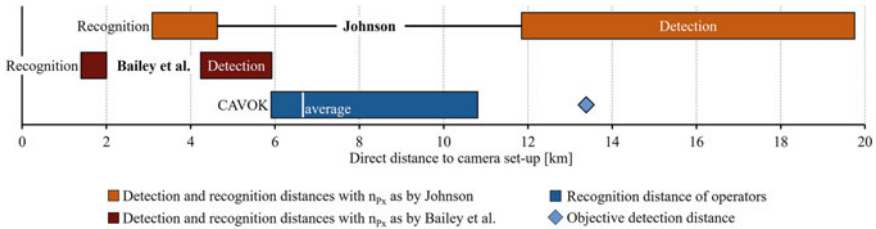
Fig. 18 Comparison of detection and recognition distances of VIS representation with theoretical values as by Johnson (1985) and by Bailey et al. (2012) for aircraft with a critical dimension of 7.24 m

recognition distances of the participants do not coincide with any of the theoretical detection and recognition ranges. The operators recognize the aircraft in higher distances than the range of detection distances calculated according to Bailey et al. and between the ranges of recognition and detection distances calculated according to Johnson.

In conclusion, Johnson's approach to determine detection and recognition distances from the Johnson Criteria is found to be partially applicable to the collected data. While it is accurate for the objective detection distance, Johnson's calculation scheme cannot be used to predict the recognition distance. The scheme used by Bailey et al. is not appropriate to predict both, the detection distance and the recognition distance, as it yields distances that are much lower than those measured in the study. Hence, the sole use of the Johnson Criteria is inappropriate to predict the detection and recognition ranges of a Remote Tower set-up. While there are newer revisions that also account for different weather conditions, any use in the Remote Tower context should be accompanied by an empirical evaluation.

7 Summary and Outlook

Remote Tower Operation is a concept which, at the moment, is of high interest for air navigation service providers (ANSPs) worldwide. The main challenge is to provide operators with the relevant information they need. A video panorama solely based on VIS camera sensors has physical limitations, so it could be thought of implementing additional sensors like IR cameras to fuse them for better integration and thus to enlarge the performance spectrum.

In this chapter, the user-centered design process for a fusion of VIS and IR camera data and main results are described. First, requirements were developed together with the end-users, leading to a novel fusion approach, called "Hot-Spot Only"-fusion. Second, this approach has been be realized by a series of image processing principles. And third, the result was systematically evaluated with end-users.

A fusion of two video streams for air traffic control context should be as close to the VIS data as possible and should only highlight relevant objects. Following

this, there are also lessons learned for the image or video processing expert: (1) In particular in image fusion, most diligence and efforts should be spent on investigating the desired functionality, the user requirements, and the purpose in detail. Interviews and cooperation with a representative sample of prospective users helps most of all. The solution then can be a simple processing chain. (2) There are image fusion applications where a strong asymmetry exists between the two modes. One mode may be much more important, and thus the largest part of resulting image (e.g., over 99%) may just be copied from the more important channel. Yet, fusing in the remaining small parts will be an important step forward, justifying the efforts.

The benefit of the Hot-Spot Only fusion approach is determined from the objective detection performance and the participants' recognition performance with the visual, infrared, and fused panoramas. The visual panorama was found to perform best in adverse weather conditions while the infrared panorama is best suited for good weather. The fusion process is capable of combining both benefits, allowing for an operator's performance that is always as close as possible to the performance with the best single panorama throughout all weather conditions. Arrivals in the Bright Light weather condition were found to be the sole exception, explained by a low contrast between the VIS pixels and the fused in IR pixels.

For the design of Remote Tower Optical Systems, it is concluded that the addition of an infrared panorama based on uncooled cameras provides benefits only in certain situations—in particular not in adverse weather conditions—and may therefore be unnecessary, reducing the total costs of the system. In such cases where an infrared panorama is part of a Remote Tower, the proposed Hot-Spot Only fusion offers benefits as it ensures the best visibility of all relevant information while removing the need for operators to switch between two panoramas, therefore freeing mental resources.

The empirical results and qualitative feedback of this evaluation are discussed with regards to acceptance and the design of the fusion. Overall, results show that operators would prefer the fusion of IR and VIS camera streams against each single sensor. When VIS and IR video data is foreseen in a Remote Tower concept, the results of INVIDEON show that operators perceive a fusion as usable as it provides the best of both data streams. The fusion provides more relevant information than a single sensor and the level of automation is adequate in terms of trust in the system. As decisions made by operators are safety–critical, their requirements with regards to time delay of video images, image noise and physical effects like parallax are technically demanding. Nevertheless, it has been shown that the found INVIDEON solution has a satisfying usability and acceptance and should be further advanced.

With an approach as described in this paper, the benefit of technical solutions can be quantified and consequently can be a basis for air navigation service providers, when they define their Remote Tower CWP. Therefore, the test environment at the BWE and the partwise automated data gathering and preparation process was practical to gather representative weather situations. This data basis could also be used to further advance the fusion approach. For instance, the color scale for superimposing the IR data can be based upon the brightness of VIS data ensuring that fused-in IR pixels always contrast well with the underlying visual panorama. Furthermore,

METAR data could be analyzed, to pre-define parameters for the fusion, based upon visibility conditions.

References

Bailey, R. E., Wilz, S. J., & Arthur III, J. J. (2012). Conceptual design standards for eXternal visibility system (XVS) sensor and display resolution. NASA Langley Research Center, Hampton, Virginia.

Bavachan, B., & Krishnan, P. (2014). A survey on image fusion techniques. *International Journal of Research in Computer and Communication Technology, 3*(3), 48–52.

. Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. Taylor and Francis.

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies, 8*(2), 29–40.

Canga, E. F. (2002). *Image fusion*. Dept. Electronic and Electrical Eng., Univ. of Bath, Bath, UK.

Dehn, D. M. (2008). Assessing the Impact of automation on the air traffic controler: The SHAPE questionnaires. *Air Traffic Control Quarterly, 16*(2), 127–146.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64.

European guidance material on all weather operations at aerodromes—EUR Doc 013. 2016.

Firooz, S. (2005). Comparative image fusion analysis. In *IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition* (pp. 1–8).

Hagl, M., et al. (2018). Augmented reality in a remote tower environment based on VS/IR fusion and optical tracking. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics* (pp. 558–571).

Hanafy, M. E. (2014). *Characterization of the atmospheric effects on the transmisson of thermal radiation*. Dissertation, Michigan Technological University.

Hilburn, B. (2004). Head-down times in aerodrome operations: A scope study. Technical Report. Center for Human Performance Research (CHPR), Den Haag, The Netherlands.

Jagalingam, P., & Hegde, A. V. (2014). Pixel level image fusion—A review on various techniques. In *3rd World Conference on Applied Sciences, Engineering and Technology*.

Johnson, J. (1985). Analysis of image forming systems. In *Select papers on infrared design, part one and part two* (p. 761). SPIE-The International Society for Optical Engineering

Khidse, S., & Nagori, M. (2014). Implementation and comparison of image enhancement techniques. *International Journal of Computer Applications, 96*(4).

Krebs, W. K., & Sinai, M. J. (2002). Psychophysical assessments of image-sensor fused imagery. *Human Factors, 44*(2), 257–271.

Minimum aviation system performance standard for remote tower optical systems ED-240A change 1. Saint-Denis, France.

Moehlenbrink, C., Papenfuss, A., Friedrich, M., & Jipp, M. (2011). Monitoring behaviour of tower controllers. In: D. d. Waard, N. Gérard, L. Onnasch, R. Wiczorek, & D. Manzey (Eds.), *Human centred automation*. Shaker Publishing.

Mogford, R. H. (1997). Mental models and situation awareness in air traffic control. *The International Journal of Aviation Psychology, 7*(4), 331–341.

Naidu, V., & Raol, J. R. (2008). Pixel-level image fusion using wavelets and principal component analysis. *Defence Science Journal, 58*(3), 338–352.

Omar, Z., & Stathaki, T. (2014). Image fusion: An overview. In: *Presented at the Fifth International Conference on Intelligent Systems, Modelling and Simulation*, Langkawi, Malaysia, 27–29 January 2014.

OSED for remote provision of ATS to aerodromes. SESAR JU Deliverable D94. SJU2016

Papenfuss, A., & Friedrich, M. (2016). Head up only—A design concept to entable multiple remote tower operations. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, Sacramento, CA.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140–160.

Self, H. C. (1969). Image evaluation for the prediction of the performance of a human observer. In *NATO Symposium on Image Evaluation*, Munich, Germany.

SESAR HP repository: SATI—SHAPE automation trust index. https://ext.eurocontrol.int/ehp/?q=node/1594

Sjaardema, T. A., Smith, C. S., & Birch, G. C. (2015). History and evolution of the Johnson Criteria. Sandia National Laboratories, Albuquerque, New Mexico.

Stein, K., Gladysz, S., Seiffer, D., & Zepp, A. (2014). Atmospheric limitations on the performance of electro-optical systems: A brief overview. In *Laser communication and propagation through the atmosphere and oceans III*, San Diego, California, United States, Vol. 9224.

Taylor, J. H. (1969). Factors underlying visual search performance. In *NATO Symposium on Image Evaluation*, Munich, Germany.

Toet, A. (1989). Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters, 9*(4), 245–253.

Toet, A., Van Ruyven, L. J., & Valeton, J. M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering, 28*(7), 287789.

Valeton, J. M., & Bijl, P. (1995). Target acquisition: Human observer performance studies and TARGAC model validation. TNO Human Factors Research Institute.

Zhang Jun, L. W., Zhu, Y. (2011). Study of ADS-B data evaluation. *Chinese Journal of Aeronautics,* 461–466.

Planning Remote Multi-airport Control—Design and Evaluation of a Controller-Friendly Assistance System

Rodney Leitner and Astrid Oehme

Abstract A number of research projects aim at air traffic control independent from the controller's location and his outside view. In the context of one of these projects—VICTOR (Virtual Control Tower Research Studies), which was initiated by the German air navigation service provider Deutsche Flugsicherung (DFS), a completely new concept of Aerodrome Remote Control Center (ARCC) was investigated. In contrast to previous approaches, the ARCC-concept broadened the monitoring and controlling capabilities of the tower controller towards several airports at the same time. It thereby created new requirements for air traffic control, i.e. an eminent need for planning the air traffic flow of multiple airports. For this additional task the concept of a planning tool was developed, taking into consideration a user-centered approach, the guidelines for usable interfaces and a well perceived user experience. Following these Human Factors standards, our planning tool was developed to be useful and ensure safe handling, but also to look and feel good. For its evaluation, an analytical inspection method, i.e. heuristic evaluation, has been used as well as a questionnaire assessing the aesthetics of the graphical user interface. Eight usability experts assessed the tool, taking notes of any peculiarities and usability problems and carrying out the associated severity-rating. With the help of this methodology, 56 issues were identified and corrected. Furthermore, results from additional qualitative statements of the experts for development and optimisation of the user interface were subsequently used for re-design. In terms of looks, the planning tool scored above average in aesthetics ratings. This chapter briefly introduces the tool and its design, and subsequently focuses on our evaluation procedure and results.

Keywords Multiple airport control · Remote control · Air traffic planning · Usability · User experience · UX · Design · Evaluation · Human factors

R. Leitner (✉) · A. Oehme
HFC Human-Factors-Consult GmbH, Köpenicker Strasse 325, 12555 Berlin, Germany
e-mail: rodney.leitner@human-factors-consult.de

A. Oehme
e-mail: oehme@human-factors-consult.de

489

1 Introduction

While air traffic has continuously been increasing to 3.3 billion passengers in 2014 and is likely to more than double in the next 15 years (AIRBUS S.A.S., 2013; IATA, 2014), a growing fragmentation of the European airspace was identified as a major challenge as early as the late 1990s. Reacting to this obvious trend, the Single European Sky (SES) initiative initiated a reorganisation of the European airspace based on traffic flows instead of national boundaries and proposed additional measures for air traffic management to achieve key objectives: improving and increasing safety, enhancing efficiency and integrating air traffic control services.

At the same time, an intensified liberalisation in Europe had an impact on its air traffic management as well: The air navigation services regulation (the service provision Regulation (EG) 550/2004) opened air navigation services in European states for additional providers. SES regulation as well as SES's focus on efficiency have increased both cost pressure and competition at regional airports and require new, innovative air traffic control (ATC) concepts to tackle multiple challenges. Many air navigation service providers (ANSPs) focus on cost efficiency and have introduced assistance systems and automation to further minimise personnel expenses. In addition to these efforts, several international projects have attempted to realise an ATC workplace independent of location and weather by including a synthetic outside view to increase control capacities at airports.

This chapter reports on the development and first evaluative steps of an assistance system, which serves a novel concept of operations for regional airports. Addressing unique issues of these airports, such as highly inhomogeneous traffic density, the system is conceptualised for a new kind of controller working position. The chapter focuses on the user-centred design process we followed during system realisation and especially dwells on an expert evaluation carried out during prototyping. However, as a start we give a short introduction on the operational concept the assistance system is designed for and its origination background, before focussing on user requirements towards the system and the evaluation process.

2 Multi-airport Control

Traffic density at regional airports fluctuates highly and depends on a series of factors like time of the day and weather conditions. Except for the usual peak times in the morning hours and in the late afternoon/early evening, traffic density usually is very low. In particular at smaller airports, this uneven capacity utilisation decreases efficiency. Furthermore, the tasks of an air traffic controller (ATCO) are reduced to tediously surveilling the airport ground and the respective control zone. One way to implement a more even distribution of workload thus is to bundle the controlling activities of ATCOs in one control centre especially during phases of expected low traffic density. A range of projects follow this approach and work towards an ATC

that is independent of outside view. In general, the realisation aims at substituting the outside view. Sensor-based data, which complement video information and provide a clear view on the air traffic area and the nearer surroundings, can overlay the displayed output and support the ATCO with supplemental information.

It is this development of remote tower services that has been supported in the frame of the SES ATM research programme of the public–private partnership SESAR Joint Undertaking and that has helped to realise the first remote tower prototype in Sweden. The project Advanced Remote Tower (ART) established an ATCO working position independent of outside view and location outlined above and was realised with the Saab Remote Tower System (r-TWR). This concept is limited to remote control of one single airport, which is why we proposed an expansion of the control towards several airports applying a so-called Aerodrome Remote Control Centre (ARCC) (Oehme & Schulz-Rückert, 2010; Oehme et al., 2013). Obviously, this approach requires an altered, novel operational concept, the development of a new working position and the development of novel controller-assistance systems. We will sketch this operational concept that is used e.g. in VICTOR (Virtual Control Tower Research Studies) in the following paragraphs.

2.1 Concept Behind VICTOR

VICTOR was conducted on behalf of the German air navigation service provider DFS within the German aviation research programme LuFo IV. Its Concept of Operations (ConOps) envisioned two controller working roles: a master controller (MC) and a remote controller (RC) (Oehme et al., 2013; Wittbrodt, 2012).

The RC's mode of operation differs from common controllers in one aspect only: The RC has to rely on the video- and sensory-based outside-view substitute provided, because an outside view of the tower is not available. There currently is no job position similar to the MC, which is why a detailed operational concept and appropriate assistance systems have to be developed. The new ATCO working position will offer the MC the opportunity to monitor several airports and to actively control one flight movement at a time. Thereby, the concept patently aims at increasing efficiency during low capacity utilisation. In case of rapid traffic increase and the accompanying increase in the MC's workload, one airport will be handed over to a RC. Consequently, the MC only controls the remaining airports and additional RC working positions will be opened, depending on situation-related demand in case of additional traffic increase. During decreasing traffic, the MC will eventually repossess the responsibility of the airports from the various RCs and the respective RC positions will be closed.

2.2 Necessity and Elements of a Planning Tool

Since traffic balancing and traffic flow management are demanding tasks, the assistance system used has to assist the ATCO by providing a favourable workload distribution and related attention allocation. Useful and accepted arrival and departure management systems are already available for single airport control (e.g. Bergner et al., 2009).

Compared to the role of current ATCOs, the MC's role contains newly defined role aspects and tasks. In addition to an ATCO'S monitoring and control, the MC has to carry out administrative and planning tasks. The planning tasks include sequencing the flight movements, rearranging those movements according to situational demand and organising RC positions by opening them up or closing them again. The MC has to carefully balance the total number of RCs and the respective airport they are responsible for. In this context, both economic and operational factors need to be taken into consideration in order to increase safety and efficiency. For these planning tasks, the MC needs a tool supplying the relevant information and thereby supporting the decision making. It should e.g. provide an overview of all movements so that the controller can analyse traffic movements and density in order to optimise sequences for controlling the movements one by one in case of overlaps.

Following this ConOps as a first basis, our assistance system supports the MC in these planning tasks. Relating to the novel working position 'MC' it is called MasterMAN (see Fig. 1).
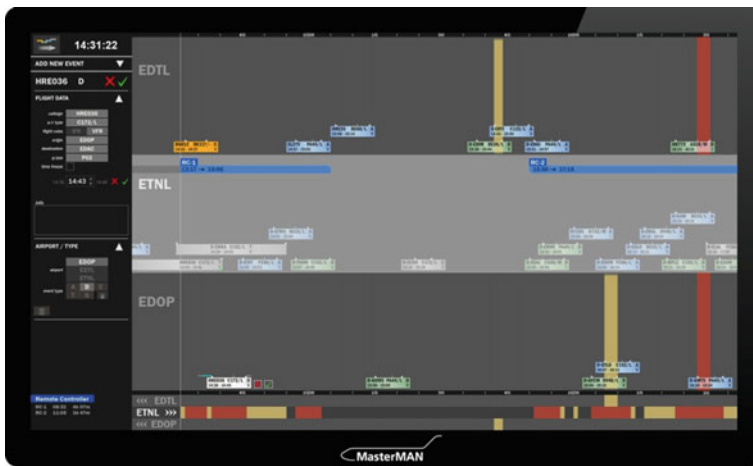

Fig. 1 User interface of MasterMAN

2.3   Requirements for MasterMAN Planning Tool

The success of new systems depends mainly on how well they suit the task they are built for and on the users' acceptance. Consequently, the approach of user-centred design was applied in the development process of the planning tool described here in order to allow for end users to be systematically integrated into the system development process (cp. Moser, 2012). Due to their professional expertise, ATCOs can detail best what expectations and requirements they have regarding technical devices for airport control.

Following this approach, we integrated ATCOs in the development process by conducting a focus group with them to identify their specific requirements for such a planning tool and to thus boost their acceptance of the new system. The results were transferred into a requirements matrix of three different fundamental classifications: functional requirements, data requirements and quality requirements.

Data requirements describe the information end users receive from the system and which is essential to their tasks and needs to be readily accessible, e.g. data for a specific call sign, aircraft type or flight. The functional requirements reflect the actions end users want to carry out using this information and result from the interaction of the user with the system. Quality requirements detail how these functional requirements have to be implemented (cp. Leitner and Jürgensohn 2014b) and significantly impact the usability of the system. In total, these user requirements form the basis for the development of the planning tool and are used to develop use cases, stipulate an information and interaction design, and develop a graphical user interface. A comprehensive overview on the planning tool is provided in Leitner and Jürgensohn (2014a), while Leitner et al. (2011) report on its related functional, data and quality requirements.

3   Usability/User Experience

3.1   Usability

It is obvious that a system high in usablity is easy to use, which is why user friendliness has increasingly gained importance for system development in the last decades. Usability criteria support developing systems with a user-friendly and ergonomic design (Sarodnick & Brau, 2011, p. 18).

The international standard *Ergonomics of human-interaction systems* (DIN EN ISO 9241-11:1999-01) describes usability as 'extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use'. The usage context includes the user, the task and the means to fulfil the task within the setting in which a product is used. Effectiveness is described as the accuracy and completeness with which the usage goals can be achieved, whereas efficiency is a measure of effective goal achievement in relation

to the rescources needed. Usability not only aims at the suitable/appropriate usage of a system but also sets the requirement that the system supports the user in reaching his/her goals in the respective field of application. Usability is a quality feature of products or systems and describes the goal of interface development in incorporating ergonomic findings (Sarodnick & Brau, 2011). A systematic assessment of user satisfaction can be carried out by applying a variety of usability evaluation methods (Lehr, 2011).

Designing and developing systems with a high usability holds many advantages. These systems are technologically and commercially successful with customers willing to pay more for this quality standard (DIN EN ISO 9241-210:2010-06, p. 8), mainly because usability increases user productivity and thus acceptance and on the provider side reduces expenses for customer support services and training.

Preceding the system development, the usage context has to be defined and subsequently user requirements have to be derived (c.f. Sect. 2.3). These are derived from needs, desires and conditions of the user and describe which goals a user wants to reach with a system.

3.2 User Experience

User Experience (UX) is a broad psychological and human-factors related construct that maps the perception and response of a person resulting from an actual or expected usage of a product, system or service (ISO 9241-210, 2010). According to Hassenzahl and Tractinsky (2006), UX is 'a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.)'. For these authors, a system comprises pragmatic and hedonic qualities with the former concentrating on a product's utility and usability in relation to potential tasks. The latter however focusses on the user, i.e. it addresses his/her feeling of so called 'be-goals' (e.g. 'being competent', 'being related to others', 'being special') and general needs (e.g. for novelty and change, personal growth, self-expression and/or relatedness) (Hassenzahl, 2008). Hedonic attributes of a product or service thus refer to the user being stimulated (personal development, new impressions), being able to communicate his identity to others (social recognition), and by the products' ability to evoke memories (Hassenzahl, 2003). They contribute directly to the core of positive experience, while pragmatic quality does so only indirectly by facilitating its fulfillment. In a similar fashion, the CUE-Model (Components of User Experience; Thüring & Mahlke, 2007) distinguishes between two qualities: Instrumental qualities refer to the experienced amount of support provided by the system and its ease of use (i.e., pragmatic qualities), while non-instrumental qualities address the look and feel of the system. Emotions elicited by the system use are a third component. These

three constituents result in the user's overall appraisal of the system and thus influence future decisions and behaviour, e.g. their decisions to use the system regularly, if at all, or their intention to migrate to a different system with potentially similar capabilities.

Despite this obvious major importance, UX is often not considered in the working context. In contrast to other approaches, MasterMAN embraces UX in an aesthetically appealing graphical design in order to facilitate stimulation, user acceptance, and ultimately usability.

4 Evaluation of MasterMAN

The development of the system's user interface should hold iterative testing intervals. Preferably, early and regular assessment detects initial shortcomings or even maldevelopments, so those can be remedied quickly and at low costs (Baumann & Lanz, 1998, p. 8f). Applicable methods range from formal-analytic methods (analytic task analysis, expert guidelines) through inspection methods (heuristic evaluation and walkthrough methods) to usability tests (inductive or deductive) and surveys using questionnaire (e.g. ISONORM, QUIS or SUMI) (Sarodnick & Brau, 2011). We evaluated MasterMAN at an early stage in the form of a *working partial system* prototype (Rosson & Carroll, 2002). This was an operative system version which featured the majority of the final system's functionalities in a so-called mixed prototype. While horizontal prototypes provide the entire, but only partially implemented set of functionalities of the human–machine interface, vertical prototypes are limited to fully operating parts of the system (Dumke, 2001; Sarodnick & Brau, 2011). Integrating the advantages of both kinds of prototypes was used to give early testers the impression of a complete system and to thus suggest its real-life performance in order to be able to include their feedback at later stages of the development process.

4.1 Test Design and Procedure

Besides trying to uncover undetected errors in functionality, the tests were also meant to collect feedback on MasterMAN's quality of experience and UX. As a *working partial system* prototype it featured all basic functionalities like event adding, editing and deletion as well as several additional functions such as time/clock setting or selecting airports and aircraft types via adaptive selection lists.

We decided to carry out a heuristic walkthrough (Moser, 2012) with human factors experts followed by a pluralistic walkthrough (Karat et al., 1992; Wilson, 2014), the latter usually being a group discussion including all system stakeholders, i.e. among others developers, users, usability experts, marketing. This evaluation procedure can quickly identify usability shortcomings, which then can be remedied promptly in the subsequent design process. In the Heuristic Walkthrough, the expert assessment was

conducted one at a time in order to receive independent results. The lack of domain specific knowledge was compensated by a thorough introduction of the participants to the usage context.

The basic functions of MasterMAN were at the scope of this evaluation phase. In order to make sure that the experts explore and test all functions, four comprehensive scenario tasks were prepared. Participants had to note detected usability issues in systematic categories linked to usability heuristics. These also had to be assessed on a severity scale. VisAWI (visual aesthetics of websites inventory by Thielsch & Moshagen, 2011) was used to access the visual attractiveness of the system to test for hedonic, non-instrumental qualities of UX.

In a final session, a pluralistic walkthrough was conducted, so that the evaluators could openly discuss their impression of the prototype and could propose improvements. As the heuristics used in the first walkthrough were the basis for the subsequent Pluralistic Walkthroughs and were pivotal in establishing the experts' first impression of MasterMAN, we will discuss their selection.

4.1.1   Selection of Heuristics

The sheer number of usability guidelines and rules is increasingly confusing for both developers and evaluators. For this reason, Nielsen and Molich (1990) have developed heuristics reflecting basic usability principles which can be applied easily during an evaluation. Heuristics support the evaluator with a categorisation of usability issues and indicate problem fields of an application (Nielsen, 1994). Besides the detection of usability problems, accumulating individual problems into broader, but considerably fewer categories leads to a comprehensive understanding of a system's shortcomings and helps to prioritise adjustments of interaction and graphical design.

The original list by Nielsen and Molich (1990) encompasses nine heuristics which were later amended by one additional heuristic as a result of numerous revisions and a factor analysis (Nielsen, 1994; Sarodnick & Brau, 2011). One could have applied these heuristics one by one to evaluate MasterMAN or could alternatively have substituted them with a more appropriate set. Since an extensive comparison of different usability heuristics by Somervell and McCrickard (2005) concluded that there are no significant differences between the various sets of heuristics and because they pointed out that a target-oriented pre-selection and self-developed heuristics might have a positive impact on the evaluation of an application, we selected an individual set of heuristics.

Usability expert Donald A. Norman's focus on man–machine interfaces as well as on everyday objects renders his interaction principles applicable in a larger context. Norman's principles offer a differentiated view on visual attributes of MasterMAN. In addition to Norman's heuristics, we adapted Shneiderman's heuristics (Shneiderman, 2002, Shneiderman & Plaisant, 2010) and the dialogue principles of the respective standard to avoid user problems (DIN EN ISO 9241-110:2008-09) to our purpose, because the applicability and usefulness of each principle can vary strongly and is context-based.

Following a comparison of all stated principles and taking into consideration the specific field of application (ATC), we selected the following heuristics:

- **Suitability for the task**
  The principle is suitable for the task if an implementation of a dialogue enables the user to accomplish his task effectively and efficiently. For this criterion, among others, an emphasis of task-relevant information and a reduction of non-task-relevant information to a minimum would be useful.
- **Conformity with user expectations**
  Compliance to generally acknowledged conventions (DIN EN ISO 9241-110:2008-09) and a certain level of predictability are expected from a well-designed human-machine interaction. This also includes vocabulary the users are acquainted with well. As these expectations differ depending on the user group, establishing a consistent dialogue based on the experiences, expectations and knowledge of the users is of prime importance.
- **Self-descriptiveness**
  To implement this ergonomic principle, a consistent and constant information flow indicating to the user at which stage of the working process she/he currently is has to be established. For example, upcoming working steps could be indicated until dialogue closure.
- **Visibility**
  Following Norman's (1988) definition, visibility describes the visible arrangement of control and other interaction elements. This means that users cannot make use of HMI elements which are not visible to the user, i.e. all context-relevant information have to be placed visibly on the software surface and overlaps or other visual disturbances must not occur.
- **Affordance**
  To prevent usability difficulties, all elements of the user interface should be designed (affordance) implying their respective use at a glance.
- **Clearly marked exits**
  This design principle is essential in order to hand control of any process over to the user. Consequently, it should be possible to exit as many user interface dialogues and interactions as possible. This heuristic encourages the user to independently explore the system, because the user can revoke unintentional actions and processes at any time and can effortlessly return to a former state.
- **Suitability for learning**
  This principle supports and guides the user in learning to use of the system adequately aiming at minimising learning efforts. As it in general is much easier for users to recognise visually than to recall the same information from memory (Nielsen, 1993), the system should provide dialogue elements and allow the users to choose.
- **Feedback**
  Feedback should be implemented in context-sensitive ways: In most low-persistence situations users will need feedback only during the process itself, while in other situations with medium persistence a confirmation may be required

of the user. Eventually, very important situations require continuous feedback, which hence has to be a substantial part of the user interface. In general, feedback should disappear automatically when it is no longer needed and its extent should be adjusted to the importance and frequency. It should inform the user about what the system is doing or what interactions are necessary especially for comprehensive and complex tasks. The visual presentation of changes can be an adequate feedback as well (Shneiderman & Plaisant, 2010).

- **Error tolerance**

  The system has to address two main points. First, user actions must not lead to system crashes or incorrect user inputs. Second, the system should support users in identifying and correcting errors. According to this heuristic, incorrect user inputs should be marked and a constructive feedback to correct the error should be provided.

- **Prevent errors**

  One main strategy of error control is attempting to design a fail-safe system that avoids error-prone situations (Nielsen, 1993). Asking users to reconfirm their actions before moving ahead can reduce the frequency of errors especially in situations with grave consequences. One can also adapt the options related to different operations, e.g. by providing radio buttons, shortlists or drop down menus to prevent the risk of spelling mistakes.

- **Good error messages**

  Error messages should support the user in solving critical situations quickly, effortlessly and reliably. To achieve that, the wording of the messages should be brief, clear, and comprehensible. However, the user should have facile access to a detailed explanation of the problem in the form of 'multiple-level messages': Instead of overloading user's cognition by putting all potentially useful pieces of information in one message, a combination of a short first message that upon user demand is replaced by a more elaborate message will allow for both quick reactions and detailed comprehension of the problem when necessary.

- **Controllability**

  A well-controllable system allows the user to influence the progress of a task process regarding direction and speed (DIN EN ISO 9241-110:2008-09). Each intervention should be available independently at all times and offer options to correct preceding interactions.

- **Suitability for individualisation**

  Users should be allowed to adapt the interface design to agree with their personal preferences, needs, tasks, working conditions, and skills. For common users, this in particular means individually defined shortcuts in order to reduce the number of interactions with the system and to increase speed. Experienced users often profit from using abbreviations, shortcuts and hidden macros.

The heuristics *visibility*, *affordance* and *clearly marked exits* can be subsumed under the heuristic *self-descriptiveness* (DIN EN ISO 9241-110:2008-09). However, the granularity of this heuristic should be increased which is why these three have explicitly been set out during the evaluation. The same holds true for the heuristics

Shneiderman	ISO 9241-110	Nielsen	Norman	Heuristics for the evaluation
	Suitability for the task			Suitability for the task
Consistency	Conformity with user expectations	Speak the users language, Consistency	Consistency	Conformity with user expectations
	Self-descriptiveness	Simple and natural dialogue		Self-descriptiveness
			Visibility	Visibility
			Affordance	Affordance
		Clearly marked exits		Clearly marked exits
Reduce short-term memory load	Suitability for learning	Minimize user memory load		Suitability for learning
Offer informative feedback		Feedback	Feedback	Feedback
Offer simple error handling	Error tolerance			Error tolerance
		Prevent errors		Prevent errors
		Good error messages		Good error messages
Permit easy reversal of actions	Controllability			Controllability
Enable frequent users to use shortcuts	Suitability for individualisation	Shortcuts		Suitability for individualisation

Fig. 2 Comparison of usability principles and overview of the heuristics used

prevent errors and *good error messages* in the field of *error tolerance*. Figure 2 provides a summary and comparison of the usability heuristics described above.

4.1.2 Severity Rating

The heuristics described above constitute categories of design recommendations/principles. Within these categories, problems are likely to be created during user-interface development, e.g. a system may lack comprehensive and meaningful error messages. This is why human factors experts are asked to assess during system exploration whether the recommendations are met. Instances where principles are reneged on highlight usability problems. Within our study, all problems were recorded on specifically designed documentation sheets. The evaluators were encouraged to think aloud during system exploration and their output was recorded and reviewed by a trained examiner.

Table 1 Categories of severity rating adapted from Nielsen (1993)

Severity rating (SR)	Desciption	Meaning
1	Cosmetic problem	Solving the problem if additional resources are available
	There is no interference of the functionality	
2	Minor problem	Solving the problem if it is often mentioned
	Problem is avoidable	
3	Medium problem	Solving the problem should be implemented
	User notes the problem and gets used to it	
4	Major problem	Solving the problem is urgently necessary
	User has big problems with accomplishment of the task	
5	Disastrous problem	Solving the problem is compulsory
	User cannot accomplish the task	

Usually it is not feasible to eliminate all detected usability problems during the subsequent system development phases. Therefore, problems were not only categorised by topic but also prioritised, i.e. the evaluators also provided a severity rating for each detected problem (Nielsen, 1993). The severity rating took various criteria into account, such as in how far the problem would impair task completion, how frequently the problem occurred or in how far it impacted the further working process.

The severity rating for any problem discovered in this process is a subjective assessment by the respective evaluator and is not necessarily reliable. Nielsen (1993) therefore recommends to not rely on the ratings of a single evaluator. Following this approach, we ensured that each evaluator assessed the system independently and we subsequently aggregated these individual assessments in order to increase the validity of the evaluation (Table 1).

4.2   Test Preparation

4.2.1   Sample Description

Carefully selecting test participants is crucial to obtaining relevant, objective results (Tullis & Albert, 2008). Interestingly enough, the number of evaluators needed for a study also influences the quality of the results. While basically a single evaluator should suffice, various investigations have concluded that this setting cannot identify most usability problems and fails to detect between 70% (Tullis & Albert, 2008, p. 119) and 65% (Nielsen, 1993, p. 156) of heuristics violations. Woolrych

and Cockton (2001) identified intra-personal and external factors to influence the evaluator's detection rate to a large extent. Put plainly, different evaluators uncover different problems, which renders aggregating several evaluators' assessments worthwhile. Especially complex evaluation objects require several evaluators (Tullis & Albert, 2008, p. 118f.). Tan and colleagues (2009) reported an asymptotic trend of detected problems with approximately seven to eight evaluators, i.e. the amount of usability problems detected increases degressively with the number of evaluators, resulting in the rule of thumb that five evaluators (magic number 5) are sufficient in order to uncover more than 80% of problems (Nielsen & Landauer, 1993; Tullis & Albert, 2008; Virzi, 1992).

In addition to the number of evaluators, their expertise plays an important role in problem detection (Karat, 1994, p. 224). Nielsen (1992) investigated three groups at different levels of expertise in the usability domain: novices, regular (usability) evaluators, and double experts. The latter additionally held domain expertise, i.e. they were not only experienced in usability but also in the respective field of application the user interface was to be used in. The novice evaluators unsurprisingly held the lowest detection rate for usability problems with an average of 22%, followed by the regular experts at 41% and the double experts at 60%. Taking into consideration that sometimes there simply are no double experts for a combination of domains and that double experts usually are scarce and expensive, regular usability and human factors experts are the commonly used, reasonable and most suitable alternative (Karat et al., 1992).

For the heuristic evaluation described here, we recruited eight evaluators (five male), who had an average professional experience in human factors of 10.1 years (range: 2–25 yrs). For this sample size Nielsen and Landauer (1993) estimate a detection rate for usability problems of 85–99%, which we deemed highly suitable for our purposes. The youngest participant was 27, the oldest 59 years old (Ø 40.6 yrs). Half of the sample had a professional background in psychology and the other half in engineering. One participant was an aviation engineer and held a private pilot licence; he thus accounted for a double expert.

The evaluators' affinity for technology was assessed using TA-EG by Karrer et al. (2009). The questionnaire consists of 19 items with a 5-level Likert-scale (1 = 'strongly disagree' to 5 = 'strongly agree'). Overall, the evaluators stated a high competency and a slightly positive attitude towards technology (see Fig. 3).

4.2.2 Development of Traffic Simulation

Screen Displays

A proper heuristic evaluation of MasterMAN required a traffic simulation, supporting the basic functions necessary for carrying out ATCO planning tasks. In order to reach a substantial level of reality, a simulation unit of a working position was developed. It included the planning tool itself and additional screens on which aircraft movements at an airport and its immediate airspace were visualised.
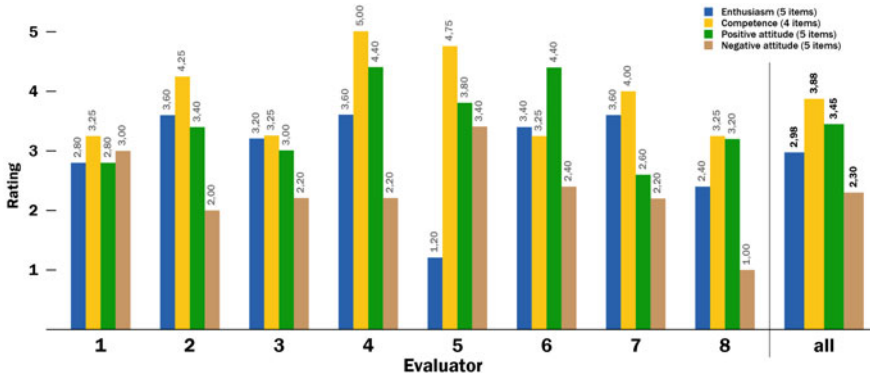
Fig. 3 Evaluators' affinity for technology

The flight movements were based on the flight plans of three regional airports, thus realistically simulating real-life situations. They included Schwerin-Parchim (EDOP), Rostock-Laage (ETNL) and Black Forest Airport Lahr (EDTL). For displaying the movements, several views have to be provided to the ATCO, which closely resembles a working position in an Aerodrome Remote Control Center. Since the simulated working position is independent of location and tower view, the view outside the tower window (e.g. on the taxiway) does not have to be emulated; however, the controller has to be enabled to observe and control movements on the taxiway. The overview of the airport and its operational airfield was realised via planar top view display (planport depiction) in a schematic way. The views of all three airports were scaled to include the runway(s) and landing strip(s), all taxiways, the apron, as well as the immediate vicinity of the airport. The views displayed on the monitor used for evaluation comprised a width of about 12 kms in reality.

Besides the planport views of the three airports, the respective radar views were required. A radar display already is a fundamental component of an ATCO's working position, both for tower and centre controllers. It will remain being fundamental in remote control. In our test set-up, the radar displayed a width of about 80kms and provided the whole area of responsibility (control zone CTR) of the 'controller'. Further areas covered were the broader periphery of the airport with prominent points for orientation such as beacons, villages, motorways, lakes and rivers, which are required for small aircraft navigating via visual flight rules.

Flight Routes and Taxi Movements

In order to be able to simulate and display flight movements on radar and planport view, movement paths were required and accordingly realised. Since the main focus was on evaluating MasterMAN's planning tool, simplified approach and departure paths were generated for the radar display. The radar movement paths extended the runway view beyond the immediate control zone of the respective airport and

followed one of three different directions up to the simulated border of the generated radar display. In addition to these general paths, three crossing movement paths were prepared for the radar screen to simulate crossing traffic in the respective scenarios.

For the planport view, further movement paths were needed in order to display landing and departing aircraft, but also aircraft taxiing on the airfield. Three different movement paths were generated, each of which simulated an aircraft land and finally taxi to one of three pre-defined, real-life parking positions. Additionally, three take-off paths were prepared, which started at one of the parking positions each and went via taxiways to the take-off position, where they initiated the take-off. Finally, we also included movement paths for traffic circuits, which are operated frequently at smaller airports.

For all airports a total of 54 movement paths were defined to simulate realistic aircraft movements on the two displays. Landing aircraft received an approach, landing and taxiing path each and departing aircraft correspondingly received a taxiing, take-off and departure path. The paths could be combined randomly in order to simulate numerous varying flight movements. A change in the runway 's operational direction was not necessary for our test purposes and was fixed at the beginning of the test.

Task Scenarios

The tests comprised of the two test procedures heuristic walkthrough and (sample-wise) reduced pluralistic walkthrough. Evaluation tasks for a total of four scenarios were defined for the heuristic walkthrough. For each of these evaluation tasks, the action steps necessary to fulfil the task effectively and efficiently were carefully defined, so that a deviation from the action steps provides clear indication of usability issues.

The evaluation tasks were furthermore classified according to their difficulty. Tasks low in difficulty mostly comprised of a small number of action steps while a considerably larger number of action steps was usually attached to difficult tasks. Because the system was novel to the evaluators, we arranged the tasks in the scenarios so that the complexity of the tasks continuously increased during the evaluation session (cp. Table 2). Since the tasks were consecutive, each completed scenario resulted in learning effects thereby enabling evaluators to realise even complex tasks consisting of several steps. Additionally, some task types were repeated in the follow-up scenarios (tasks marked in light grey in Table 2) to further deepen these learning effects.

Materials

Each of the evaluators received the following material:

- task instructions for the four scenarios,

Table 2 Tasks for the evaluation

Scenario 1	• Postponement of an event and confirmation • Cancellation of an event • Undo of postponements • Immediate handover of one airport to a remote controller • Immediate takeover of one airport of a remote controller
Scenario 2	• Independent solving of time conflicts • Creation of a new event • Modification of a data set of an event and confirmation • Cancellation of an event • Creation of a new event
Scenario 3	• Creationof a non relocatable event • Planning of an handover of one airport at a specific time • Runway closure with immediate effect • Cancellation of an event • Creation of a new event • Independent solving of time conflicts
Scenario 4	• Planning of a runway closure at a specific time and independent solving of conflicts by postponing events • Performing a manual optimization by postponing events and handover of one airport if necessary • Modification of a data set of an event and confirmation • Cancellation of an event • Creation of a new event

- note pads for each of the ten heuristics for writing down usability problems discovered,
- a questionnaire on demographic data,
- a questionnaire on visual aesthetics, and
- a questionnaire on affinity for technology.

4.3 Conduction

In preparation of the evaluation, a basis for a common understanding of the usage context of MasterMAN's planning tool and of the evaluation goal was established by giving the evaluators a short overview of VICTOR as well as detailed information on multi-airport control and working procedures of a MC. Additionally, a detailed introduction was given on the graphical user interface and the functionalities of the planning tool as well as the available controller assistance monitors (e.g. the planport). The evaluators were encouraged to ask comprehension questions before the evaluation started.

The predefined tasks for the first evaluation step (heuristic walkthrough) included working instructions, which were designed to consecutively lead evaluators through the system functions. Thus, evaluators learned about the planning tool in a stepwise

manner and used each function at least once. The tasks did not have a time limit, so evaluators were completely self-paced and able to note all conspicuousnesses using the defined heuristics.

An examiner attended the evaluation and documented the evaluators' task-based usage paths whenever evaluators departed from a predefined ideal path. After experiencing the planning tool based on the tasks and scenarios, evaluators had the opportunity to investigate and assess parts of the system in detail. No time limit was set and evaluations took 70–120 min. Finally, data on demography, visual aesthetics and affinity to technology were gathered.

The second evaluation step (pluralistic walkthrough) was planned as a group discussion amongst evaluators and developers in order to scrutinise discovered usability problems. The issues were categorised into 'unique problems' and 'shared problems', where the latter were usability problems discovered by at least two evaluators. The group discussion started with these shared problems regardless of how their severity had been rated. MasterMAN was used live to reproduce each problem and to display it on the spot, which allowed for collecting severity ratings even from evaluators who had previously not experienced said problem. In addition to all shared problems, all unique problems with severity ratings of four and above were assessed and discussed. In the discussion, first attempts for solutions were established.

4.4   Results and Analysis

The eight evaluators found a total of 58 usability issues, 15 of which were shared problems. In Fig. 4, those shared problems are marked with a black square with the difficulty of detecting a problem increasing from left to right and the evaluators' ability to detect usability problems increasing from top to bottom. Figure 4 provides an overview on the number of unique and shared problems as well as the evaluators' mean severity ratings. The evaluators discovered between 6 and 14 problems. Their average severity rating ranges from 2.17 to 3.79 (Fig. 5).

Figure 6 depicts the 13 heuristics used for evaluation as well as the number of problems discovered by the evaluators. Evaluators identified no issues in the heuristics *good error messages* and *suitability for individualisation* and just one issue respectively in the heuristics *clearly marked exits* and *suitability for learning*. Most of the issues were found in the categories *conformity with user expectations*, *suitability for the task*, and *feedback*. Especially considering the problems detected in these categories, a subsequent re-design of the user interface is mandatory. Additionally, all problems with high severity ratings, i.e. of four and above, will have to be reviewed in detail. Solution approaches for these problems, which already have been developed during group discussion have to be substantiated further and will be implemented in the user interface accordingly.

Aesthetics is a central element of UX and influences, amongst others, usability (Moshagen et al., 2009) as well as user satisfaction (Lindgaard & Dudek, 2003). We assessed layout aesthetics via VisAWI questionnaire on the subscales *simplicity*
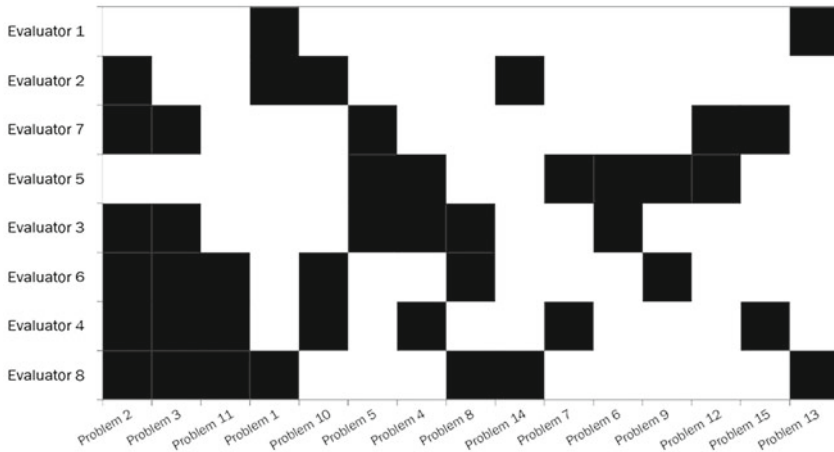
Fig. 4 Distribution of shared problems. Ability to detect usability problems is plotted for individual evaluators versus difficulty of problem detection. Black squares depict detected shared usability issues
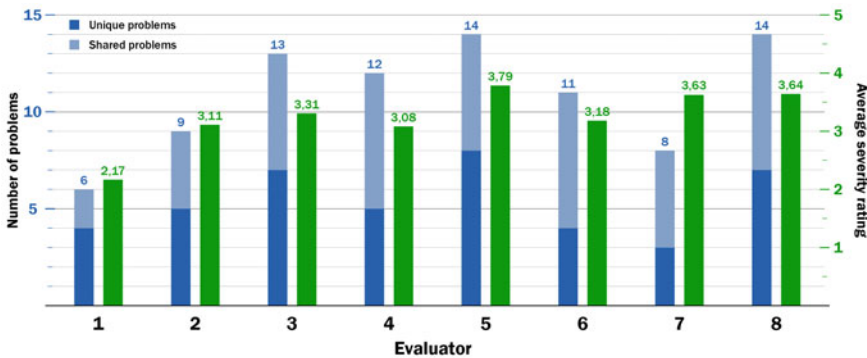
Fig. 5 Number of *unique* and *shared problems* and mean severity rating

(clearliness and structuredness), *diversity* (inventiveness and dynamics), *colourfulness* (color composition, choice and combination), and *craftsmanship* (topicality, sophistication and professionalism of design) as well as the overall layout impression the assistance system makes.

Figure 7 provides the mean values for the four subscales and the overall rating. In each case two of the subscales consist of four or respectively five items. According to VisAWI threshold analysis (Hirschfeld & Thielsch, 2015) a user interface is perceived positively with an average overall rating of 4.5. For the planning tool evaluators assigned ratings of 5.11–6.56 (Ø 5.77) (cp. black bars in Fig. 7), which is well above the established threshold. Thus, we assume there is no demand for action regarding the aesthetics of the assistance system's user interface.
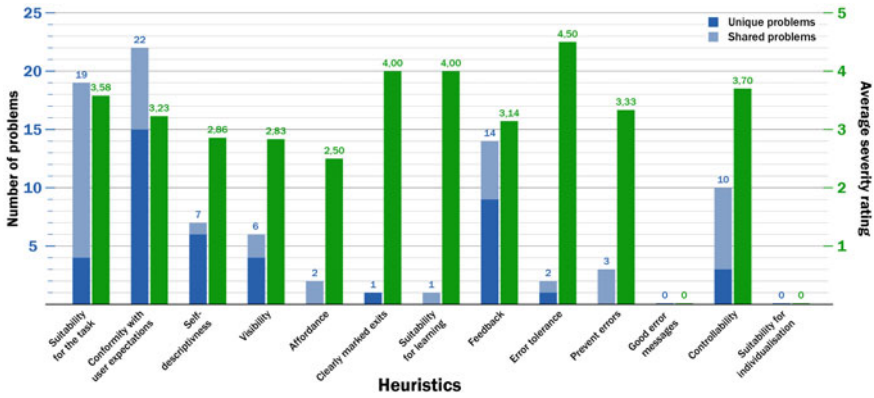
Fig. 6 Number of *unique* and *shared* problems as well as average severity rating of the heuristics
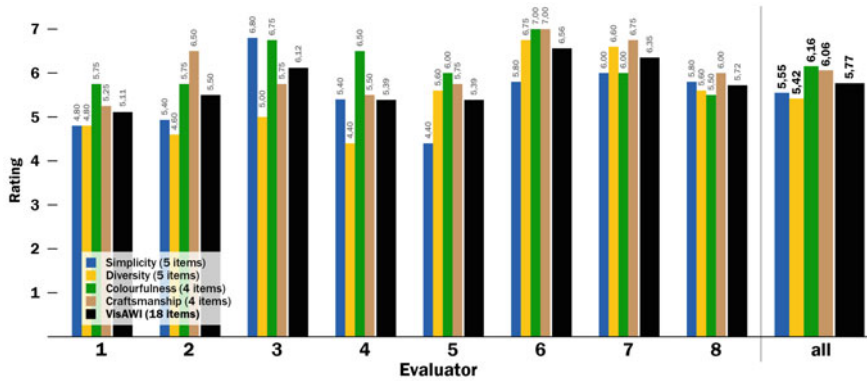

Fig. 7 Mean values of the four VisAWI-subscales and mean overall rating of all evaluators

5 Summary and Outlook

A user-centred development process demands an integration of user needs as well as further stakeholder and expert assessments in order to finally achieve high user acceptance, usefulness and usability. In following this approach, we collected user requirements in a first step, which built the basis for the system and interface design of MasterMAN's planning tool. After implementing the respective graphical user interface and basic functions, the usability evaluation reported here was carried out in order to discover user-system interaction problems. For our purposes, heuristic evaluation once again proved to provide valuable and nuanced input for mandatory as well as optional re-design. Complemented with UX-related system assessment, we concluded a number of re-design approaches from this first step of evaluations. In this process, MasterMAN already scored well in providing good error messages

and clearly marked exits as well as being suitable for learning and individualisation. The overall aesthetic appearance was rated well above the set threshold and thus presumably supports a positive user experience.

In a next step, re-design measures will be implemented in a timely manner. Valuable input for this step was provided in the group discussion. Additional to the human-factors expert evaluation, user input from ATCOs will be collected in a simulator study. In this study, ATCOs will work on different scenarios, which represent common situations occurring during a day shift in an ARCC. The results of this study will likewise facilitate the development of the planning tool. In case the users deem comprehensive changes on the tool necessary, further re-design and evaluation steps will be initiated in order to develop a most useful and effortlessly usable system for a MC.

The work of ATCOs, especially at small and regional airports will change fundamentally within the next decade. With MasterMAN they will have the opportunity to control traffic flows of several airports in an integrated way and actively anticipate and plan a whole day of air traffic. A further step to planning and optimisation of air traffic is a precise personnel planning assistance, which includes deployment availabilities of ATCOs. The combination of these very different tasks is to date not provided by any assistance system. In terms of MasterMAN functionalities, this will be our next developmental step.

References

AIRBUS S.A.S. (2013). Global Market Forecast 2013–2032. Retrieved January 4, 2014, from http://www.airbus.com/company/market/forecast/

Baumann, K., & Lanz, H. (1998). *Mesch-Maschine-Schnittstellen elektronischer Geräte.* Springer.

Bergner, J., König, C., Hofmann, T., & Ebert, H. (2009). An integrated arrival and departure display for the tower controller. In *Proceedings of 9th AIAA Aviation Technology, Integration, and Operations Conference*, Hilton Head, SC, USA.

DIN EN ISO 9241-11:1999-01. Teil 11: Leitsätze an die Gebrauchstauglichkeit. Beuth-Verlag.

DIN EN ISO 9241-110:2008-09. Teil 110: Grundsätze der Dialoggestaltung. Beuth-Verlag.

DIN EN ISO 9241-210:2010-06. Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme. Beuth-Verlag.

Dumke, R. (2001). *Software engineering* (3rd ed.). Vieweg-Verlag.

EU. Regulation (EC) No. 550/2004 of the European Parliament and of the Council of 10 March 2004 on the provision of air navigation services in the single European sky (the service provision Regulation). Official Journal of the European Union No. L 96/10, 31.03.2004 (p. 10).

Hassenzahl, M. (2003). The thing and I: Understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology: From usability to enjoyment* (pp. 31–42). Kluwer.

Hassenzahl, M. (2008). User experience (UX): Towards an experiential perspective on product quality. In *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine* (pp. 11–15). https://doi.org/10.1145/1512714.1512717

Hassenzahl, M., & Tractinsky, N. (2006). User experience—A research agenda. *Behaviour & Information Technology, 25*, 91–97.

Hirschfeld, G., & Thielsch, M. T. (2015). Establishing meaningful cut points for online user ratings. *Ergonomics, 58*(2), 310–320.

IATA. (2014, 16 October). New IATA passenger forecast reveals fast-growing markets of the future. Press Release No.: 57. Retrieved June 6, 2015, from http://www.iata.org/pressroom/pr/pages/2014-10-16-01.aspx

Karat, C.-M. (1994). A comparison of user interface evaluation methods. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 203–233). Wiley.

Karat, C., Campbell, R., & Fliegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In P. Bauersfeld, J. Bennett, G. Lynch (Eds.), *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems* (pp. 397–404). ACM.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen - der Fragebogen TA-EG. *Der Mensch Im Mittelpunkt Technischer Systeme, 8*, 196–201.

Lehr, A. T. (2011). *Lernen am Computer. Reihe Technologie-gestütztes Lernen, Band 1-2011* (Steinbeis-Edition).

Leitner, R., & Jürgensohn, T. (2014a). Entwicklung einer grafischen Benutzeroberfläche zur Fernüberwachung mehrerer Flughäfen. In A. Wendemuth, M. Jipp, A. Kluge, & D. Söffker (Eds.), *3. Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten*, 25–27 März 2014, Magdeburg.

Leitner, R., & Jürgensohn, T. (2014b). MasterMAN - Assistenzsystem zur Fernüberwachung von Flughäfen im Multi-Airport-Betrieb. In Grandt, M., Schmerwitz, S. (Eds.), *56. Fachausschusssitzung Anthropotechnik der Deutschen Gesellschaft für Luftund Raumfahrt - Lilienthal-Oberth e.V.: Der Mensch zwischen Automatisierung, Kompetenz und Verantwortung*, 14./15.Oktober 2014, Ottobrunn.

Leitner, R., Oehme, A., & Schulz-Rückert, D. (2011). Planning multi-airport traffic control—requirements and design implications. *Reflexionen und Visionen der Mensch-Maschine-Interaktion—Aus der Vergangenheit lernen, Zukunft gestalten.-9. Berliner Werkstatt Mensch-Maschine-Systeme, Berlin, 5*(7).

Lindgaard, G., & Dudek, C. (2003). What is this evasive beast we call user Satisfaction? *Interacting with Computers, 15*, 429–452.

Moser, C. (2012). *User experience design.* Springer.

Moshagen, M., Musch, J., & Göritz, A. S. (2009). A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics, 52*, 1311–1320.

Nielsen, J. (1993). *Usability engineering.* Academic Press.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25–62). Wiley.

Nielsen, J. (1992). Finding Usability Problems Through Heuristic Evaluation. In Proceedings ACM CHI'92 Conference, S. 373–380. 3. -7. Mai, Monterey, CA.

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of ACM INTERCHI'93 Conference,* 24–29 April, Amsterdam, Niederlande (pp. 206–213).

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings ACM CHI'90 Conference*, 1–5 April, Seattle, WA (pp. 249–256).

Norman, D. A. (1988). *The psychology of everyday things*. Basic Books.

Oehme, A., Leitner, R., & Wittbrodt, N. (2013). Challenges of multiple airport control—Experimental investigation of a multiairport control concept. In D. Harris (Ed.), *Aviation psychology and applied human factors* (pp. 1–8). Hogrefe.

Oehme, A., & Schulz-Rückert, D. (2010). Distant air traffic control for regional airports. In *11th IFAC-Symposium: Analysis, Design, and Evaluation of Human-Machine Systems*, Valencienne, France (Vol. 11(1), pp. 141–145).

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: Scenario-based development of human-computer interaction.* Morgan Kaufmann.

Sarodnick, F., & Brau, H. (2011). *Methoden der Usability Evaluation* (2 Überarbeitete Auflage). Verlag Hans Huber.

Shneiderman, B. (2002). *User interface design* (3rd Auflage). (Deutsche Ausgabe). mitp-Verlag.

Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: Strategies for effective human-computer interaction* (5th ed.). Addison-Wesley.

Somervell, J., & McCrickard, D. S. (2005). Better discount evaluation: Illustrating how critical parameters support heuristic creation. *Interacting with Computers, 17*(5), 592–612.

Tan, W.-S., Liu, D., & Bishu, R. (2009). Web evaluation: Heuristic evaluation vs. user testing. *International Journal of Industrial Ergonomics, 39*, 621–627.

Thielsch, M. T., & Moshagen, M. (2011). Erfassung visueller Ästhetik mit dem VisAWI. *Usability Professionals,* 260–265.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics, and emotions in human-technology interaction. *International Journal of Psychology, 42*(4), 253–264.

Tullis, T., & Albert, B. (2008). *Measuring the user experience. Collecting, analyzing, and presenting usability metrics.* Morgan Kaufmann.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors, 34*(4), 457–468.

Wilson, C. (2014). *User interface inspection methods: A user-centered design method*. Morgan Kaufmann.

Wittbrodt, N. (2012). Bord-Boden-Kommunikation im Multi-Flugplatz- Betrieb. Dissertation, Technische Universität Berlin. http://opus4.kobv.de/opus4-tuberlin/frontdoor/index/index/docId/3588

Woolrych, A., & Cockton, G. (2001). Why and when five test users aren't enough. In *Proceedings of IHM-HCI 2001 Conference*, 10–14 September, Lille, Frankreich (Vol. 2, pp. 105–108).

The Certification Processes of Multiple Remote Tower Operations for Single European Sky

Wen-Chin Li, Peter Kearney, and Graham Braithwaite

Abstract The European Union project of Single European Sky initiated a reorganization of European airspace and proposed additional measures for air traffic management to achieve the key objectives of improving efficiency and capacity while at the same time enhancing safety. The concept of multiple remote tower operation is that air traffic controllers (ATCOs) can control several airfields from a distant virtual control centre. The control of multiple airfields can be centralized to a virtual centre permitting the more efficient use of ATCO resources. The current research was sponsored by the Single European Sky ATM Research Program (SESAR) and the ATM Operations Division of the Irish Aviation Authority. A safety case was developed for migration of multiple remote tower services to live operations. This research conducted 50 large scale demonstration (LSD) trials of remote tower operations from single tower operations to multiple tower operations for safety assessment by air navigation safety regulators in 2016. The provision of air traffic services at two airports at the same time utilizing innovative technological solutions from a virtual location by a single air traffic controller was the first of its kind in the world. The implementation of this innovative technology requires a careful balance between cost-efficiency and the safety of the air traffic control in terms of capacity and human performance. No safety occurrence was reported nor did any operational safety issue arise during the conduct of the fifty live trial exercises. Multiple remote tower operations show potential in air traffic services as an alternative to traditional Local Towers. The novelty and flexibility of the advanced technology allow regulators to be creative in adapting to fit safety regulations, also has the potential to fundamentally change the way operators provide ATS. The evolution of implementation of the innovative technology requires a careful balance between cost-efficiency and its potential impact on safety, capacity, and human performance.

W.-C. Li (✉) · G. Braithwaite
Safety and Accident Investigation Center, Cranfield University, Martell House, Cranfield, Bedfordshire MK43 0TR, UK
e-mail: wenchin.li@cranfield.ac.uk

P. Kearney
Chief Executive Officer, Irish Aviation Authority, 11-12 D'olier Street, Dublin, Ireland

Abbreviations

AMC	Air Movement Control
ANSP	Air Navigation Service Provider
ATCO	Air Traffic Controller
ATM	Air Traffic Management
ATS	Air Traffic Services
ASD	Aeronautical Services Department
Con-Ops	Concept of Operation
CWP	Controller Working Position
DAA	Dublin Airport Authority
DoT	Department of Transport
EASA	European Aviation Safety Agency
EFS	Electronic Flight Strips
FAA	Federal Aviation Administration
FHA	Functional Hazard Assessment
HCI	Human-Computer Interaction
HMI	Human-Machine Interface
IAA	Irish Aviation Authority
ICAO	International Civil Aviation Organisation
IDP	Information Data Processing
LAX	Los Angeles International Airport
LSD	Large Scale Demonstration
MRTO	Remote Tower Operations
NextGen	Next Generation
NSA	National Supervisory Authority
OIS	Operational Improvement Step
OTW	Out the Window
PSSA	Preliminary System Safety Assessment
PTZ	Pan-Tilt-Zoom
RDP	Radar Data Processing
RTC	Remote Tower Centre
RTM	Remote Tower Modules
SAA	Shannon Airport Authority
SAM	Safety Assessment Methodology
SES	Single European Sky
SESAR	Single European Sky ATM Research Program
SMC	Surface Movement Control
SMU	Safety Management Unit
SRD	Safety Regulation Division

SSA   System Safety Assessment
UN   United Nations
VCS   Voice Communications System
WP   Work Package

1   Introduction

The initial concept of remote tower operations was started by the research proposal of Virtual Control Tower over 20 years ago (Kraiss & Kuhlen, 1996). The paradigm of remote tower operation will allow Air Traffic Services (ATS) be delivered remotely without direct observation from a local tower. The emerging technology of remote tower operations developed slowly during the early stages but in recent times has taken a leap forward with some single airport virtual tower operations. Based on the concept of remote tower operations, Multiple Remote Tower Operations (MRTO) offer further solutions for cost efficiency of air traffic services for small and medium size of airports. The new technology will allow one Air Traffic Controller (ATCO) control two or more airports at the same time during low traffic volumes. ATCO's use Out the Window visualization (OTW) supported by Radar Data Processing (RDP), Electronic Flight Strips (EFS) and a communications network to provide air traffic services. The feasibility of controlling two airports in parallel was demonstrated successfully with a special focus on the visual attention of ATCOs and the Controller Working Position design (CWP) related to ATS task (Moehlenbrink & Papenfuss, 2011). It is likely that ATCOs' monitoring performance is influenced by the system design of Remote Tower Centre (RTC) and the performance of multiple tasks simultaneously would require the sharing of cognitive resources of the controllers. The concept of distributed cognition seeks to understand the structure of cognitive system and extends the application to encompass interactions between resources and information in the operational environment (Hollan et al., 2000). The motivations of this research are to understand the limitations of controlling parallel traffic at two airports by a single ATCO, to demonstrate how the implementation of the new technology impacts safety, capacity and human performance, and how to conduct safety assessment of MRTO in order to secure regulatory approval for operations.

This chapter consists with six sections. The first section is an introduction of initial concept on developing remote tower operations and the leap to multiple remote tower operations thanks to advanced technology. The second section is the evolution of remote tower operations which can be divided as single tower, contingency tower and multiple tower operations. The third section is the demonstration of live exercises which can be separated as three batches involving 500 aircraft. The fourth section is the findings of live exercises related to air movement, surface movement, air movement plus surface movement on single aerodrome and multiple aerodromes. The fifth section is the discussion of innovative technology impacts to human–computer interactions, safety, capacity and human performance. The final section

is general discussion and recommendations on multiple remote tower operations and certification.

2 Background of Policy and Practice

The current Multiple Remote Towers project was sponsored by the Single European Sky ATM Research Program (SESAR) and the ATM Operations Division of the Irish Aviation Authority (IAA). The RTC was located at Dublin Air Traffic Services Unit in excess of 100 nautical miles away from the two airports at Shannon and Cork where the services were provided simultaneously (Fig. 1). Cork airport is a H24 international airport with aircraft types up to medium weight category such as Boeing 737 and Airbus 320. Total movements in 2016 were 50,242. Shannon is a H24 international airport with aircraft types up to the heavy weight category such as Airbus A330, it handled 25,059 movements in 2016 (Irish Aviation Authority, 2017). This research will contribute to the objectives for in sequence and simultaneous remote provision of ATS for multiple aerodromes as outlined in the Operational Improvement Step (OIS) SDM-0205 linked to SESAR Work Package (WP) 06.09.03 of the EU ATM Master Plan.


Fig. 1 The multiple remote tower operational centre equipped with PTZ, EFS, RDP and OTW located at Dublin Airport provided air traffic services for both Shannon and Cork airports simultaneously by SAAB digital tower systems

2.1   The Evolution of Remote Tower Operation

Air traffic in Europe has constantly increased since the 1990s. The Single European Sky (SES) initiated a reorganization of European airspace based on traffic flows instead of national boundaries and proposed additional measures for air traffic management to achieve key objectives of enhanced efficiency and capacity while improving safety performance. SES regulations focussed on efficiency, capacity and safety have increased cost pressure on air navigation service providers and require them to be more innovative in their approach to the provision of air traffic management services. Many Air Navigation Service Providers (ANSPs) have developed automated systems using video-panorama cameras for synthetic outside view, to increase capacity at airports and to improve cost efficiency by minimising personnel to meet cost efficiency targets (Leitner & Oehme, 2016). This has seen increased attention in remote tower research over the last 20 years. The concept of remote tower operations is that an ATCO can control any airfield from a distant virtual control centre. The view of the airfield under control is displayed in real time on screens and air traffic movements can be controlled. This concept is predominantly appropriate for the lower volume airports. Therefore, the control of multiple airfields can be centralised permitting capital costs savings. Consequently, the visual features of cues and objects which ATCOs shall be able to identify for safety operations are significant influencers of the requirements of surveillance cameras, the data-communication links and the display systems in a remote tower centre (Van Schaik et al, 2016). The concept of an advanced remote tower was developed for airports with fewer than 25 movements at the mean busy hours with a mix of visual flight rules and instrument flight rules (Saab, 2010). Technology advances can facilitate the image-video resolution for visual detection but not for recognition. The German Aerospace Center also has focused on the development enhancing video resolution to visual recognition on remote air traffic services (Friedrich & Mohlenbrink, 2013). The multiple remote tower research and implementation fall under SESAR Operational Step 3, the timeline shown as Fig. 2 (Irish Aviation Authority, 2016).

On the other side of the world, a review of air traffic control in the United States shows a rapid progress of technologies by FAA's Next Generation (NextGen) program which focuses on preventing aircraft collisions and providing a safe, orderly and expeditious flow of traffic. NextGen is concerned with the diverse aspects of tower control including human–computer interaction, situation awareness, cost of airport control tower, safety management and capacity variation (Federal Aviation Administration, 2012). Due to budget reduction measures under the sequestration cuts in the Budget Control Act, FAA closed 149 ATC towers at small airports and faced major financial constraints in the building and maintenance of control towers. Generally, there is a need to develop an innovative technology which will be able to provide alternative solutions to address such financial issues in ATM provision. The concept of remote tower operations has been addressed as a suitable solution and is being developing in many countries. NASA has also examined remote tower operations by studying alternative approaches for improving runway safety at Los
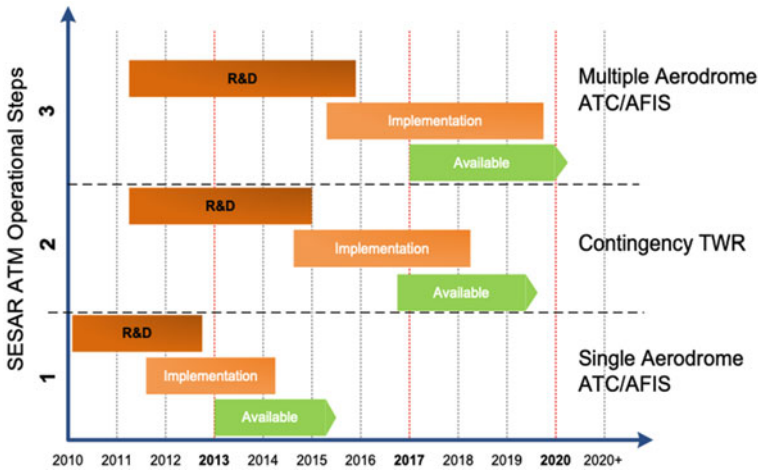
Fig. 2 The timelines of SESAR ATM operational steps for single aerodrome, contingency tower and multiple aerodrome ATC/AFIS

Angeles International Airport (LAX) under future flight central program (Dorighi & Rabin, 2002). The preliminary research demonstrated that remote tower operations can provide substantial economic benefits compared with traditional operations of local air traffic control centre, as NextGen proposed an innovative concept to address airport capacity problem by introducing an integrated tower information display providing weather and surveillance data as decision support tools (Nene, 2008).

2.2 The Cost Efficiency of Multiple Remote Tower Operation

The emergence of multiple remote tower operations is due in part to the changing operational environment in air transportation which had rapid expansion by low cost carriers at smaller airports. The cost constraints required ANSPs to develop new concepts and new technologies to fit the new business environment. Multiple Remote Tower Operations is an alternative solution to enhance safety and capacity at small/medium airports in a cost-efficiency manner. This new technology allows one ATCO control one or more small airports from a remoted location without direct visualization of the airport under their control (Fürstenau, 2016). Over 75% of regional airports with lower than one million passengers are currently making a loss. Cost of ATC services present the major portion of a regional airport's overall operating costs. The operational services of regional airports are similar, so the costs can be shared by relocation of the ATC function of two or more airports to a shared facility of multiple remote tower centre. The introduction of multiple remote tower paradigm is mainly driven by cost efficiency to reduce ATS operational cost. However, the SESAR assessment report of remote tower for multiple airports added additional

Table 1 The comparison of cost-efficiency between existing tower and remote tower

	Build	Equipment	Manpower
Existing Tower	Roughly cost £12M to Build. To assume 10% annual running cost for the building is reasonable £1.2M a year	Usual Communications, Navigation, Surveillance and Flight Data Processing Systems	Typical manning is 8 to 10 staff per H24 position
Remote Tower	Build costs will reduce significantly as only a mast needed to house the cameras. Estimated cost of mast £2M saving £10M To assume 10% annual running cost for the Mast is reasonable e.g. £200K a year saving £800K a year In summary if the tower is depreciated over 30 years, saving is (12–2)/30 = £333K in CAPEX, plus £800K in OPEX so £1.33M a year	Additional CAPEX is £2M. If the remote tower system is depreciated over 8 years, additional costs is 2/8 = £250K in CAPEX, plus £200K in OPEX so £450K a year There should be potential to save on some of the Communications, Navigation, Surveillance and Flight Data Processing Systems Costs via centralisation which will offset some of the increase in network costs	Remote Towers will facilitate staffing efficiency. The objective is to crew to workload such that operational staff are always busy within allowable safety limits For the IAA example of Cork and Shannon controlled from Dublin we anticipated a saving of 4 ATCO's or £400 K a year

safety specifications as requirements (Ziegler, 2017). IAA has conducted an analysis of the total costs of building and operating a physical tower by compared with the costs of remote tower. The result demonstrated that remote tower reduces costs significantly on the building the infrastructure and operational manpower 1.3 million per year (Table 1). Furthermore, Federal Aviation Administration (2012) revealed that the construction of a single control tower under federal contract might take three to five years with approximately 4.2 million dollars, plus the average annual operational costs and maintenance costs of $185,000 and several hundred thousand dollars for annual controllers' compensation.

2.3 Safety Assessment Methodology of Multiple Remote Tower Operations

The Safety Assessment Report for Multiple Remote Towers (SESAR Joint Undertaking, 2015) contemplated the availability of surveillance data to support ATCOs task performance in bad weather conditions (Ziegler, 2017). The IAA ANSP as sponsor and project coordinator is the ANS provider for Dublin, Cork and Shannon airports. The Dublin Airport Authority (DAA) as the airport operator for Cork Airport and Stobart Air, an international commuter airline. The Shannon Airport Authority (SAA) was involved as a stakeholder. A Demonstration Plan was prepared to describe how the live trial exercises would be organised, conducted, supervised, and assessed

focused on safety, capacity, cost efficiency and human performance. The safety case report is a structured argument, supported by a body of evidence that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given environment. It provides a comprehensive and structured set of safety documentation which is aimed to ensure that the safety of a specific system or equipment can be demonstrably safe. It will also establish the requirements for safety monitoring following transition into operation and for the entire life cycle of the system through to decommissioning (European Aviation Safety, 2015b; Agency, 2014).

For delivery of a safety argument for approval the IAA project team developed a "safety case" by applying the Eurocontrol Safety Assessment Methodology (SAM) to provide safety assurance that the introduction of any new technological systems or changes to these systems are proven to be tolerably safe for service provision. The safety assessment methodology of current research follows a structured step wise process as followings (Fig. 3);

(1) Safety Plan defines a safety programme that is planned, integrated and developed in conjunction with other design, development, production and quality control activities. It details safety activity timelines and deliverable in accordance with the higher project plan. It requires regulatory endorsement and approval.
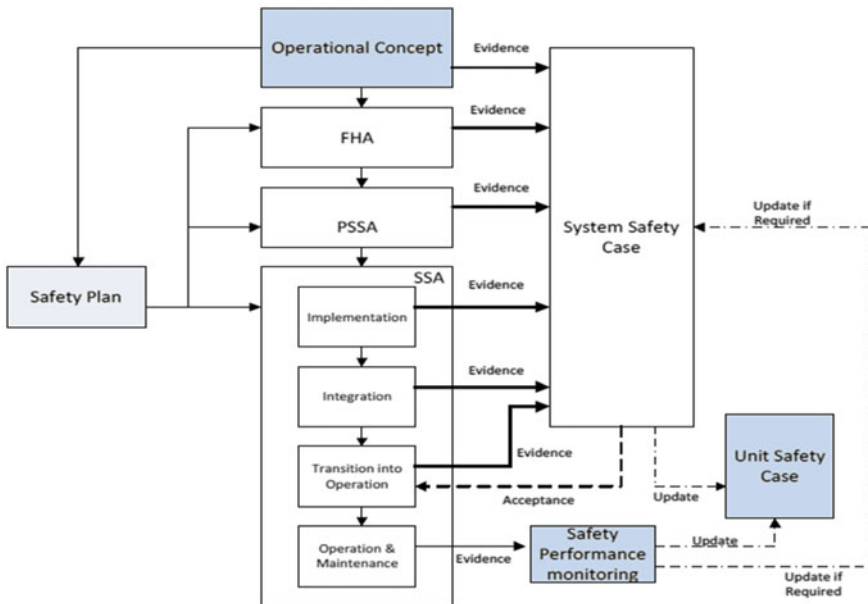

Fig. 3 Safety assessment methodology applied by IAA terminal services

(2) Functional Hazard Assessment (FHA) records the functions to be performed by the system, the effects of identified hazards on operations, including assessment of the severity of the hazards effects and also records the derived safety objectives, i.e. determines their acceptability in terms of the hazards maximum frequency of occurrence, derived from the maximum frequency of the hazards effects.

(3) Preliminary System Safety Assessment (PSSA) produces Safety Requirements and Assurance Levels for the system elements and records the evidence, arguments and assumptions to verify that the proposed solution will meet its Safety Requirements. It also provides the arguments to support the claim that the system will not affect the safety of the ATM system during installation and commissioning.

(4) System Safety Assessment (SSA) records the evidence, arguments and assumptions to verify and validate that the system design configuration will meet its Safety Requirements. It also describes specific operating and maintenance requirements necessary to assure safety and provides arguments to support the claim that the system will not affect the safety of ATM during the transition to operational use. In addition, the SSA provides details of the Transition Plan for introducing the system into service.

2.4 The Processes of Regulatory Approval for Practical Implementation

The regulatory body responsible for the regulation of aviation in Europe is the European Aviation Safety Agency (EASA) based in Cologne, Germany with offices in Brussels. It has been providing safety regulation for member states in Europe since 2002. It is an agent of the European Union, its mission is to ensure the highest common level of safety protection in aviation for EU citizens. EC Regulation 549/2004 'The Article-4 of Framework Regulation mandates that each European Communities State establish a National Supervisory Authority (NSA) with responsibilities for the supervision and safety oversight of Air Navigation Service Providers which provide air traffic control, airspace management and air traffic flow management services (Pellegrini & Rodriguez, 2013). The Irish Department of Transport (DoT) have designated the Safety Regulation Division (SRD) of the Irish Aviation Authority as the NSA for Ireland with Aeronautical Services Department (ASD) specifically charged with the oversight of all ANSP's nationally. External oversight of the IAA is carried out by two independent bodies, namely; the International Civil Aviation Organisation (ICAO) a body of the United Nations (UN), who conduct safety oversight audits of all States' safety regulation authorities worldwide and EASA, who routinely audit the IAA regulatory Body. As internal ANSP safety processes the safety case report was submitted to the Safety Management Unit (SMU) to ensure all evidence and arguments were met and that all identified hazards and their subsequent effects were assessed, documented and safety requirements implemented prior to operational

usage of the multiple remote tower concept. Any open issues were highlighted and detailed in the safety case report ahead of the trial being conducted. During the trial, safety levels were monitored by the implementation of a shadow operation whereby the actual towers of Cork and Shannon were manned by appropriately qualified and competent controllers at the stations while service delivery was being provided from Dublin RTC (European Aviation Safety Agency, 2014; SESAR Joint Undertaking, 2015).

Part of the safety case required to deployment of live Large Scale Demonstration (LSD) trials of remote tower operations for multiple usages in order to provide evidence to the regulator for approval. These involved the provision of air traffic services at two airports at the same time utilising innovative technological solutions. The first of its kind in the world, a dedicated team of operations and technology experts completed 50 trials demonstrating multiple remote tower operations in real time, specifically, Air Movement Control (AMC) and Surface Movement Control (SMC) at Shannon and Cork airports simultaneously from the remote tower centre at Dublin Airport. Trials were only permitted following the submission of a detailed and comprehensive safety argument, submitted by Terminal Services Operations (European Aviation Safety Agency, 2015a; SESAR Joint Undertaking, 2013) to the Irish National Supervisory Authority.

3 Demonstration of the Live Exercises

The project was supported by a safety case which was approved by the NSA for Ireland. Fifty live trial exercises involving up to 500 aircraft were conducted between June and September 2016. A consortium was established to ensure all aspects of relevant aviation activity was represented in the project. The project team consisted of Project Manager, an ATM Specialist, a Human Factors Expert and two appropriately rated Controllers who were present for the live trials. This large-scale demonstration involved the provision of aerodrome control service (air movements control and surface movements control) and flight information service for Cork and Shannon airports from a RTC located at Dublin Air Traffic Services Unit. The concept of operation (Con-Ops) for the conduct of the live trials was provision of ATS from the RTC in Dublin with the local towers at Cork and Shannon shadowing with an immediate intervention capability. The RTC contained two panoramic OTW displays in a 220-degree configuration and two airports could be displayed on one 14 screen display with a number of screens assigned to each airport depending on the operational scenarios to be trialled (Figs. 1 and 4). Therefore, one ATCO could provide services for both Cork and Shannon using one OTW display. An Electronic Strip System and a feed from the radar system were provided to the ATCO. The Demonstration Plan describes how the 50 live trial exercises were divided into three batches and how the trials built from extremely low traffic levels to increased traffic volumes in an iterative and progressive manner with a comprehensive review of each live trial
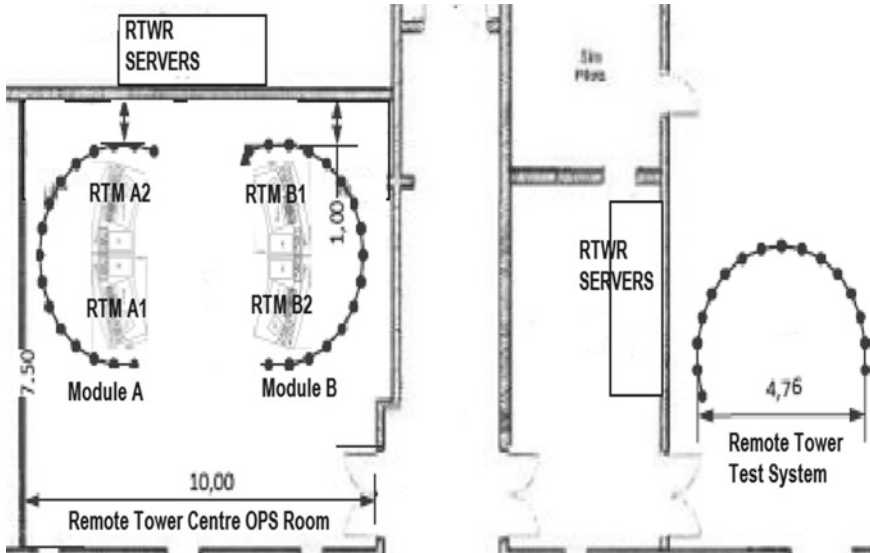
Fig. 4 Remote tower centre OPS room and test and validation room

exercise before proceeding to the next. The Demonstration Plan also specified the success criteria for each live trial exercise.

3.1 The Development of Safety Case for 50 Live Exercises

The Safety Case for multiple remote tower live trials followed the SESAR standard four-part safety case approach beginning with the production of a safety plan. This framework outlines the safety case activities to be conducted for the entire Remote Tower System (people, procedures, and equipment), the specific deliverables applicable and the timescale for submission to the NSA. This was followed by the production of a functional hazard analysis which formed the basis for the setting of safety objectives and requirements for the system. A preliminary system safety assessment document was then developed leading to a final system safety assessment. Each deliverable was submitted to the NSA as it reached maturity. A hazard log was also developed which remained open for the duration of 50 live exercises so that any previously unidentified hazards could be recorded and mitigated appropriately. Following exchanges via an NSA comments response document and meetings with the project, the NSA issued its acceptance of the safety case signifying approval to proceed with the first part of exercises as outlined in the Demonstration Plan. The project provided a report to the NSA on progress after reviewed the report and an updated safety case the NSA issued acceptance to proceed with further exercises. This arrangement was in line with the Demonstration Plan which provided for

concurrent signoff of safety deliverables at key milestones thereby causing no delay to the demonstration schedule.

3.2   Preparations of Live Exercises

The RTC contained two identical Remote Tower Modules (RTM), each with two identical working positions. Each RTM has comprised of 15 screens, 14 screens for the Out of the Window view at any one time and one spare screen. All of the Remote Tower Trials were conducted in the Remote Tower Centre OPS Room (Fig. 4). Depending on which airport the Controller was responsible for the screen configuration was changed to suit the scenario for that exercise and could be dynamically changed during the exercise. Each day the project team assembled and performed the following tasks, a systems check both from a technical and operational point of view; coordination with the Local Towers to brief them on the planned activities and get initial feedback from them on any Local issues that may impact the trials; analysis of the scheduled traffic into both airports using the EUROCONTROL CHMI tool; analysis of any weather issues that may impact the trial; decide on the targeted times to take control depending on the predicted sequence of traffic and the objectives for the upcoming exercise. Controller Working Position (CWP) on a Remote Tower Module comprises with Out the Window (OTW) visualization supported by Radar Data Processing (RDP), Electronic Flight Strips (EFS), Information Data Processing (IDP), and Voice Communications System (VCS) to provide air traffic information to ATCOs (Fig. 5).

3.3   Exercise Execution

Due to the fact that the operation of the Remote Tower Centre was for a live trial period only and at all times during the trial the Local Tower was fully staffed with ATC staff ready to re-assume control, it was decided to reduce the complexity and data line costs to provide a single non redundant video data line and a connection to the A channel radios only for voice communications. Obviously stand-alone operations, fully redundant systems and data lines would be an absolute requirement. It is critical to follow the procedures of 20 min before transfer of Control. ATCOs cross check information from Local Tower strips against information on electronic strips at the RTM. This was initiated through a phone call from the Remote Tower. The cross check also permitted the co-ordination of information on aircraft stand allocations, transponder codes and any upcoming local training details by training school aircraft on Cork airport. These cross checks were followed by a detailed handover of position(s) to the Remote Tower ATCO in accordance with current handover procedures including current weather data, airfield lighting status and nav-aid status (10 min before transfer of Control). Deselect frequency transmission on COMPAD
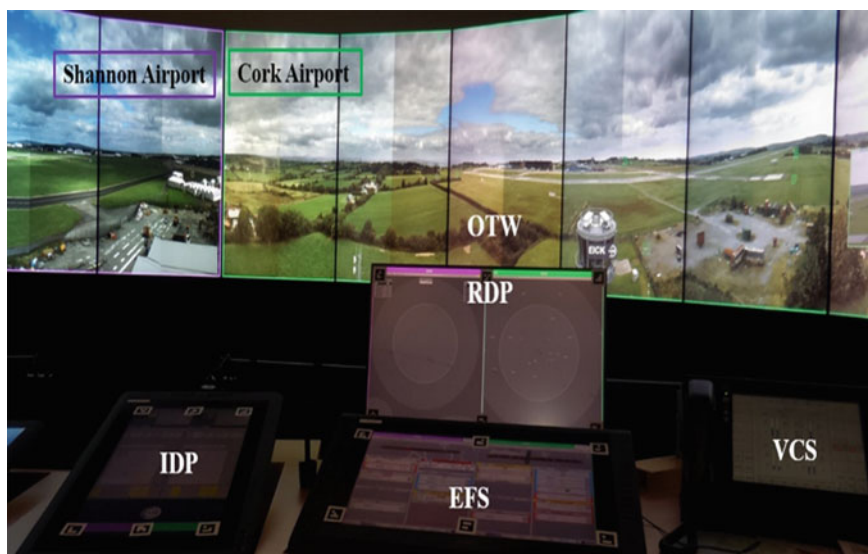
Fig. 5 Controller working position (CWP) comprises with OTW, RDP, IDP, EFS and VCS on a remote tower module

in the local tower and temporarily operate on Radio Backup System in the Local Tower. This was to avoid simultaneous transmissions from two locations on a single transmitter which may cause transmitter failure. Local Tower Controller shall advise Approach of position(s) transferred to Remote Tower.

4 Findings of Live Exercises

The operational significance of the demonstration results was assured because the live exercises were conducted during summer busy schedule of aircraft operations and vehicular activities at Cork and Shannon airports without any restrictions. Therefore, the live trial exercises were totally representative of real-time ATS provision during low to medium traffic density and complexity and were more than adequate to provide valid measurements against research objectives. In addition, documented exercise results recording aircraft movements and the timing of Controller actions enabled the project team to verify incrementally how remotely provided air traffic service for multiple aerodromes could be conducted and identify limiting circumstances. After each batch of exercises, the objectives and success criteria were recorded and examined, the status of live exercise was updated before moving on to the next batch of exercises.

Table 2 The description of batch-1 live exercises 1–5

Exercise ID	Exercise description
001	Shannon (SNN) SMC only in module RTM-A1
002	Cork (CRK) SMC only in module RTM-A2
003	Control of SNN SMC in RTM-A1 & Cork SMC in RTM-A2
004	CRK SMC first then SNN SMC combined on a single position
005	CRK SMC in RTWR A2 SNN SMC in RTWR 1 with different screen configuration to exercise 003

4.1 Findings on Batch-1 Demonstration

The demonstration plan also described the success criteria for each live exercise and how the 50 live exercises were further divided into three batches. The first batch of Number 1 to 5 exercises had the objective to familiarise operational and technical ATS and airport personnel with the procedures to be used and the environment in which they will be operating for SMC (Table 2).

The batch-1 live trials focus on the Surface Movement Control which is the air traffic control service provided to aircraft, vehicles and personnel on the manoeuvring area of Cork and Shannon aerodrome excluding the runway in use. In certain cases, the SMC controller may provide an advisory service to aircraft on the aerodrome apron. The first batch of 1 to 5 exercises is the demonstration objectives to familiarise operational and technical ATS with the procedures to be used for SMC. The critical findings of live exercises are summarized as Table 3.

4.2 Findings on Batch-2 Demonstration

The second batch comprising Number 6–20 exercises had the objective of demonstrating the applicability of integrated SMC and AMC with incrementally increased traffic movements mixing arrivals and departures at both Cork and Shannon airports. Flexibility in the timing of exercises was applied to maximise the variability of scenarios to be used with regard to runway in use, type of approach (instrument or visual). During this phase the simultaneous scenario (Cork and Shannon) was introduced with low traffic movements (Table 4).

The second batch is to demonstrate the applicability of integrated both SMC and AMC with incrementally increasing movements mixing arrivals and departures at both Cork and Shannon airports. The suitability of the equipment has been assessed with a number of comments. The suitability of the procedures has been assessed during exercise 6 to 20 and there was only one slight procedure change recommended after exercise 9. During this phase the simultaneous scenario (Cork and Shannon) was introduced with low traffic movements. The critical findings of batch-2 live exercises are summarized as Table 5.

Table 3 The critical findings of live exercise 1–5 reflect to objectives and criteria

ID	Objective	Criteria	Results of exercise
001	To demonstrate the state of readiness of the remote tower concept for industrialisation and subsequent deployment	The equipment, procedures and people elements of the remote tower system have been measured and analysed for deployment	The procedures defined for Shannon and Cork SMC control were sufficient The system was suitable for Shannon and Cork SMC operation by Dublin ATS Unit
002	To evaluate the human performance related to Human–Computer Interaction (HCI) in a sequenced or simultaneous scenario	Human performance and human factors have been measured and assessed for 'in sequence' and 'simultaneous' scenarios	There were two minor HF issues (1).The other minor issue was the Controller, on one occasion, made an incorrect selection of a button on the COMPAD; (2) The Mouse pointer in the Out of the Window view is a shared mouse pointer which, from time to time, resulted in one Controller waiting for the other to manoeuvre the Zoom Camera. However, with practice during the trials the Controllers became adept at co-ordinating the use of the mouse
003	To optimise collaborative airports operations management at Cork and Shannon in terms of scheduling, push-back, taxi out etc	The interface between the RTC and the two airport operators have been measured and assessed to identify opportunities for efficiencies in operations	This exercise only took Shannon and Cork SMC, it was better to have all airport interaction routed via the local tower to ensure that there was no confusion created with the airport authority units. Therefore, this objective is not applicable due to the fact that multiple airports were not performed
004	To support the proof of concept and demonstrate the state of readiness of the remote tower initiative for industrialisation in the case of ATS provision for multiple airports	An assessment of the live trial demonstrations to support the proof of concept and readiness for industrialisation of remote towers for multiple airports has been conducted and assessed as positive	Cork and Shannon SMC only therefore no assessment of multiple airports was possible in this batch of exercises. Therefore, this objective is not applicable due to the fact that multiple airports were not performed Furthermore, these exercises the main focus was that the RTC took control of the SMC at both airports. Each airport SMC control was performed by a dedicated Controller in the RTC. There was no impact to service observed when SMC was performed from the RTC compared to SMC being operated from the Local Tower

(continued)

Table 3 (continued)

ID	Objective	Criteria	Results of exercise
005	To confirm operational procedures for the provision of SMC for aircraft have been assessed	The operational procedures for the provision of SMC for aircraft have been assessed as suitable for the provision of SMC	The operational procedures for the provision of SMC for aircraft have been assessed as suitable for the provision of SMC ATS
006	To identify shortcomings and limitations in order to identify corrective actions required before next batch of exercises	Any shortcomings and limitations have been identified and assessed and detailed for further examination	In local Tower the AMC and SMC can easily monitor each other's activities whereas in this exercise it was more difficult. This presented a change in the working relationship between the SMC and AMC in terms of more intercom work was required. It is unlikely that in a future operation the AMC and SMC would be in a different location. However more evaluation would be required on this impact if this was to become the normal situation

Table 4 The description of batch-1 live exercises 6–20

Exercise ID	Exercise description
006	Control of SNN SMC in RTM-A1 & Cork SMC in RTM-A2
007	Control of SNN SMC & SNN AMC from a single position in RTM-A1. No Cork Positions
008	Control of SNN AMC & SNN SMC in RTM-A1 and Cork SMC in RTM-A2
009	Continuation of exercise 08. Hand back CRK and split SNN SMC onto RTM-A2
010	Control of SMC & AMC from a single position RTM-A1
011	Control of SNN AMC in RTM-A1. Cork AMC in RTM-A2
012	Merge SNN AMC and Cork AMC in RTM-A2 This exercise is a continuation of exercise 11 whereby we kept control of both SNN & CRK AMC Roles but merged them onto a single position thereby making this exercise the first time Multiple AMC Control was performed from a single Remote Tower position
013	Control of SNN AMC in RTM-A1 and CRK AMC in RTM-A2 The plan is to merge the two positions as soon as traffic allows
014	Control of SNN AMC in RTM-A1. Cork AMC in RTM-A2
015	Control of SNN AMC in RTM-A1. CRK AMC in RTM-A2 Later SNN and CRK AMC combined in RTM-A2
016	Control of SNN AMC RTM-A1 CRK AMC RTM-A2 Later SNN & CRK AMC combined in RTM-A1
017	Control of SNN AMC in RTM-A1. CRK AMC in RTM-A2 Later SNN & CRK AMC combined in RTM-A2
018	Control of SNN AMC in RTM-A1. CRK AMC in RTM-A2
019	Control of SNN & CRK AMC combined in RTM-A2
020	Control of SHA AMC in RTM-A1. No Control of CRK AMC due to Low visibility in Cork which needed to be aware before actively Controlling in these conditions

4.3 Findings on Batch-3 Demonstration

A further 30 exercises (Number 21–50 live trials) were conducted with the objective of building on the experience gained from previous exercises and increased traffic movements as appropriate in the sequenced and simultaneous scenarios (Table 6).

The third batch was conducted building on the experience gained from previous exercises and increased traffic as appropriate in the sequenced and simultaneous scenarios. The suitability of the equipment and procedures of MRTO were assessed, and there have been a number of comments in relation to the system which are captured in the IAA Remote Tower System Operational Evaluation document. The procedures of MRTO have been assessed with no additional procedure changes required. However potential changes were discussed to operating methods in any future RTC environment such as better cooperation between airports involved in a Multiple Tower Operation whereby vehicle activity at each airport is coordinated so as to manage the workload of the MRTO controllers. The critical findings of batch-3 consisting 30 live exercises are summarized as Table 7.

Table 5 The critical findings of live exercise 6–20 reflect to objectives and criteria

ID	Objective	Criteria	Results of exercise
007	To demonstrate that the full range of ATS as provided from on-site control towers can be provided without degradation from the RTC	The tasks and duties of the ATCO providing services from the RTC have been measured and assessed in line with the developed procedures to ensure that there was no degradation of service when providing a service from the RTC	As observed in the exercises 6 to 20 there are obvious differences between the Local Tower Operation and the RTC Operation. This mainly relates to the fact that the view from the Local Tower is better than the RTC. Some examples of this are: In exercise 10, the EFS is a fantastic tool with measurable safety benefits that the current paper strips don't provide In exercise 15 there is a discussion about the difficulty in seeing smaller aircraft In exercise 16 there is a discussion about rapid climbing aircraft In exercise 17 there is a discussion about challenging lighting conditions All of the above can be mitigated by a combination of experience of RTC Operation, changes to Operating methods/procedures however in some cases there may have to be a change to how the service operates
008	To evaluate various aircraft movement rates in order to establish the optimum number consistent with a safe and effective service for a remote tower—multiple airports scenario	The level of safe and effective provision of services to multiple airports under varying traffic levels has been measured and assessed to establish the optimum number of aircraft movements	From exercise 12 onwards we begin to get a picture of the workload increase required to manage a MULTI RTC environment. The project team have a good system in place for monitoring workload and a decision process in place to decide or not to continue in MULTI Operation or Split into separate airport control However, at this stage it is very difficult to actually put an exact figure or establish the "optimum" number of movements because it very much depends on the type of movement, e.g. a solo student VFR pilot in the circuit will be totally unpredictable in the workload increase to the Controller

(continued)

Table 5 (continued)

ID	Objective	Criteria	Results of exercise
009	To evaluate the human performance/factors element from the ATCO's and other human operators perspective in a sequenced or simultaneous scenario	Human performance and human factors have been measured and assessed for 'in sequence' and 'simultaneous' scenarios related to attention distributions, situation awareness and perceived workload during performing multiple remote tower operations	Based on the eye tracker data analysis, the Controllers are conscious of the fact that due to Multiple tasks having to be done at the same time that the level of service is not the same as in the Local Tower and this adds to the pressure on the RTC ATCO. The main reason for this is that the Controller knows that if he was just performing a task for a single airport that e.g. this Vehicle would not have been delayed but because he was engaged in another task for the other airport he is delaying something in the other airport. This is alien to the Controller because they would be used to very rarely having to delay replying to a Vehicle when Operating in the Local Tower
010	To assess the results of the demonstration exercises with respect to sequencing and metering to support 'in sequence' and/or 'simultaneous' operations	The application of sequencing and metering processes as applied to two airports was measured and assessed	This aspect is one of the aspects of the exercises so far that was most interesting. The project team are formulating good opinions on the impact of in sequence and simultaneous. In these series of exercises, a number of exercises with good examples of "in sequence" such as Ex 14 & 16 and also some good examples of aircraft departing in one airport and an arrival <1 Min later in the other airport such as Ex 15 & 17. In addition exercise 18 which demonstrated where it was difficult to measure due to too many predicted simultaneous tasks

Table 5 (continued)

ID	Objective	Criteria	Results of exercise
011	To assess the findings of the demonstration exercises in order to optimise collaborative airports operations management at Cork and Shannon aerodromes in terms of scheduling, push-back, taxi out etc	The interface between the RTC (one ATS unit) and the two airport operators have been measured and assessed to identify opportunities for efficiencies in operations	For simplicity and to ensure there was no confusion for the airport agencies and the Shannon ACC, all telephone interaction was conducted via the Local Tower Controllers. However, even though this aspect was not performed, it is anticipated that there would be almost no difference in the interaction between an RTC and airport operators and the Local Tower and airport operators because it is all interactions are performed by telephone and consequently no difference would be experienced. Therefore, this objective is not applicable
012	To assess the service efficiency aspect of ATS provision from the RTC for in sequence operation	The service efficiency aspect of ATS provision from the RTC for in sequence ATS provision has been assessed	The service efficiency really depends on what else the Controller is doing at the time. As stated above if there are only two aircraft in sequence and nothing else, it would be very efficiently done by one Controller, however if that single Controller is performing other tasks because he is one his own then there could be a degradation of service resulting is less efficiency for one of the two aircraft

Table 6 The description of batch-3 live exercises 21–50

Exercise ID	Exercise description
021	Control of SNN AMC in RTM-A1. CRK AMC in RTM-A2 initially then later in the exercise SNN & CRK AMC combined in RTM-A2
022	Control of SNN AMC in RTM-A1 CRK AMC in RTM-A2 initially then later in the exercise SNN & CRK AMC combined in RTM-A2
023	Control of SNN & CRK AMC combined in RTMA1
024	Control of SNN AMC in RTM-A1 CRK AMC in RTM-A2 Later SNN & CRK AMC combined in RTM-A2
025	Continuation of exercise 024. Control of SNN & CRK AMC combined in RTM-A2
026	Control of SNN AMC in RTM-A1 CRK AMC in RTM-A2 initially then later in the exercise SNN & CRK AMC combined in RTM-A2
027	Control of SNN AMC in RTM-A1 CRK AMC in RTM-A2 initially then later in the exercise SNN & CRK AMC combined in RTM-A2
028	Control of SNN & CRK AMC combined in RTM-A2
029	Control of SNN & CRK AMC combined in RTM-A2
030	Control of SNN & CRK AMC combined in RTM-A2
031	Control of SNN & CRK AMC combined in RTM-A2
032	Control of SNN & CRK AMC combined in RTM-A1 Control of SNN & CRK SMC combined in RTM-A2 initially then later in the exercise Control of SNN & CRK AMC & SMC combined in RTM-A2
033	Control of SNN AMC in RTM-A1. CRK AMC in RTM-A2 initially then later in the exercise SNN and CRK AMC combined in RTM-A2
034	Control of SNN & CRK AMC combined in RTM-A2
035	Control of SNN AMC in RTM-A1 CRK AMC in RTM-A2 initially then later in the exercise SNN & CRK AMC combined in RTM-A2
036	Control of SNN AMC& SMC in RTM-B2 Control of CRK AMC& SMC in RTM-A2
037	Control of SNN & CRK AMC & SMC combined in RTM-A2
038	Control of SNN AMC in RTM-A1, then Control of SNN SMC in RTM-A2 Next series of exercises is to follow the progression of workload starting with 2 Controllers at one airport and in the next exercise moving to 1 Controller per airport
039	Control of SNN AMC& SMC in RTM-B2 Control of CRK AMC& SMC in RTM-A2
040	Control of SNN & CRK AMC & SMC combined in RTM-A2
041	Control of SNN & CRK AMC & SMC combined in RTM-A2
042	Control of CRK SMC in RTM-A1 & CRK AMC in RTM-A2
043	Control of SNN SMC in RTM-A1 SNN AMC in RTM-A2
044	Control of SNN and CRK AMC combined in RTM-A2
045	Control of SNN and CRK AMC combined in RTM-A2

Table 6 (continued)

Exercise ID	Exercise description
046	Continuation of Ex.45 Control of SNN and CRK AMC combined in RTM-A2
047	Control of SNN and CRK AMC combined in RTM-A2
048	Control of SNN and CRK AMC combined in RTM-A2
049	Control of SNN AMC on RTA 1 and CRK AMC combined in RTM-A2 initially then later in the exercise Control of SNN & CRK AMC & SMC combined in RTM-A2
050	Control of SNN & CRK AMC & SMC combined in RTM-A2

5 Discussion of Impacts

The management of incoming and outgoing traffic at airports is a major function of ATCO who follows procedures and guidance established by past practice, industry guidelines and regulatory policies. The operational procedures seek to ensure the safety while enabling efficiency operations (MacLean et al., 2016). The demonstration of 50 live trial exercises represented real-time ATS provision during low to medium traffic density at both Cork and Shannon airports. The design philosophy of multiple remote tower operations is to keep the impact of remote tower centre control onto the ATM system to a minimum with no degradation in safety levels, no negative impact on capacity and human performance (European Aviation Safety Agency, 2015a; SESAR Joint Undertaking, 2013).

5.1 New Technology Induced Unexpected Behaviours Related to HCI

It was observed that depth perception was a potential issue (exercises 15, 16 & 17) in the RTC, as it was easier to judge the position of an aircraft in relation to another aircraft from the local tower than the RTC (Howard, 2012). Where there were more than two aircraft in the vicinity of an airport, controller's awareness of aircraft size was advantageous in determining which aircraft was closer e.g. a medium size aircraft 5 DME from an airport may be represented on the OTW view as being the same size as a small aircraft closer to the airport. The height and location of the RTC cameras, compared to the location and height of the Local Tower, particularly in Shannon where the cameras are lower than the Local Tower, made difficult to clearly differentiate between traffic on Taxiway C and D2. When a single controller is responsible for four tower roles AMC/SMC in two airports, there is a requirement to actively use four frequencies in addition to monitoring two separate approach unit frequencies (for situational awareness). Consequently, there is an increased likelihood of the controller missing a transmission by an aircraft or vehicle (Bailey et al., 2001). The organisation of whole communications systems in an RTC needs to be

Table 7 The critical findings of live exercise 21–50 reflect to objectives and criteria

ID	Objective	Criteria	Results of exercise
013	To assess cost efficiency of the provision of ATS for multiple airports from the RTC	The cost efficiency of the provision of services from the RTC remotely controlling Cork and Shannon local towers has been measured and assessed	It is the opinion of the Project team that RTC Operations do not reduce workload on a Controller, if anything there is actually a marginal increase in workload. This means that if during certain hours in the current Local Towers the two Controllers are busy then this will directly translate to two busy Controllers in the RTC. However, there were also relatively quiet times during the trials when a single RTC Controller was performing the tasks of two Controllers in the Local Tower. This resulted in a moderately busy RTC Controller compared to a Local Controller with a relatively light workload. This would imply that there are cost efficiencies to be gained. Significant work would have to be done to analyse typical traffic patterns before a shift schedule could be designed
014	To evaluate the human performance/factors element from the ATCO's and other human actors perspective in a sequenced or simultaneous scenario	Human performance and human factors have been measured and assessed for 'in sequence' and 'simultaneous' scenarios	MRTO is the future for safety and capacity of air traffic control of small airport. However, there is a trend of increasing mental workload across all the trials based on NASA-TLX. There is a raising need for dealing with ATCO's perceived workload. There were 3 scenarios which were broken down by HTA, all scenarios were proved applicable for multiple remote tower operations. The step by step of operations of multiple remote tower research had included the human–computer interaction on EFS, OTW, RDP, and IDP
015	To support the proof of concept and demonstrate the state of readiness of the remote tower initiative for industrialisation in the case of ATS provision for multiple airports	An assessment of the live trial demonstrations to support the proof of concept and readiness for industrialisation of remote towers for multiple airports has been conducted and assessed as positive	The document "LSD 02 04 IAA Remote Tower System Operational Evaluation" (Annex 3 to this report) provides the project teams full assessment of the state of readiness of the systems provided for the provision of ATS provision for multiple airports. The document lists a number of suggestions for changes to systems which should be considered in advance of any potential future deployment

Table 7 (continued)

ID	Objective	Criteria	Results of exercise
016	To demonstrate that that the current range of ATS can be provided from the RTC for simultaneous operation without degradation	The provision the current range of ATC service, flight information service and alerting service from the RTC for simultaneous operation has been assessed to confirm no degradation of service	The provision the current range of ATC service, flight information service and alerting service from the RTC for simultaneous operation has been assessed and as stated above on a number of occasions, simultaneous operations of aircraft is possible in a multiple airport RTC environment but on occasion there may be a delay to a vehicle that would not occur to the extent in the local tower operations
017	To evaluate various aircraft movement rates in order to establish the optimum number consistent with a safe and effective service for a remote tower – multiple airports scenario	The level of safe and effective provision of services to multiple airports under varying traffic levels has been measured and assessed to establish the optimum number of aircraft movements	A picture of the workload increasing required to manage a multiple aerodrome environment. In addition to what we learned in Batch 2 in relation to workload management we also noted in later exercises in Batch 3 the impact of environmental factors such as high winds and night operations The workload issue is also very much dependant on the setup of the RTC in terms of what Roles the single Controller is assigned, i.e. dual AMC or dual AMC/SMC Roles. The project team need more time to discuss the merits of all of the combinations of Roles to determine the optimum set up
018	To assess the exercises with respect to sequencing and metering to support 'in sequence' and 'simultaneous' operations	The application of sequencing and metering processes as applied to two airports was measured and assessed	As outlined in Batch 3 the project team have gained a very good understanding of what is possible in a Multiple airport 'in sequence' and 'simultaneous' aircraft operations. The new technology of PTZ, OTW and EFS did facilitate ATCOs task performance on multiple remote tower operations (exercises 46, 47, 48)

explored further for the new Human–Computer Interaction issues reflect to ATCO's visual response, auditory response, spoken response, manual response and cognitive processing information.

The view of OTW displays objects at a smaller size compared to the object size when viewed from the Local Tower, this results in it being difficult to see smaller objects far away from the camera. For areas of the airfield such as runway incursion hotspots further than 1.5 km from the cameras continuous use of the Pan-Tilt-Zoom (PTZ) is required to get a clear view of the area. When two controllers were working in the RTC, as AMC or SMC controllers, at times both controllers required the use of the PTZ, due to current system design simultaneous interaction with another different PTZ was not possible and created a situation where one controller was not able to use PTZ. This PTZ operation is a new Human–Computer Interaction (HCI) issue and increases workload on ATCO, an induced workload by MRTO which does not exist in Local Towers to the same extent, i.e. PTZ is used more frequently than binoculars (Marchitto et al., 2016). In order to try to mitigate and reduce this workload in future MRTO, the IAA had discussed system revisions on PTZ manipulation with the supplier including: (1) Automatic PTZ tracking of certain Objects as determined by the controller; (2) Explore Human–Machine Interface (HMI) adjustments to the PTZ manipulation; (3) Hotspot Cameras set up on targeted distant areas of the airfield displayed permanently on separate displays.

5.2   Impact on Safety

At airports, tower controllers are responsible for the safety and efficiency of the air movements and ground movements. Therefore, the monitoring of traffic within the control zone by ATCO's visual attention resources is an important safety mechanism (Papenfuss & Friedrich, 2016). To conduct the live exercises of multiple remote tower operations, the safety case has to be approved by regulators. The IAA has listed a fundamental principle which is that at least the same level of safe ATS provision during remote tower operations must exist as during local tower operations. The safety case provided the evidence, arguments and assumptions to support this principle. During the trials the controllers and the RTC project team were governed by the same safety management policies, principles and procedures that exist in the Local Tower operations. There were no safety occurrences during the 50 live exercises where there was a reduction in safety barriers which was not anticipated or provided for during the Safety Case development and update. The visual presentation of remote tower system in relation to the HCI functions shall not exceed the 1,000 ms of end-to-end delay in order to fit the requirements of safety assessment (European Aviation Safety Agency, 2015a). The results of eye tracking data analysis demonstrated the average fixation durations on OTW (201 ms), EFS (310 ms) and RDP (219 ms) for the scenario of aircraft in sequence departing and landing to Cork. The values of eye movement parameters were identified for the low density aerodromes which correspond to the simplest scenarios for the application of the remote tower concept.

Fig. 6 ATCO's fixation and pupil dilation on the EFS and OTW with PTZ recorded by Eye Tracker to investigate HCI design and Human Performance on Multiple remote tower operations

There is a need to conduct further investigation of ATCO's visual attention related to situation awareness and HCI with in the Remote Tower Module (Fig. 6).

The project team concluded therefore that there was no adverse impact on safety while conducting the Remote Tower Trials from the RTC and conditions for the grant of project acceptance by the NSA were successfully maintained. The live trial exercises demonstrated that the ATS provided by the RTC for a single airport and two medium airports by a single controller with 'in sequence' and 'simultaneous' aircraft operation was at least as safe as the ATS provided by the Local Towers at both Cork and Shannon aerodromes. No safety occurrence was reported nor did any operational safety issue arise during the conduct of the fifty live trial exercises. Based on the exercises of Demonstration Plan, the objective of there is no degradation in safety levels was achieved.

5.3 Impact on Capacity

In advance of commencing the Remote Tower trials, it was agreed that there would be little or no change to, or deviation from the air traffic services that the aircraft operators would normally experience when these services were provided from the Local Towers. In addition, when the RTC had control of the Shannon and Cork AMC positions, predicted traffic was monitored to determine if the two AMC positions could

be merged. On occasion when the two AMC positions were merged and controlled by a single controller it was necessary to ensure that ATCO didn't suffer from high workload and consequently impact safety and capacity. While there were a number of comments during the debriefing of these exercises in relation to workload impacts (exercises 12–20). The control of a single Local Tower with both SMC and AMC positions from an RTC was applicable for the levels of traffic in the exercise scenarios. One of the live exercise 'in sequence' presentation of traffic posed no problems for the controller' workload when controlling both Cork and Shannon AMC position. In advance of the positions being combined there were seven scheduled arrivals to the two airports in addition to a number of VFR aircraft. This situation initially has increased ATCO's workload for combining and performing the two AMC roles. Although the benefit of remote tower provision of ATC services for multiple remote towers was predicted increasing 60% of efficiency at some locations (Ziegler, 2017), it might have trade-off effects by increasing ATCO's perceived workload. In one exercise it was demonstrated that two 'in sequence' arrival flights into the two airports were manageable but it was recorded that there was potential for delay at one airport due to activities at the other airport, particularly if that activities are unexpected or non-routine. In that particular exercise two vehicles were delayed. The controller managed his workload in the exercise and was able to prioritise which work had to be done and which work could wait (Kearney et al., 2016).

Operating innovated technology of RTM to maintain safe separation of aircraft both in the air and on the ground is not only an issue of technical skill performance but also of real-time decision-making involving situation awareness and risk management within a time-limited environment (Li, 2011). There may be time lag based on the MRTO compared to Local Tower operations, due to low cloud and moisture impacting the cameras and not impacting the Local Tower. For future application, workload capacity must be monitored to ensure that unplanned pop up aircraft such as Search and Rescue Helicopters can be accommodated without delay. However even with the protections outlined above there were a number of occasions where there was a delayed response (<60 s) to a vehicle and two occasions where an aircraft was slightly delayed because the controller was dealing with traffic at the other aerodrome. Based on the exercises of Demonstration Plan, there is no significant negative impact on capacity on multiple remote tower operations observed.

5.4 Impact on Human Performance

Air traffic control is a complex cognitive task, it involves perceived information, processing information and decision-making (Papenfuss & Friedrich, 2016). During the trials a number of human factor issues were encountered, most of which were anticipated (e.g. the operation of new equipment and associated HCI) and some which were not anticipated such as the level of noise in the RTC when a single controller was operating four frequencies and monitoring an additional two frequencies. Multiple Remote Tower Operations might be the good solution for increasing

safety and capacity of ATS at small/medium airports. However, there was a trend of increasing ATCO's performance with by-products of also increasing workload on multiple remote tower operations compared to local tower operations. There is a requirement to address the issue of controllers' perceived workload for performing MRTO either by training, staffing, designing new standard operating procedures or interface design (European Aviation Safety Agency, 2015c; SESAR Joint Undertaking, 2013, 2015), as workload can negatively affect a controller's situation awareness and increase the potential for error. However, suitable human-cantered design RTM including OTW, EFS & PTZ systems can significantly improve controllers' situation awareness and reduce their cognitive workload (Laois & Giannacourou, 1995; Tobaruela et al., 2014), and increasing capability to process the information (Wickens et al., 1996).

A framework of situation awareness based on information-processing model which proposed situation awareness as 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the future' related to task performance (Endsley, 1995). ATCO's visual behaviours provide an opportunity to investigate the relationship between eye movement patterns and task performance. Eye scan pattern is one of the most powerful methods for assessing human beings' cognitive processes (Ahlstrom & Friedman-Berg, 2006). The air traffic flows of Cork and Shannon were combined in a data set describing traffic situation, consisting of time, type of event and time distance. Only data set with time distances less than 60 s were regarded as valid 'simultaneously' condition. Whilst the landing is represented in the flight movement data by a single timestamp, the actual process of landing occupies the ATCO's attention longer by monitoring closely the last mile of the final approach until the aircraft touches down and brakes on the runway. Therefore, it is assumed that two landings within the 60 s time span can be considered as simultaneous regarding the monitoring task of the controller (Schier et al., 2011). Based on the exercises of Demonstration Plan, there is no negative impact on human performance on multiple remote tower operations occurred.

6 Conclusion and Recommendation

The concept of remote tower operations has been addressed as a suitable solution and is being developed in many countries. This research provided scientific evidence that multiple remote tower operations can achieve the objectives of Single European Sky ATM Research program. The implementation of multiple remote tower operations is promising as new technologies can assist ATCO's situation awareness. The findings can be reflected to both ATCO's training and system design. In terms of standardisation, the results of 50 live demonstrations can provide feedback to develop standards for both single and multiple modes of remote towers systems. The Demonstration provided confirmation that a multiple remote tower solution provided the potential for more cost-efficiency deployment of human resources during periods

of low aircraft movements, particularly when combined with other initiatives such as the centralisation of Approach Control Service and for contingency purposes. There is no negative impact on human performance observed on the demonstration of multiple remote tower operations. However, there are some issues to be aware of for future implementation, including the impacts of monitoring different radio frequencies and perceived workload. Workload management in the provision of ATS for multiple towers is a new challenge for controllers, and practice is required to even out the workload by distributing tasks more evenly where possible. In Summary, the multiple remote tower operations show potential in air traffic services as an alternative to traditional Local Towers. The novelty and flexibility of the advanced technology allow regulators to be creative in adapting to fit safety regulations, also has the potential to fundamentally change the way operators provide ATS. The evolution of implementation of the innovative technology requires a careful balance between cost-efficiency and its potential impact on safety, capacity, and human performance. From a regulatory perspective the results of these live trials may contribute to EASA rulemaking activity for single and multiple airports remote tower operations.

References

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36*(7), 623–636. https://doi.org/10.1016/j.ergon.2006.04.002

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2001). The effects of interruptions on task performance, annoyance, and anxiety in the user interface. *Interact, 1*, 593–601.

Dorighi, N., & Rabin, B. (2002). *NASA uses virtual reality to target runway incursions at LAX*. In *2002 Federal Aviation Administration Technology Transfer Conference*, Atlantic City, New Jersey.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65–84. https://doi.org/10.1518/001872095779049499

European Aviation Safety Agency. (2014). *Term of reference for a rulemaking task: Technical requirements for remote tower operations.*

European Aviation Safety Agency. (2015a). *Guidance material on the implementation of the remote tower concept for single mode of operation.*

European Aviation Safety Agency. (2015b). *Notice of proposed amendment 2015–04: Technical and operational requirements for remote tower operations.*

European Aviation Safety Agency. (2015c). *Requirements on air traffic controller licensing regarding remote tower operations.*

Fürstenau, N. (2016). Model-based analysis of two-alternative decision errors in a videopanorama-based remote tower work position. In N. Fürstenau (Eds.), *Virtual and remote control tower*. Research Topics in Aerospace. Springer. https://doi.org/10.1007/978-3-319-28719-5_11

Federal Aviation Administration. (2012). *Airport traffic control tower alternatives.* Fort Collins.

Friedrich, M., & Mohlenbrink, C. (2013). Which data provide the best insight? A field trial for validating a remote tower operation concept. In *The 10th USA-Europe Air Traffic Management Research and Development Seminar*, Chicago, USA.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI), 7*(2), 174–196. https://doi.org/10.1145/353485.353487

Howard, I. P. (2012). *Perceiving in depth, volume 1: Basic mechanisms* (pp. 93–128). Oxford University Press.

Irish Aviation Authority. (2016). *Total movements for Cork and Shannon 2015 and 2016*. Dublin, Ireland.

Irish Aviation Authority. (2017). *Annual report of total movements for Cork and Shannon 2016*. Dublin, Ireland.

Kearney, P., Li, W. C., & Lin, J. J. H. (2016). The impact of alerting design on air traffic controllers' response to conflict detection and resolution. *International Journal of Industrial Ergonomics, 56*, 51–58. https://doi.org/10.1016/j.ergon.2016.09.002

Kraiss, K., & Kuhlen, T. (1996). Virtual reality: Principles and applications. In N. Furstenau (Ed.), *From sensors to situation awareness* (pp. 187–208). DLR-Mitteilung, 112-96-02.

Laois, L., & Giannacourou, M. (1995). Perceived effects of advanced ATC functions on human activities: Results of a survey on controllers and experts. In *Proceedings of the Eighth International Symposium on Aviation Psychology*, Ohio.

Leitner, R., & Oehme, A. (2016). Planning remote multi-airport control—Design and evaluation of a controller-friendly assistance system. In: N. Fürstenau (Ed.), *Virtual and remote control tower*. Research Topics in Aerospace. Springer. https://doi.org/10.1007/978-3-319-28719-5_7

Li, W. C. (2011). The casual factors of aviation accidents related to decision errors in the cockpit by system approach. *Journal of Aeronautics, Astronautics and Aviation, 43*(3), 147–154.

MacLean, L., Richman, A., & MacLean, S. (2016). Benchmarking airports with specific safety performance measures. *Transportation Research Part a: Policy and Practice, 92*, 349–364. https://doi.org/10.1016/j.tra.2016.06.016

Marchitto, M., Benedetto, S., Baccino, T., & Canas, J. J. (2016). Air traffic control: Ocular metrics reflect cognitive complexity. *International Journal of Industrial Ergonomics, 54*, 120–130. https://doi.org/10.1016/j.ergon.2016.05.010

Moehlenbrink, C., & Papenfuss, A. (2011). ATC-monitoring when one controller operates two airports research for remote tower centres. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 76–80). SAGE Publications.

Nene, V. A. (2008). *A proposed operational concept for nextgen tower*. Mitre Corp.

Papenfuss, A., & Friedrich, M. (2016). Head up only—A design concept to enable multiple remote tower operations. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, Sacramento, CA (pp. 1–10). https://doi.org/10.1109/DASC.2016.7777948.

Pellegrini, P., & Rodriguez, J. (2013). Single European sky and single European railway area: A system level analysis of air and rail transportation. *Transportation Research Part a: Policy and Practice, 57*, 64–86. https://doi.org/10.1016/j.tra.2013.09.004

Saab, A. B. (2010). *Advanced remote towers: Final activity report*. Author.

Schier, S., Papenfuss, A., Lorenz, S., Walther, J., & Moehlenbrink, C. (2011). An approach to support controller work-place design in a multi-airport environment using fast and real-time simulations. *CEAS Aeronautical Journal, 2*(1), 185–193. https://doi.org/10.1007/s13272-011-0030-8

SESAR Joint Undertaking. (2013). *SESAR human performance assessment process V1 to V3-including VLDs* (Pro. No. PJ19 SESAR2020/P16.4.1/16.06.05). EUROCONTROL.

SESAR Joint Undertaking. (2015). *Remote towers demonstration report* (Pro. No. LSD 02.04). Irish Aviation Authority.

Tobaruela, G., Schuster, W., Majumdar, A., Ochieng, W. Y., Martinez, L., & Hendrickx, P. (2014). A method to estimate air traffic controller mental workload based on traffic clearances. *Journal of Air Transport Management, 39*, 59–71. https://doi.org/10.1016/j.jairtraman.2014.04.002

Van Schaik, F. J., Roessingh, J. J. M., Lindqvist, G., & Fält, K. (2016). Detection and recognition for remote tower operations. In N. Fürstenau (Ed.), *Virtual and remote control tower*. Research Topics in Aerospace. Springer. https://doi.org/10.1007/978-3-319-28719-5_3

Wickens, C. D., Miller, S., & Tham, M. P. (1996). The implications of data-link for representing pilot request information on 2D and 3D air traffic control displays. *International Journal of Industrial Ergonomics, 18*(4), 283–293. https://doi.org/10.1016/0169-8141(95)00065-8

Ziegler, J. (2017). Cut the cost: How can remote tower and virtual tower solutions improve the efficiency and reduce the operational costs of ATC services for regional airports? *The International Review of Air Traffic Technology Manage, 2017*, 1–3.

Designing a Low-Cost Remote Tower Solution

Fabian Reuschling and Jörn Jakobi

Abstract In recent years, Remote Tower Optical Systems became a well-established means of providing cost-effective remote surveillance capabilities to aerodromes operated by air traffic control (ATC) officers. However, smaller, non-ATC operated (uncontrolled) aerodromes, often affected by very low revenues, may also benefit from a Remote Tower solution, but still cannot afford the implementation and maintenance costs just like that. The aim of the work presented herein is to design a low-cost Remote Tower solution specifically tailored towards the requirements and budget constraints of uncontrolled aerodromes. The approach taken consists of a site survey at an uncontrolled aerodrome identifying two constraints faced and three potential operational use-cases, followed by the definition of a suitable camera set-up analogous to standard Remote Tower systems consisting of two panoramic cameras each capturing a 180-degree panorama and a pan-tilt-zoom-camera. Two novel visualization concepts based on a head-mounted device (HMD) are developed. The results of a validation study are presented indicating the camera set-up to be sufficient for remote control of uncontrolled aerodromes and pointing towards the potential of an HMD-based visualization. Finally, the proposed low-cost solution is compared to standard Remote Tower Optical Systems regarding the costs and design of the controller working position.

Keywords Remote tower · AFIS · Low-cost

1 Introduction

Currently, Remote Tower Optical Systems are well-established as a means to provide cost-efficient aerodrome air traffic control (ATC). By visual surveillance technology Remote Towers allow the traffic to be remotely controlled from any location decoupled from the direct Out-of-the-Window (OTW) view of the physical tower at the aerodrome. Combining the remote control of multiple aerodromes into one

F. Reuschling · J. Jakobi (✉)
Deutsches Zentrum Für Luft- Und Raumfahrt (DLR) E.V, Institute of Flight Guidance,
Lilienthalplatz 7, 38108 Braunschweig, Germany
e-mail: joern.jakobi@dlr.de

Remote Tower Center or even one Controller Working Position (CWP) (referred to as a multiple Remote Tower Module [MRTM]) furthermore allows for a flexible allocation of personnel.

Comprehensive research is conducted exploring the requirements for Remote Towers (Ellis and Liston, 2016; Jakobi and Hagl, 2020, van Schaik, 2016), validation approaches (Schaik, 2016), design and benefits of Multiple Remote Tower solutions (Papenfuss and Friedrich, 2016), and detailing various supporting systems, like fusing visual and infrared panoramas (van Schaik et al., 2010; Reuschling et al. 2020; Hagl et al., 2018). Nevertheless, all state-of-the-art Remote Tower Optical Systems, even though they are already very cost-effective systems, inherit costs to install, operate, and maintain—usually a six-digit Euros amount or more—that generally make them affordable only for aerodromes providing ATC services with higher revenue margins. The benefits of Remote Towers, however, also extend to the great number of non-ATC operated, uncontrolled aerodromes. Hence, a low-cost Remote Tower solution is needed, designed for the requirements and budgets of uncontrolled aerodromes. These provide either an aerodrome flight information service (AFIS) or UNICOM.[1] For simplicity and readability, the term "AFIS" takes into account UNICOM services in the context of this work.

In this work, the results of a site survey at the AFIS-operated airport Schönhagen are reported with respect to the current situation of German non-ATC operated aerodromes and potential use-cases. Then, a possible design comprising a camera set-up and two visualization concepts is outlined, followed by the results of a preliminary evaluation. Finally, the proposed low-cost solution—costing a lower a five-digit Euros amount—is compared to standard Remote Tower systems.

2   The Need for a Low-Cost Solution

The defining benefit of Remote Tower Optical Systems is the possibility to control multiple aerodromes from one Remote Tower Center or even a single CWP, thereby enabling a flexible allocation of operators depending on the traffic situation. This is of particular benefit for aerodromes with high traffic volume fluctuations over the day, week, or year, often the case for small regional, ATC-controlled aerodromes. Uncontrolled aerodromes may also face this and other constraints that can be resolved with a Remote Tower Optical System.

Therefore, a site survey was performed at the start of the study at the Schönhagen aerodrome in Germany. It is equipped with two runways, allows for instrument flight rules (IFR) arrivals and departures, and operated by one AFIS officer (AFISO) per working shift. It serves as an example for a multitude of similar

---

[1] UNICOM (Universal Communications) is implemented at non-ATS aerodromes where a two-way radio communication is not mandatory. Pilots state their intentions via a dedicated frequency which may be accompanied by a ground station acting as informal, non-certified facility for information and advisory exchange. (30; 31).

airports in Germany—about 406 non-ATC aerodromes of which 21 are operated by AFISOs (Wikipedia, 2021; Bundesaufsichtsamt für Flugsicherung (BAF), 2021). In the following, observed constraints for uncontrolled aerodromes and use-cases discussed with the AFISOs are outlined.

2.1 Constraints for Uncontrolled Aerodromes

In contrast to air traffic control officers (ATCOs), who are tasked with granting clearances to aircraft in their responsibility, AFISOs provide aircraft with relevant information in their area of responsibility, like weather, other aircraft in the vicinity, runway condition, etc., to support the pilots to safely act in the AFIS airspace. With a few exceptions, they also have fewer and less sophisticated tools at their disposal, e.g. no approach or surface radar, and rely more on pilots' reports and the direct OTW view.

The two defining constraints observed during the site survey are the following:

1. Limited or no OTW view.
   Instead of continuously being on the tower with a good view of their area of responsibility, AFISOs often have secondary tasks like billing or provision of other services for pilots on ground. Within these times, AFIS is provided from a secondary controller working position with limited or no OTW view, reducing the ATS service level in such situation.
2. Limited resources.
   As the overall Air Traffic Service (ATS) level at uncontrolled aerodromes is lower that at ATC-controlled ones, the charges for providing these services are as well. This consequently leads to limited resources for improving the service level in terms of opening hours or offering a better communication, navigation and surveillance service. Customer wishes or requests for IFR operations or extended opening hours cannot be met.

2.2 Operational Use-Cases

Building on the constraints outlined in the previous section, three potential use-cases for a remote control solution were developed and discussed with airport stakeholders. In this work, they are presented from an operational and technical point of view and not with respect to current regulations. The use-cases are:

1. **On-demand provision of higher quality ATS (Air Traffic Services)**
   By temporary handover of control to Remote Tower Centers (RTCs), ATCOs at the RTC can provide ATC for an AFIS airport or ATC/AFIS for an UNCIOM operated airport, which again increases service level, attractiveness, and revenues. This would enable a flexible switching between UNICOM, AFIS,

and ATC optimized for the actual demand, increasing the cost-effectiveness and service level.

2. **Extension of the daily opening times of an AFIS airport**
   Particularly at the edges of the airport opening hours, synergies can be exploited by handover of on-site AFIS services to a not on-site Remote Tower Center to extend opening hours of an airport and increase the service level, which in turn would make the airport more attractive for their customers and further in turn, would increase the airport's revenues.

3. **Ad-hoc provision of AFIS, decoupled from the OTW Tower view**
   As described in the first constraint, AFISOs may not be able to have an OTW view at every time, even when being at the airport site. A compact and portable system—opposed to multi-monitor set-ups used for Remote Tower systems— would allow the AFIS officer to get an ad-hoc visual overview of the traffic even when temporarily fulfilling secondary tasks and therefore raise the overall service level.

3 Design for Low-Cost Remote Surveillance

Summarizing the conducted research, requirements for Remote Tower Optical Systems are defined in EUROCAE's Minimum Aviation Systems Performance Standards (MASPS) for Remote Tower Optical Systems (ED-240A) (European Organisation for Civil Aviation Equipment (EUROCAE), 2018). While some minimum performance requirements (e.g. video update rates above 1 frame per second [FPS]) are explicitly stated, the specific requirements, like the actual video update rate or the distances in which aircraft shall be detectable or recognizable—defining the cameras' resolution—are derived from the operational requirements. These are specified by the individual operators for their aerodrome and use-case, therefore depending on the type of provided service, airspace classification, and importance of visual observation dependent tasks.

As outlined in Sect. 2.1, operators providing AFIS or UNICOM services are generally not allowed to give clearances. Instead, they are tasked with relaying relevant information to aircraft in the vicinity so that the pilots can make informed decisions, self-separate, and navigate the airspace safely by see and avoid principles. Due to the limited area of responsibility and lower service level, it can be assumed that the operational requirements for the provision of AFIS (or UNICOM) are lower than for ATC, in turn affecting performance requirements and design of a low-cost Remote Tower solution.

In this work, less strict requirements are therefore placed on the general camera set-up—i.e. the number and siting of cameras and the quality and kind of display system—as well as the technical details of the system, like overall sophistication of the video feed in terms of resolution, frame rate, field of view (FOV), network bandwidth, etc. This reduces the overall cost of the system and would make it affordable for uncontrolled low revenue aerodromes. In adjusting the various parameters listed,

the camera set-up described in Sect. 3.1, the visualization concepts presented in Sect. 3.2, and interaction concepts described in (Hofmann et al., 2020) are designed.
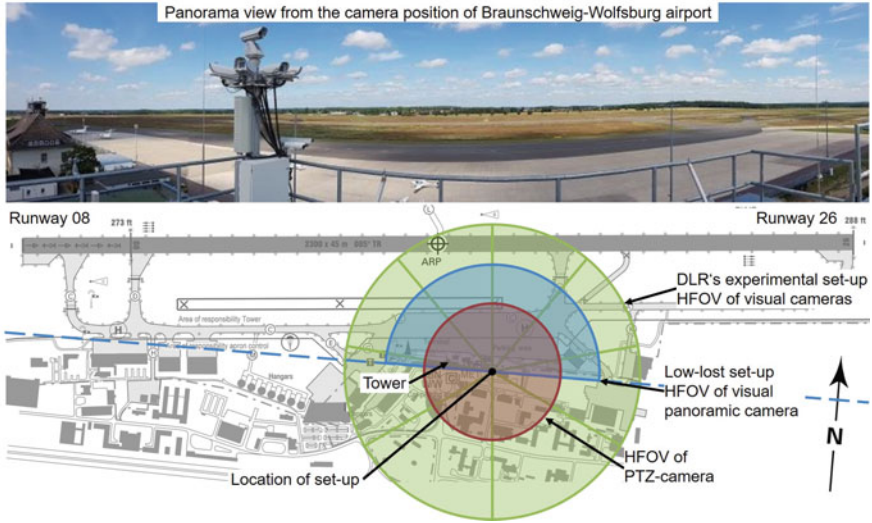
3.1 Camera Set-Up

The camera set-up is designed in analogy to standard Remote Tower camera set-ups. As an example, DLR's research system at the Braunschweig-Wolfsburg aerodrome pictured in the left part of Fig. 1 is used. It consists of two elements:

The first is ring of eight visual spectrum cameras totaling an 360° horizontal ($6 \times 40°$ and $2 \times 60°$) and 66° vertical FOV. Each camera has a resolution of 1920 × 1080 pixels at 30 FPS. The videos of the cameras are externally stitched into a seamless 360-degree visual panorama. For the low-cost solution, the ten cameras are reduced to one panoramic camera covering 180° horizontally and 90° vertically. It has a resolution of 3648 × 2052 pixels at 15 FPS and provides a fully stitched 180-degree panorama captured by four sensors. The camera is orientated in such a way that the entire arrival and departure sectors as well as the apron are covered (see Fig. 2).

The second element is a pan-tilt-zoom-camera (PTZ-camera). It is used as a replacement for binoculars and allows operators to enlarge parts of the panorama. The camera used in DLR's experimental camera set-up has a resolution of 1920 × 1080 pixels at 30 FPS. It has a horizontal (pan) range of 360° and can be vertically tilted in a range of 130° with a maximum speed of 140°/s. The FOV varies depending on the up to 30 × zoom level. The PTZ-camera is considered an integral part of the Remote Tower set-up and therefore not replaced by an inexpensive version, but reused from the experimental set-up.
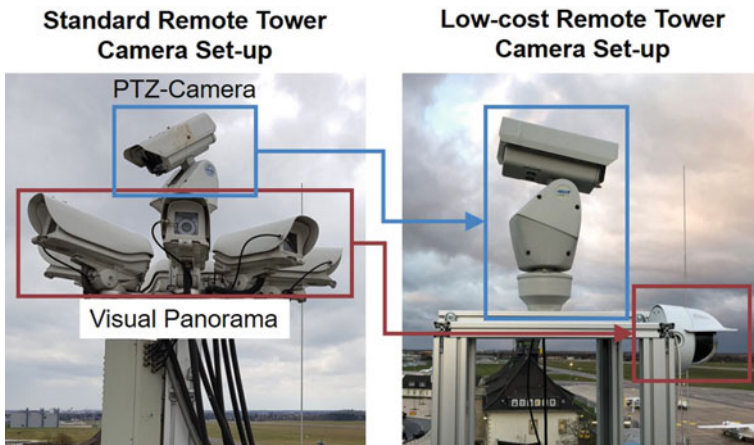
Fig. 1 Elements of DLR's experimental Remote Tower set-up in comparison with the proposed low-cost solution

Fig. 2 Aerodrome chart of Braunschweig-Wolfsburg aerodrome with marked position of cameras and HFOVs of DLR's experimental set-up and low-cost set-up, panorama view from camera position presented in top part

Additionally, some Remote Tower camera set-ups also are equipped with one or several sensors capturing an infrared panorama (Reuschling et al., 2020). The primary reason for such an addition is to improve the visibility in adverse weather conditions enabling operators to maintain visual contact. As an infrared panorama would be useful in conditions in which most flights (i.e. those following visual flight rules [VFR]) would not be allowed to depart or arrive, no such addition is considered for the low-cost system.

The so designed low-cost Remote Tower (see right part of Fig. 1) is erected at the Braunschweig-Wolfsburg aerodrome. The siting of the cameras and the horizontal field of views (HFOVs) of the cameras included in DLR's experimental set-up and the designed low-cost set-up are presented in Fig. 2. The technical parameters of the set-ups are compared in Table 1.

Table 1 Comparison of the specifications of DLR's experimental Remote Tower set-up and the proposed low-cost solution (Axis Communications AB, 2020; Pelco Inc., 2020)

		DLR's experimental set-up	Low-cost solution
Visual Panorama			
	Camera type	AXIS Q1625	AXIS P3807-PVE
	Number of cameras	8	2
	Resolution [px]	1920 × 1080	3648 × 2052
	Framerate [FPS]	30	15
	Covered area [°]	6 × 40 & 2 × 60 horizontally 66 vertically	180 horizontally 90 vertically
PTZ-Camera			
	Camera type	Pelco Esprit Enhanced	
	Resolution [px]	1920 × 1080	
	Framerate [FPS]	30	
	Field of view [°]	between 63.7 and 2.3 horizontally between 35.8 and 1.3 vertically	
	Rot.-Speed [°/s]	140	
	Optical Zoom	30x	
Infrared Panorama			
		Optional to maintain visual contact in low visibility	Not applied

3.2 Visualization Concepts

With the camera set-up of the low-cost Remote Tower solution being fixed, different concepts of how to visualize the video feeds are designed. In addition to a baseline based on standard Remote Tower display solutions, novel concepts are developed with the aim of providing a compact, simple to use—and potentially portable— system as is required for the outlined use-cases (see Sect. 2.2). In order to fulfil these aspects, the concepts are designed to use a head-mounted device in which the visualization and interaction is integrated.

As first step towards the design of suitable visualization concepts, a design-space analysis is first carried out following the QOC (Questions, Options, Criteria) method (MacLean, 1991). Its purpose is to construct a comprehensive, visual overview of all design options by asking specific questions (e.g. "How is the PTZ-camera positioned?") to which possible answers ('options') are then formulated. In Fig. 3, part of the so outlined design-space is presented.

In order to explore a possibly large area of the design-space, options for two concepts are selected representing opposite ends: For visualization concept 1 (marked in blue in Fig. 3), the video feeds from both cameras are used and the positioning of the PTZ-camera is decoupled from the operator's head turns. For the more straight-forward visualization concept 2 (marked in red in Fig. 3), only the PTZ-camera's
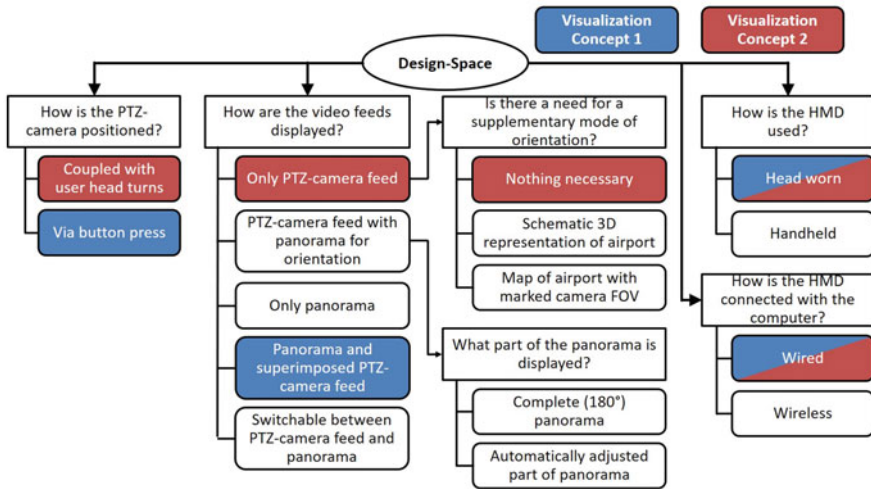
Fig. 3 Design-space for visualization concepts

video feed is used which also moves in conjunction with performed head turns. For both concepts, the HMD is designed to be head worn and wired in order to support a fast and easy development of the visualization concepts. In the following paragraphs, the properties of the baseline and the two novel visualization concepts are described.

Baseline. For comparison with the novel visualization concepts utilizing an HMD, a baseline visualization is defined. It is designed after the display solutions of standard Remote Tower Optical Systems and utilizes two standard computer monitors. The video feeds of the panoramic camera and the PTZ-camera are each displayed on one of the monitors, pictured in the top of Fig. 4. For control of the PTZ-camera, the manufacturer-provided interface and a standard computer mouse is used.

Concept 1 (PTZ & Panorama). In the first visualization concept, the video feed of the panoramic cameras is stretched out around the operator to form a seamless 180-degree panorama, as shown in the bottom left of Fig. 4. This panorama is fixed in the Virtual Environment (VE) so that the operator can look around by turning her/his head. During this, no lag is experienced as no mechanical movements are necessary. The video feed of the PTZ-camera is superimposed on the panorama and serves as virtual analogy of binoculars. It is positioned via a button press after which the camera automatically aligns with the gaze direct of the operator and then remains fixed at the respective position.

Concept 2 (PTZ only). In contrast, the second visualization is solely based on the video feed of the PTZ-camera which is displayed across the entire FOV of the HMD (see bottom right of Fig. 4). The movements of the camera are directly coupled with the operator's head turns so that the direction the PTZ-camera is facing constantly equals the gaze direction of the operator.
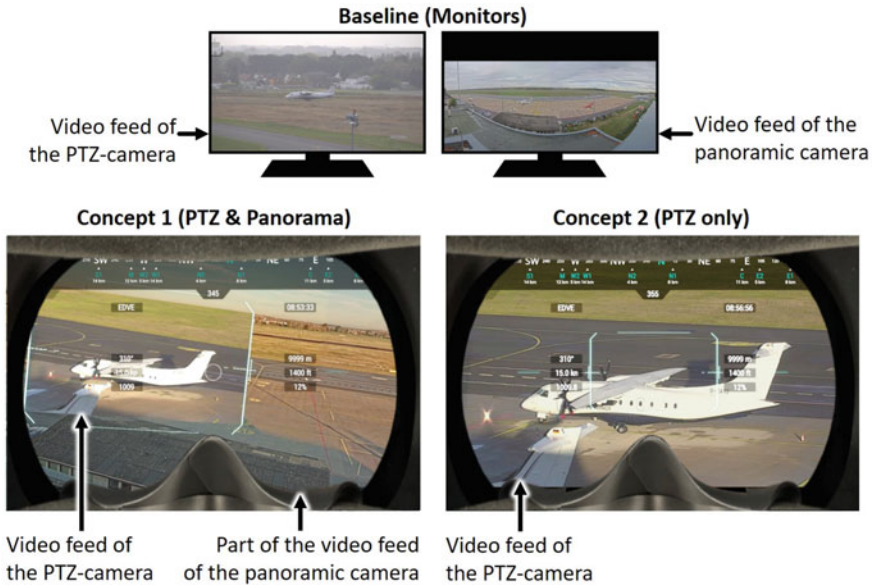
Fig. 4 Outline of the baseline and two different visualization concepts

3.3 The Problem of Cybersickness

With the use of a head-mounted device, as in visualization concept 1 and visualization concept 2, opposed to standard computer monitors, the issue of potential unintended side effects of the immersion into a virtual environment arise. The most prominent of these is 'cybersickness', a special form of motion sickness that is visually induced, i.e. by what a user sees, not by what they feel. The symptoms primarily experienced affect the oculomotor apparatus and include eyestrain and difficulty focusing. Further symptoms may be nausea, headache, dizziness and, in extreme cases, vomiting (Barrett, 2004; Rebenitsch and Owen, 2016).

Critically high levels of cybersickness clearly pose a health and safety issue. It is also expected that even non-critical cybersickness may degrade the user's performance (Stanney and Key, 1995). Hence, special consideration during the design phase and the validation study described in the next chapter is placed on avoiding factors inducing cybersickness. The measures taken include avoiding fast movements and automatic movements without user control, seated use, limiting the duration of exposure—which also is in line with the ad-hoc and on-demand use-cases described in Sect. 2.2—and prior training, all of which are suggested by literature. Especially the last measure—prior training—is assumed to reduce the experienced cybersickness as it is known that users habituate to VEs so that fewer symptoms are felt after repeated exposures (Ultraleap Limited, 2021; Barrett, 2004; McCauley and Sharkey, 1992; Regan, 1995; Yuan, 2018].

4 Validation

Within the work described herein, a validation study for the developed low-cost Remote Tower solution focusing on the visualization concepts described in Sect. 3.2 was performed. It aimed to determine in how far they solve the constraints uncontrolled aerodromes face and their suitability for the operational use-cases. In this chapter, the results of this study are discussed.

4.1 Method

The validation study was performed at the DLR Institute of Flight Guidance in Brunswick between the 28th October and 5th November 2020. In it, two AFISOs and two ATCOs working at two uncontrolled, German aerodromes and four ATCOs working at two controlled, German aerodromes, all holding valid licenses, participated. Additionally, one former ATCOs validated the system totaling nine validation participants. Of these, eight were male while one was female.

The two visualization concepts are implemented on an HTC Vive Pro Eye HMD using the software Unity (Unity Technologies, 2021). This particular HMD was chosen because its angular resolution, approximately 20 px/° horizontally and 22 px/° vertically, coincides with that of the panoramic camera (21 px/° and 23 px/°) therefore ensuring that no image information is lost because of a too low resolution. For the baseline, two 27-inch monitors with a resolution of 2560×1440 pixels are used. The participants are presented with live video feeds from the camera set-up at the Braunschweig-Wolfsburg aerodrome and the radio communication between the pilots and the tower controllers. During the entire time of the study, visibility was at least 6000 m (according to the METAR weather reports) with no significant downfall.

The validation study is set up as with-subject experiment. Each participant performed three runs in which the visualization is manipulated as independent variable between the factors 'Baseline', 'Concept 1 (PTZ & Panorama)', and 'Concept 2 (PTZ only)'. The order in which the concepts are presented is randomized between subjects to minimize any carry-over effects like learning or fatigue. Before using each concept, the participants are given ten minutes to accommodate themselves with the system.

In the validation study, five subjective metrics were collected. The first is the System Usability Scale (SUS) (Brooke, 1996), providing a general indication of the baseline's and concepts' usability. In order to evaluate the induced cybersickness, the Simulator Sickness Questionnaire (SSQ) is administered. It was originally designed for flight simulators, but is widely adopted for virtual environments as well (Kennedy et al., 1993). Furthermore, the participants were asked to complete EUROCONTROL's SHAPE Automation Trust Index (SATI) (Eurocontrol, 2012; Dehn, 2008) that is specifically designed to assess the trust air traffic controller have

in a system. Finally, two tailor-made questions on the approximate maximum time one could work with the concept, and the personal ranking of the baseline and the two concepts are asked.

4.2 Results

System Usability Scale. For a general evaluation of the suitability of the concepts, the System Usability Scale as indicator of the overall usability is evaluated. In Fig. 5, the scores for the baseline and the two visualization concepts are presented as box and whisker plots (Tukey, 1977). The rating scale developed by Bangor et al., (2009) is added in the right part of the figure.

As can be seen, the median SUS scores gradually drop from left to right. While the baseline's median is 90.0 and in the range of 'excellent' usability, that of concept 1 is lower at 75.0, just above 'good' usability, and the median of concept 2 is 65.0 equalling an 'OK' to 'good' usability according to the rating scale. The same trend observed for the median is visible for the box position—marking the range between the first and third quartile. All of these values are above a SUS score of 52, defined to equal 'OK' usability. The maximum and minimum values for the baseline and the two concepts are in similar ranges: The maximums (27.5, 25.0, and 25.0) lie between 'excellent' and 'best imaginable' usability. The minimum values are located around 'worst imaginable' usability at scores of 27.5, 22.5, and 25.0 for the baseline, concept 1, and concept 2, respectively. Since the lowest score for the baseline is more than 1.5
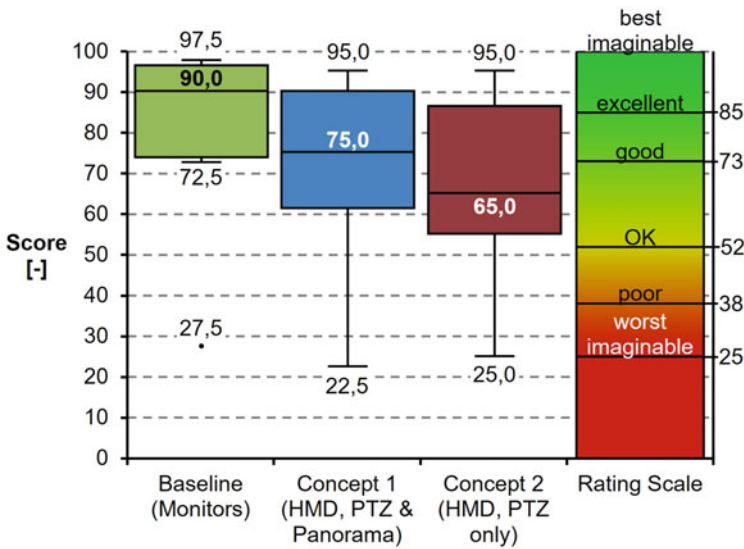
Fig. 5 SUS scores for n = 9 participants with rating scale from Bangor et al., (2009)
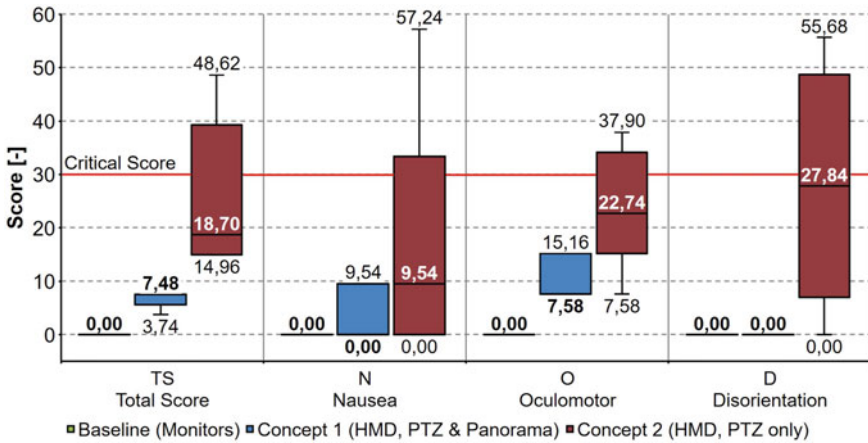
Fig. 6 SSQ scores for n = 5^2 participants with median written in bold and indication of critical score

times the interquartile distance, i.e. the height of the box (in this case 30.0), below the first quartile, it is regarded as outlier.

Preponing the discussion, this very low minimum value compared to the location of the first quartile as well as those observed for the two other concepts point towards an outlier in the data. Especially the baseline's usability being rated as 'worst imaginable' stands in contrast to the 'excellent' median score and indicate a rejection of the Remote Tower concept in general rather than the proposed low-cost solution. Similar ambiguities were also found in the other subjective metrics supporting the conclusion that the evaluation of the participant may have been influenced by a personal bias. Hence, the data for this particular participant were excluded from the results and *those subsequently presented in this section are calculated for the remaining eight participants.*

Simulator Sickness Questionnaire. The evaluation of the Simulator Sickness Questionnaire yields one Total Score (TS) and three Nausea (N), Oculomotor (O), and Disorientation (D) subscale scores. The results for the baseline and two concepts validated are depicted in Fig. 6. While the authors of the SSQ found a score of 15 to be critical, it was observed that for virtual environments, a score twice as high is still acceptable (Brooke, 1996; Stanney et al., 1997). Therefore, an SSQ score of 30 is marked as critical score in the figure.

It can be noticed that for the baseline, none of the symptoms included in the SSQ were experienced by the participants resulting in a TS and subscale scores of 0.00. For concept 1, a median Total Score of 7.48 is calculated and a median TS of 18.70 for concept 2. In the Nausea subscale, median scores of 0.00 for the 'PTZ & Panorama' concept and 9.54 for the 'PTZ only' concept are observed. In the Oculomotor subscale, the difference between the medians is larger with the one

---

[2] Due to technical problems, only the scores of five participants were recorded and evaluated.

of concept 1 being 7.58 and the median score of concept 2 being 22.74. This gap widens for the medians in the Disorientation subscale. There, the median of concept 1 is 0.00 and the median of concept 2 is 27.84. In all scales and for all concepts, the median values lie below the critical score of 30.

The spread of the SSQ scores is larger for the 'PTZ only' concept than for the 'PTZ & Panorama' concept for the TS and all three subscales. The Total Scores for concept 1 range between 3.74 and 7.48 and all lie below the minimum TS for concept 2, which is 14.96. The highest TS for the 'PTZ only' concept is 48.62 and above the critical value. In the Nausea subscale, the scores for concept 1 lie in the interval of 0.00 to 9.54. The scores for concept 2 range between 0.00 and 57.24, extending to above the critical score of 30. While the minimum values in the Oculomotor subscale is 7.58 for both concepts, the maximum scores differ at 15.16 for concept 1 and 37.90 for concept 2, which again is above the critical score. For the Disorientation subscale, only 0.00 scores were calculated for concept 1. The values for concept 2 range between 0.00 and 55.68, above the critical score. In summary, the maximum scores and upper ends of the third quartile (i.e. the top end of the box) of the 'PTZ only' concept lie above the critical score of 30 throughout all scales.

The individual symptoms reported by the participants for concept 1 (top part) and concept 2 (bottom part) are presented in Fig. 7, grouped by the sub-scales they weight on. The symptoms 'general discomfort', 'nausea', 'difficulty concentrating', 'difficulty focussing', and 'blurred vision' count towards two subscales and are marked with '*'. An individual column denotes the severity reported by one
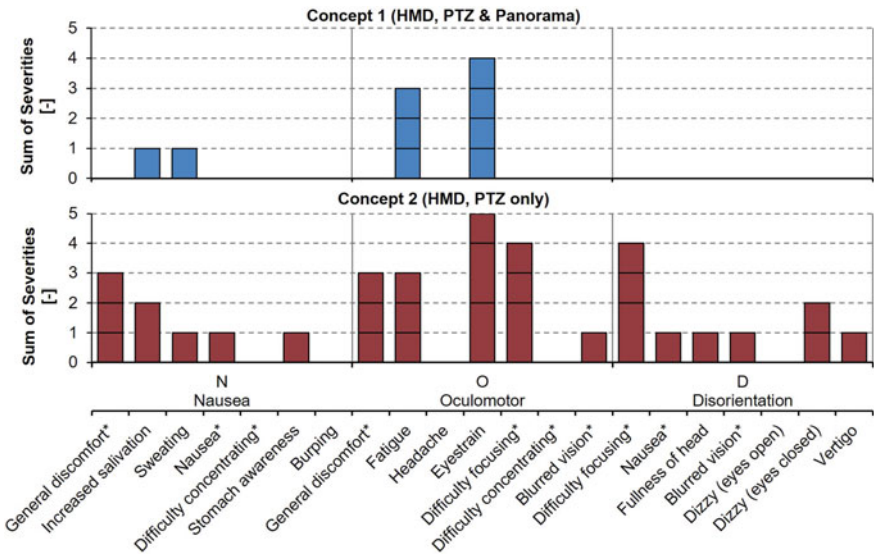
Fig. 7 Number and severity of symptoms reported by n = 5 participants answering the SSQ for concept 1 (top) and concept 2 (bottom), individual column of the stacked columns represents the severity reported by one participant, symptoms counting towards two scales are marked with '*'

participant. The number of stacked columns indicate the number of participants reporting the respective symptom and total height of the stacked columns represent the total severity.

Overall, there are more and higher severity symptoms reported for the 'PTZ only' concept than for the 'PTZ & Panorama' concept throughout all subscales. In the Nausea subscale, the symptoms 'increased salivation' and 'sweating', each reported once with a severity of one ('low') for concept 1, are still each reported once for concept 2, but with higher severity for 'increased salivation' and in addition to three more symptoms being reported. Two Oculomotor symptoms ('fatigue' and 'eyestrain') are recorded, each with low severity, for concept 1 for three and four participants, respectively. For concept 2, fatigue is reported as for concept 1 and eyestrain is reported by one less participant, twice with medium severity and once with low severity. Additionally, the symptoms 'general discomfort', 'difficulty focussing', and 'blurred vision are reported with a total severity of three, four, and one respectively. In contrast to no symptoms for the 'PTZ & Panorama' concept, six of seven Disorientation symptoms are reported for the 'PTZ only' concept. Among these, the highest total severity, a value of 4, is reported for 'difficulty focusing'.

In summary, it is observed that the highest total severities for both concepts are reported for the Oculomotor subscale. For concept 2, it is furthermore noted that the symptoms with the highest total severities in the Nausea and Disorientation subscales—i.e. general discomfort and difficulty focussing—also count towards the Oculomotor subscale.
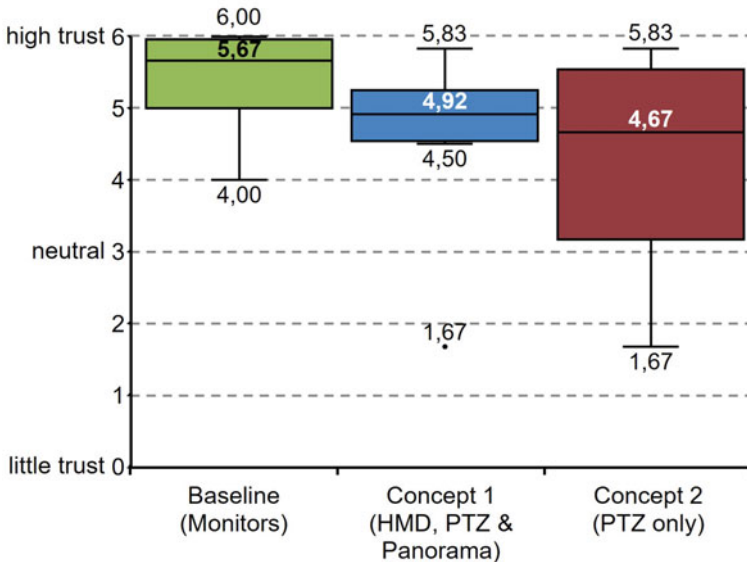

Fig. 8 SATI scores for n = 8 participants with median written in bold

SHAPE Automation Trust Index. In Fig. 8, the results of the SATI questionnaire for the baseline and two developed visualization concepts are presented. As marked on the ordinate, low scores are associated with little trust in the system while high scores equal high trust.

The highest median trust is recorded for the baseline with a score of 5.67. For concept 1, the score drops to 4.92 and to 4.67 for concept 2. All median scores lie in the area of high trust. The scores calculated for the baseline are all higher than the neutral point of three and range between 4.00 and 6.00, equalling maximum trust in the system. The spread of the 'PTZ & Panorama' concept's scores is narrower ranging from 4.50 to 5.83—also lying in the area of high trust—with one outlier present at a score of 1.67 equalling little trust in concept 1. The scores for the 'PTZ only' concept are distributed over a wide range between 1.67—denoting little trust—and 5.83—close to the highest possible trust.

Possible Worktime. In Fig. 9, the participants' estimates of how long they would be able to continuously work with the baseline and the two concepts is depicted. The presented results are determined for eight of the nine participants, with the estimates of the outlier identified on the basis of the SUS (see Sect. 4.2) excluded.

It is generally visible that the participants' estimates are the highest for the baseline and then drop to 'PTZ & Panorama' concept and the 'PTZ only' concept. The median estimate for the baseline is 300 min, while those for the two concepts are below one quarter of that at 75 min for concept 1 and 30 min for concept 2. The values' spread is the highest for the baseline ranging from half an hour to eight hours of continuous
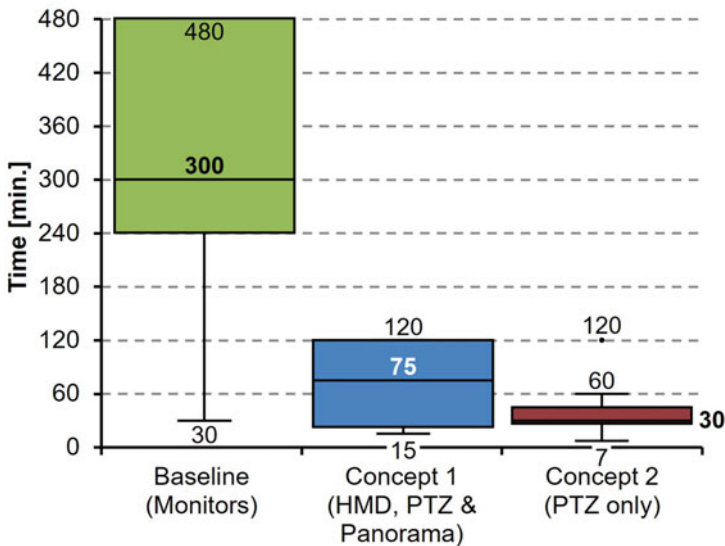
Fig. 9 Maximum possible worktime estimated by n = 8 participants for the baseline and the two concepts, median marked in bold
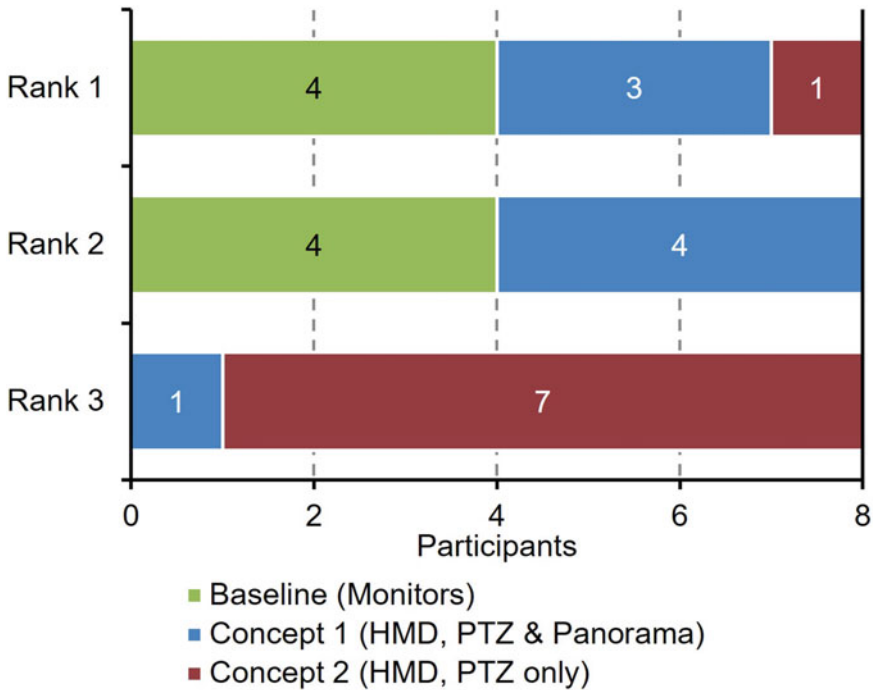
Fig. 10 Ranking of the baseline and two concepts against each other for n = 8 participants

work. The estimates for concept 1 range from 15 to 120 min and those for concept 2 from 7 to 60 min. For concept 2, there is one outlier, i.e. an estimate more than 1.5 times the interquartile distance from the top end of the third quartile, at 120 min.

Ranking of concepts. The summary of how the participants rank the baseline and concepts against each other is depicted in Fig. 10 for eight participants, excluding the identified outlier (discussed in Sect. 4.2).

As can be seen, the baseline is preferred by half the participants as visualization of the video feeds. Of the other half, three participants prefer the 'PTZ & Panorama' concept and one operator prefers the 'PTZ only' concept. The second rank is equally split between the baseline and concept 1. For the majority of participants, concept 2 is the least preferred one. One participant dislikes concept 1 the most.

4.3 Discussion

System Usability Scale. From the presented SUS scores, several conclusions regarding the operators' acceptance of the low-cost camera set-up and the proposed visualization concepts can be drawn. Firstly, the high scores for the baseline, equaling an overall good to excellent usability, indicate that the technical specifications of

the camera set-up may be acceptable for the purpose of proving remote control to uncontrolled aerodromes.

Secondly, the two concepts designed for an HMD are judged to have good usability therefore pointing towards a virtual environment being an acceptable display solution. In this context, it is relevant to consider that the ATCOs among the validation participants, who represented a majority (seven vs. two), may place higher personal requirements on the concepts due to their usual work environment than the AFISOs. Even though it was emphasized that the solution is intended for AFIS and UNICOM aerodromes and occasional use for singular aircraft movements, the ATCOs may therefore value the concepts as having lower usability than the AFISOs. In turn, the good usability found in this study allow for the conclusion that the solution will be sufficiently usable to the AFISOs for whom it is designed.

Judging from the SUS scores, the overview provided by the 'PTZ & Panorama' concept and the PTZ-camera's positioning being decoupled from the operator's head turns are preferred over the simpler 'PTZ only' concept and the PTZ-camera moving in parallel with head turns. In the subsequent sections, a detailed comparison and analysis of potential issues is carried out.

Simulator Sickness Questionnaire. Based on the SSQ scores, conclusions regarding the cybersickness experienced by the validation participants can be drawn. As explained before (see Sect. 3.3), cybersickness is a special form of motion sickness that is induced by the immersion in virtual environments. In this regard, the zero SSQ scores for the baseline system were expected as no VE was used. However, the analysis of the SSQ scores and symptom profiles determined for the two HMD-based concepts allow for conclusions about issues that not only lead to cybersickness, but also influence the previously discussed usability.

For the 'PTZ & Panorama' concept, the low SSQ scores throughout all scales—mostly below half the critical score—point towards little to no cybersickness being experienced by the operators. This is supported by the fact that the experienced symptoms are of low severity (compare Fig. 7). The symptoms most experienced by the participants are fatigue and eyestrain, both solely counting towards the Oculomotor subscale. This observation indicates that using the HMD leads to more strain on the operators' oculomotor apparatus than using the baseline, i.e. monitors, which is a commonly known side effect of immersion in VEs and not specific to visualization concept 1. While these conclusions do explain the reduced usability compared to the baseline, no health and safety issues are expected to arise from the symptoms experienced using the concept and the operators' performance may only be slightly degraded.

In contrast, the maximum SSQ scores for the 'PTZ only' concept exceed the critical score of 30.00 in all scales. This is also the case for the top parts of the third quartile leading to the conclusion that high and, for some of the participants, critical cybersickness is experienced. The spread of the scores, however, indicates that the experienced cybersickness is highly depended on the individual, so that one operator's SSQ score in the Disorientation subscale is 0.00 while another one's is 55.68 and above the critical score. This observation is in line with research on cybersickness that found a high dependency of the symptoms experienced on the individual

(Stanney et al., 1998). From comparing the individual symptoms experienced when working with the two HMD-based concepts, it can be concluded that the higher SSQ scores for concept 2 are due to more and higher severity symptoms experienced across all three subscales. As for concept 1, the symptoms with the highest total severity count towards the Oculomotor subscale. Considering that more symptoms than just fatigue and eyestrain are experienced and that eyestrain is rated as medium severe by two participants, they not only result from using an HMD in general, but are also attributed to the visualization concept. Additionally, symptoms reported in the Disorientation subscale—of which there were none for the 'PTZ & Panorama' concept—point towards conceptual issues inducing cybersickness as well. These may be the lack of overview as no panorama is provided and a delay between the operator's input (i.e. a head turn) and the corresponding movement of the camera, explaining the participant being disoriented and more stress on the oculomotor apparatus than when using concept 1. Future iterations of this concept could combat the problem with an improved technical implementation of the PTZ-camera and by presenting the operator with an overview of the aerodrome, as already considered in the design-space (see Fig. 3). In its current implementation, the 'PTZ only' concept poses a health and safety risk for the operator and also reduces their productivity, explaining the in comparison lowest usability score.

SHAPE Automation Trust Index. Although no sophisticated automations, like data fusion or automatic object following used in Remote Tower Optical Systems, were implemented in the tested prototypes, sufficient trust in the (image) data and means of control provided is important. Clearly, the trust in the baseline is sufficiently high and in median close to the highest possible trust. The reasons for this are twofold. The first reason is that the baseline is derived from state-of-the-art Remote Tower Optical Systems which are highly validated and trusted in. The second reason is the simplicity of the system using two ubiquitous tools—a computer mouse and monitors—without intermediary data processing. The high trust in the baseline concept highlights that the set-up of panoramic camera and PTZ-camera is accepted by the operators and sufficient for providing remote surveillance capabilities to uncontrolled aerodromes. This finding is in line with the conclusions drawn for the SUS scores.

In contrast to the baseline, the cameras' video feeds and the operators' input (i.e. head turns) are processed in the HMD-based concepts. While the processing of the image data is marginal—solely consisting of the steps necessary to display them in the virtual environment, without pixels being altered—the head turns are translated to PTZ-camera positioning commands in multiple steps, including data smoothing. Consequently, the quality of these steps, i.e. if the result is the one expected by the operator, highly influence the trust they place in the concepts.

As the trust the participants place in the 'PTZ & Panorama' concept is high, the processing steps seem to be understandable to the operator. One further reason for the high trust may be the largely unaltered panorama that provides a fallback if issues with the PTZ-camera arise. In summary, no major adjustments to the general concept or the processing steps are necessary.

On the other hand, the trust in the 'PTZ only' concept varies widely between participants. This may be due to issues in the translation of the operators' input into

camera positioning commands leading to delayed movements and/or unexpected results, that are not equally experienced by all participants. Furthermore, the ATCOs participating in the study may place stricter requirements on the system than the AFISOs, leading to a lower SATI score. In both cases, improvements to the concept and providing more information on how the visualization concept works in order to enhance the understandability are required.

Possible Worktime. In addition to the visualization concepts being usable, trusted, and posing no health and safety risks, the operators must be able to work with them for as long as is required for the use-cases presented in Sect. 2.2. Based on the participants' estimates, this clearly is the case for the baseline. As the median possible worktime is five hours, the baseline would be suitable not only for the on-demand and ad-hoc use-cases (use-cases 1 and 3), but also a contiguous remote surveillance needed for use-case 2. However, this conclusion stands in contrast to one operator's estimate of 30 min which cannot be explained with the observations made for the SUS and SSQ scores. Still, this estimate is sufficiently high to allow for temporary use as in use-cases 1 and 3.

In the worktime estimates for the two HMD-based concepts, the conclusions previously drawn are reflected. Even though no specific issues with the first visualization concept could be identified, the use of a Head-Mounted Device limits the participants' estimates to a maximum of 120 min. Considering that this concept was primarily designed for compactness and portability, targeting the first and third use-case, the lowest estimate (15 min) is deemed to be high enough for a temporary provision of air traffic services as in the respective use-cases. Depending on the individual adaption to using an VE, some restrictions may have to be place on the aircraft sequence as to avoid repeated exposures in close sequence and provide the operator enough time to recover from the immersion.

For concept 2, the experienced cybersickness reduces the participants' worktime estimates to half of those for concept 1. In this case, the minimum estimate (seven minutes) cannot be regarded as enough time to get an overview of the traffic situation and then guide one landing or departing aircraft. Therefore, the 'PTZ only' concept is not applicable to the use-cases outlined in Sect. 2.2. Even if it is assumed that the possible worktime increases with training and adaption to the Virtual Environment, higher restrictions than for the 'PTZ & Panorama' concept would have to be put in place to ensure a proper recovery of the operator after an exposure.

Ranking. Complementing the preceding discussions, the analysis of the participants' ranking of the concepts provides insights of how they judge the baseline and the two concepts against each other compared to what is deduced from the other metrics. In this regard, the equal preference of baseline and HMD-based concepts on the first two ranks stands in contrast to the previous conclusions. With consideration of lower usability and trust, experienced cybersickness, and shorter possible worktime, it is expected that the baseline is preferred over the other concepts. Based on the collected data, no reason for the equal preference could be deduced. One influencing factor may have been the novelty of an HMD-based approach to remote surveillance being attractive to some of the participants.

Table 2 Summary of discussion results for baseline and two concepts

Metric	Baseline	Concept 1	Concept 2
SUS	Sufficient	Sufficient	Barely sufficient
SSQ	Sufficient	Sufficient	Not sufficient
SATI	Sufficient	Sufficient	Barely sufficient
Worktime	Sufficient	Limited	Not sufficient
Ranking	Sufficient	Sufficient	Not sufficient
Overall	Sufficient	Limited	Not sufficient

The 'PTZ only' concept being least preferred, however, is in line with the previous findings of low usability, little trust, and critical cybersickness. Nevertheless, one participant prefers this concept which may be explained by the simplicity of only needing to turn one's head in the desired direction and the PTZ-camera following without additional input required.

Summary. Summarizing the previously reached conclusions, concept 1 of the investigated visualization solutions is evaluated sufficiently high to be considered as a potential future candidate for providing AFIS remotely. Even though the computer monitor-based baseline received the highest ratings throughout all metrics and caused no cybersickness, it is per se not suitable for the identified use-cases (see Sect. 2.2), especially not for an ad-hoc provision of ATS. In contrast, visualization concept 1 was specifically designed for these use-cases and has proved to have sufficiently high usability and trust scores as well as to cause no critical cybersickness. With consideration of the operators' possible worktime estimates, this concept supports all targeted use-cases. However, depending on the operators' adaption to the virtual environment, minor restrictions may have to be put in place to prevent critical cybersickness after long use of the system or repeated use in short succession. For visualization concept 2, barely sufficient usability and trust scores, critical cybersickness—especially affecting the oculomotor apparatus—and a very low estimated worktime were recorded making it insufficient to be used for any of the foreseen use-cases.

A summary of the conclusions for all metrics is presented in Table 2.

5   Comparison with Standard Remote Tower Solutions

Since the preliminary validation indicates that the developed low-cost Remote Tower solution is suited for remote surveillance of uncontrolled aerodromes, one possible set-up is presented in this section and compared to the standard Remote Tower concept in the areas of costs and controller working position.

As outlined in Sect. 3.1, the low-cost camera set-up is designed in analogy with standard Remote Towers and comprises one PTZ-camera and one panoramic camera. Due to its cost effectiveness and low bandwidth requirements, a second panoramic camera may be added to the set-up, as is pictured in the left part of Fig. 11, to capture a full 360° panorama without compromising on the low-cost idea. In the right part
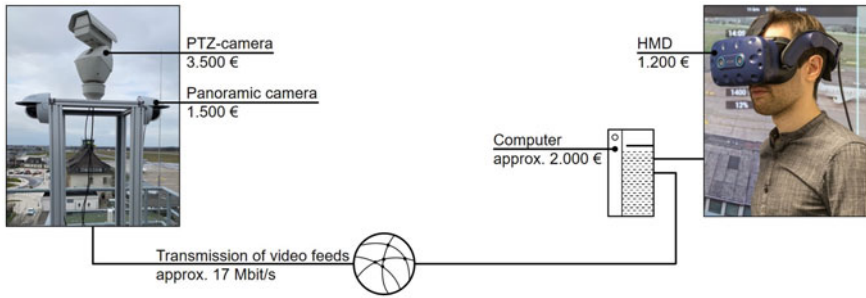
Fig. 11 Possible set-up and costs of the proposed low-cost Remote Tower solution

of the figure, an HMD and the computer driving it are shown as display solution. Between both elements, a wired or wireless network connection is established for transmission of the video feeds and PTZ-camera position commands.

This low-cost solution enables cost reductions in multiple areas compared to standard Remote Tower systems. The first reduction is in implementing cost primarily influenced by the components used. Due to the simpler camera set-up and the use of an HMD as display solution, the total component cost for the pictured set-up amount to approximately 10,000 Euros compared to usually 100,000 Euros or more for Remote Tower Optical Systems. The use of commercial off-the-shelve (COTS) components also reduces maintenance cost. Additionally, the bandwidth requirements for the transmission of the video feeds are lower given the use of fewer cameras with less demanding specifications. In the configuration currently used in the prototypic set-up at the Braunschweig-Wolfsburg aerodromes, a bandwidth of 17 Mbit per second is required so that a wireless transmission via the cellular network (at least 4G speeds) is also thinkable. This option is of special benefit for uncontrolled aerodromes as they are often located in rural areas potentially not offered with a sufficient wired connection. The lower bandwidth requirements furthermore reduce the recurring operating costs.

In the second area of comparison, the CWP, the use of an HMD as display solution offers two benefits: Portability and flexibility designing the CWP. The portability is achieved by using an HMD instead of multiple monitors as in standard Remote Tower systems. While a wired HMD and external computer were used in the prototypic set-up, the proposed low-cost Remote Tower solution can easily be implemented on a wireless HMD allowing the operators to freely move around. While this is a perquisite for the third use-case—ad-hoc establishment of visual contact with aircraft when not on the tower (see Sect. 2.2), the other use-cases also benefit from the compactness of the system, for example enabling an easier implementation at existing facilities.

In addition to eliminating the need for multi-display set-ups, the use of a virtual environment also enables flexibility in designing the CWP. One option is to rebuild a familiar, location fixed working position similar to Remote Tower display systems which may improve the acceptance and does not require a change in working procedures. However, innovative concepts, such as those presented in

(Hofmann et al., 2020), are also possible and reduce the 'head-down-times' during which the operator loses visual contact with an aircraft in order to scan other information sources.

6 Summary and Outlook

In this chapter, the design of a low-cost Remote Tower solution targeted at non-ATC, uncontrolled aerodromes is presented. It is based on a site survey performed at the Schönhagen aerodrome in Germany during which two constraints—operators having secondary responsibilities and limited budget—and three potential use-cases—an extension of opening times, ad-hoc, and on-demand provision of aerodrome flight information services—were identified by observing the AFISOs' day-to-day work and in discussion with them.

Then, a low-cost camera set-up was defined by simplifying the usual components of standard Remote Tower systems, i.e. the visual panorama, potential infrared panorama, and the PTZ-camera. Utilizing this set-up, three visualization concepts were developed. The visualization concepts consist of a baseline, defined after Remote Tower display solutions and utilizing standard computer monitors, and two novel visualization concepts. The first of these displays the video feed of the PTZ-camera layered onto the feed of the panoramic camera. The second relies solely on the PTZ-camera's video feeds and couples its positioning with the operator's head turns.

The results of a System Usability Scale, Simulator Sickness Questionnaire, and tailor-made questions of the estimated possible worktime and personal ranking of the baseline and two concepts were analyzed. High usability scores found for the baseline and the 'PTZ & Panorama' concept indicate that the technical specifications of the camera set-up are acceptable to the operators and an HMD is an acceptable display solution. The 'PTZ only' concept on the other hand was found to cause critical cybersickness likely resulting from a lack of orientation points in the VE, also resulting in higher strain on the oculomotor system than concept 1. It is furthermore concluded from the tailor-made questions that operators are able to comfortably work with the 'PTZ & Panorama' concept for as long as is required for the ad-hoc and on-demand use-cases—for which it is designed—and that an HMD-based display solution is acceptable to them.

Finally, the findings were compared to the Remote Tower concept in the areas of cost and controller working position. It was deduced that there is a high cost reduction both, in implementation and operating costs. The use of an HMD as display solution also may allow for a compact and portable system that is easy to carry around or to connect to existing Remote Tower Centers.

In summary, the presented low-cost Remote Tower solution was found to be generally suited for remote surveillance of uncontrolled aerodromes. The 'PTZ & Panorama' concept (concept 1) in particular was found to be a suitable approach for the design of a compact and portable display solution and should be further advanced.

Necessary improvements to this concept may be derived from the analysis of objective data also recorded during the validation study. In a further step, validations with more diverse traffic situations and a more sophisticated prototype have to be carried out.

References

AXIS Communications AB. (2020). Datasheet AXIS P3807-PVE network camera. [Cited: 21. August 2020]. https://www.axis.com/files/datasheet/ds_p3807pve_t10108840_en_2007.pdf.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean. *Journal of Usability Studies, 4*(3), 114–123.

Barrett, J. (2004). *Side Effects of Virtual Environments: A Review of Literature.* Edinburgh, Australia: DSTO Information Sciences Laboratory.

Brooke, J. (1996). *SUS—A quick and dirty usability scale. Usability Evaluation In Industry.* London: Taylor & Francis Ltd., pp. 189–194.

Bundesaufsichtsamt für Flugsicherung (BAF). (2021). Map of AFIS aerodromes in Germany. 01. April 2021. [Cited: 21. April 2021]. https://www.baf.bund.de/SharedDocs/Downloads/DE/Publikationen_BAF/BAF_KarteAFIS.pdf?__blob=publicationFile&v=4.

Dehn, D. M. (2008). Assessing the Impact of Automation on the Air Traffic Controller: The SHAPE Questionnaires. *Air Traffic Control Quarterly, 16*(2)*,* 127–146.

Ellis, S. R., & Liston D. B. (2016) Visual features used by airport tower controllers: some implications for the design of remote towers. In N. Fürstenau (Ed.). *Virtual and Remote Control Tower* (pp. 21–51), s.l: Springer International Publishing.

Eurocontrol. (2012). SESAR HP repository: SATI—SHAPE automation trust index. 22. October 2012. [Cited: 9. July 2020] https://ext.eurocontrol.int/ehp/?q=node/1594.

European Aviation Safety Agency (EASA). (2016). *Notice of Proposed Amendment NPA 2016–09(B).*

European Organisation for Civil Aviation Equipment (EUROCAE). (2018). *Minimum Aviation Systems Performance Standard for Remote Tower Optical Systems ED-240A.* 2018.

Federal Aviation Administration (FAA). (2021). Aeronautical Information Manual (AIM): Chapter 4—Section 1. Services Available to Pilots. 31. December 2020. [Cited: 29. March 2021]. https://www.faa.gov/air_traffic/publications/atpubs/aim_html/chap4_section_1.html.

Hagl, M., et al. (2018). Augmented Reality in a Remote Tower Environment Based on VS/IR Fusion and Optical Tracking. In Don Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics.* 2018, pp. 228–571.

Hofmann, T., et al. (2020). Design and implementation of a virtual workstation for a remote AFISO. In *HCI International 2020—Late Breaking Papers: Virtual and Augmented Reality*. Copenhagen (virtual): Springer International Publishing.

Jakobi, J., & Hagl, M. (2020). Which minimum frame rate is needed in a remote tower environment? In N. Fürstenau, J. Jakobi & A. Papenfuss (Eds.), *Virtual and Remote Control Tower* (2nd Edn.), s.l: Springer International Publishing.

Kennedy, R. S., et al. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220.

MacLean, A., et al. (1991). Questions, options, and criteria: elements of the design space analysis. s.l.: L. Erlbaum Associates Inc., 1991. *Human-Computer Interaction, 6*(3), 201–250.

McCauley, M. E., Sharkey, T. J. (1992). Cybersickness: Perception of Self-Motion in Virtual Environments. *Presence: Teleoperators and Virtual Environments, 1*(3), 311–318.

Papenfuss, A. & Friedrich, M. (2016). Head Up Only - A design to enable multiple remote tower operations. In 2016. In *Proceedings of the IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)* (pp. 1–10). Sacramento, CA, USA: Institute of Electrical and Electronics Engineers (IEEE).

Pelco Inc. (2020). Datasheet Pelco Esprit Enhanced series IP positioning system. [Cited: 21. August 2020]. https://media.pelco.com/wp-content/uploads/2020/04/17201235/C4046_Esprit-Enhanced_PTZ_Series_040820.pdf.

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality, 20*, 101–125.

Regan, C. (1995). An investigation into nausea and other side-effects of head-coupled immersive virtual reality. *Virtual Reality, 1*, 17–31.

Reuschling F., et al. (2020). Designing and evaluating a fusion of visible and infra-red spectrum video streams for remote tower operations. In N. Fürstenau, J. Jakobi & A. Papenfuss (Ed.). *Virtual and Remote Control Tower*, 2nd Edn. s.l. : Springer International Publishing.

Schier, S. (2016). Remote tower simulation environment. In N. Fürstenau (Ed.), *Virtual and Remote Control Tower* (pp. 69–85). s.l: Springer International Publishing.

Stanney and Key. (1995). Realizing the full potential of virtual reality: human factors issues that could stand in the way. In *Proceedings of the Virtual Reality Annual International Symposium (VRAIS '95)*. Research Triangle Park, North Carolina, USA : s.n.

Stanney, K. M., Kennedy, R. S., Drexler, J. M. (1997). Cybersickness is not simulator sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 41*(2).

Stanney, K. M., Mourant, R. R., Kennedy, R. S. (1998). Human Factors Issues in Virtual Environments: A Review of the Literature. *Presence: Teleoperators and Virtual Environments, 7*(4), 327–351.

Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley Publishing Company.

Ultraleap Limited. (2021). Product Page: Leap Motion Controller. [Cited: 29. March 2021]. https://www.ultraleap.com/product/leap-motion-controller/.

Unity Technologies. (2021). Unity Website. [Cited: 25. February 2021]. https://unity.com.

van Schaik, F. J., et al. (2010). Advanced remote tower project validation results. *IFAC Proceedings Volumes., 43*(13), 135–140.

van Schaik, F. J., et al. (2016). Detection and recognition for remote tower operations. In N. Fürstenau (Ed.), *Virtual and Remote Control Tower.* s.l: Springer Internation Publishing.

Wikipedia. (2021). List of commercial and special airfileds in Germany. 18. March 2021. [Cited: 21. April 2021]. https://de.wikipedia.org/wiki/Liste_der_Verkehrs-_und_Sonderlandepl%C3%A4tze_in_Deutschland.

Yuan, J. et al. (2018). The visual effects associated with head-mounted displays. *International Journal of Ophthalmology and Clinical Research, 5*(2).

Appendix A
Basic Optics for RTO Videopanorama Design

Norbert Fürstenau

In this technical appendix we will provide some basic optics for the design of a digital video panorama reconstruction of the out-of-windows view with a visual resolution comparable to the human eye, i.e. an angular resolution of the order of $\alpha_E \approx 1$ arcmin $= 1/60° = 0.3$ mrad, corresponding to 30 cm object size at 1 km distance. Besides satisfying resolution and contrast requirements the system design, including data processing and transmission infrastructure should be realizable under acceptable cost which has to be considerably below that one of a physical control tower building of several M€ for a medium size airport. Typically each single camera of the panorama system has its own display at the operator working position, plus hard- and software for high performance image processing at the transmitter and receiver side of the high bandwidth data transmission infrastructure.

A.1   Geometrical Optics for Panorama Design and Pixel Resolution

Estimates for the camera optics design including the visual (pixel-) resolution of the RTO-videopanorama may be derived to first order from the simple paraxial geometrical optics approximation (neglecting the wave character of light, see Sect. A.2 for limitation) via Newton's fundamental thin lens equation:

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b}$$

(A.1)

with f = focal width, g = lens − object distance, b = lens − image distance. Combining the ratio of image size / object size B/G = b/g with Eq. (A.1) yields

$$f = \frac{gB}{G + B} \approx \frac{gB}{G} \tag{A.2}$$

with the approximation valid for our case of large distances $g \gg f \approx b$. If we cover 190° visual angle towards the runway by five high resolution HD-cameras with portrait orientation (design for the DLR-DFS validation experiments, see chapters "Remote Tower Prototype System and Automation Perspectives" and "Which Metrics Provide the Insight Needed? A Selection of Remote Tower Evaluation Metrics to Support a Remote Tower Operation Concept Validation") each one covers a horizontal visual angle $2\Theta = 38°$. With portrait orientation of a ¾″ CCD-chip (1080×1920 pixel on (horizontal H × vertical V) = 5.9 mm × 10.6 mm active image size we get for the required focal width (B = H/2):

$$f = \frac{0.5H}{tan(\theta)} = 8.6 \; mm \tag{A.3}$$

The standard zoom camera used in the DLR-system (zoom factor $Z_{max} = 26x$) covered a focal width of $f = 3.5$ mm (wide angle 42°) to $f = 91$ mm (tele visual angle 1.7°). From Eq. (A.2) we obtain at a distance of 500 m a section of 343 m orthogonal to the optical axis covered by the visual angle 2Θ.

In the same way we obtain for the video pixel resolution (pixel width $p = 5.5 \; \mu m \gg g$)

$$g\alpha_V = G_{min} \geq gp/f = 0.64 \text{ m} \tag{A.4}$$

as estimate for minimum video resolution and object size at distance $g = 1$ km respectively, that covers at least 1 pixel. i.e., under ideal illumination and foreground/background contrast conditions it should be possible for an operator viewing a HD-display with well adjusted Gamma correction (see below) to detect at this distance the wheel of a small aircraft. 60 cm at 1 km distance corresponds to $\alpha_V = 0.6$ mrad = 2.1 arcmin (=1/30°) $\approx 2 \alpha_E$, i.e. only half the typical angular resolution of the human eye (see above).

In chapter "Remote Tower Experimental System with Augmented VisionVideopanorama" we used the approximation (A.3) to estimate the field of view and pixel resolution for given focal width as dependent on the chip size H or pixel distance B = H/2 from the optical axis

$$\theta(B) = \arctan\left(\frac{B}{f}\right) \tag{A.5}$$

This FOV-angle (FOV = 2Θ) is plotted in Fig. 5 of chapter "Remote Tower Experimental System with Augmented VisionVideopanorama" as function of focal width for the panorama and PTZ CCD-chip size. Differentiation yields the corresponding dependence of pixel-FOV $\Delta\Theta$ on distance B from the axis (p = ΔB = 7.5 μm):

Fig. A.1 Decrease of pixel FOV with distance H from optical axis for thin lens with f = 12.5 mm and image chip pixel width p = 7.5 μm
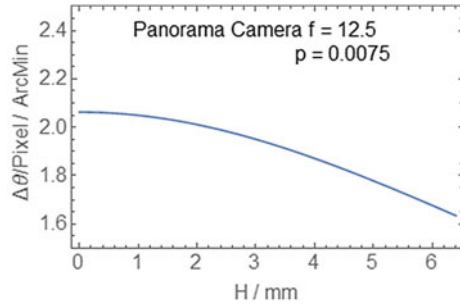
$$\Delta\theta_p = \frac{p}{f}\left[\frac{1}{1+(H/f)^2}\right] \tag{A.6}$$

which for our initial experimental system yields near the optical axis (H = 0) $\Delta\Theta_{min}$ = 7.5 μm/12.5 mm = 2.1 arcmin, corresponding to the above mentioned minimum resolvable object size. The decrease of $\Delta\Theta_{min}$ with distance H from the optical axis is depicted in Fig. A.1:

For large distance lim $\Delta\Theta(H \rightarrow \infty) = 0$ and moreover the paraxial approximation looses its validity with increasing H. Due to the decrease of $\Delta\Theta_p$ with H the received light power per pixel from the corresponding object area decreases accordingly which reduces the contrast towards the chip boundaries. This is only one simple example for (nonlinear) dependencies of image properties on viewing angle and it underlines the necessity to carefully specify, test, and characterize the selected electronic (image-chip and pixel type and size, signal-to-noise ratio, etc.) and optical camera components (lens system, quality of corrections for image distortions, MTF (see below)).

In reality the ideal pixel resolution is hardly achieved anywhere on the whole image area due to limited contrast as quantified by the modulation transfer function (MTF, see below). It corresponds to the Nyquist limit of black-white line pair resolution, with line width = pixel size. The realistic value depends on several additional camera and digital processing parameters. One important camera parameter is the f-number f# (the aperture-stop number, typically 1.4 … 22), defined as ratio of focal length to aperture stop diameter D (Hecht & Zajac, 1974):

$$f\# = f/D \tag{A.7}$$

Minimum and maximum for f = 8.6 mm are f# = 1.4 and 22 corresponding to aperture diameters D_{max} = 6.1 and 0.4 mm respectively. In practice the camera requires motor driven automatic iris control to adjust for (rapid) illumination changes (bright sun, clouds, shadow, etc.). If under bright illumination conditions D becomes small, the depth of focus becomes larger so that sharpness of image details increases.

On the other hand a resolution problem may arise due to diffraction effects originating from the wave character of the light.

The above first order paraxial approximation neglects all lens distortions (chromatic and monochromatic aberrations, e.g. (Hecht & Zajac, 1974)) which influence the imaging quality of the optical system. From a systems point of view the design task also includes an optimal combination of electronic and optical camera component, i.e. chip technology including size and signal/noise level, and type of objective (including automatic aperture control for quick adaptation to changing illumination).

A.2 Diffraction Limit and Resolution

With decreasing aperture diameter D the wave character of light (wavelength λ) plays an increasing role so that diffraction effects may begin to limit resolution. The (Fraunhofer or farfield) diffraction limited angular resolution $\Delta\alpha$ and corresponding "Airy disk" blur-radius q_1 depends on λ and is defined via the "Airy" or point-spread function (focal image of a distant light source via a well corrected (aberration free) optical system, (Hecht & Zajac, 1974))

$$q_1 = f\Delta\alpha = 1.22\lambda \, f/D = 1.22\lambda \, f\# \tag{A.8}$$

It describes the spread of the distant point light source (blurring) as distance between intensity maximum at the optical axis and the minimum between the axis and the first of a series of circular diffraction fringes (first minimum of dark ring) of a circular aperture. The resolution improves ($\Delta\alpha$ decreases) with shorter wavelength and with decreasing f#. This means that for the visible spectrum ($\lambda \approx 0.4 - 0.7 \, \mu m$) with green light (0.6 μm) we get an Airy disk (blur diameter) of the order of f# μm, i.e. between 1 and 20 μm. For f# > ca. 6 the spread is larger than the pixel size (for λ > 0.7 mm = green) so that for small aperture diameter (bright light conditions) with constant exposure time the image quality decreases. Of course with slow movement scenarios longer exposure times may be selected with smaller f# in order to keep illumination constant and improve resolution.

A.3 Contrast and MTF

Contrast is defined via the object (foreground) and background intensity or brightness which for human perception has to be transferred into the subjective luminance value (see below). The classical contrast measurement is based on alternating black-white line pairs of different width (spatial frequency) where contrast c is defined as

$$c = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{A.9}$$
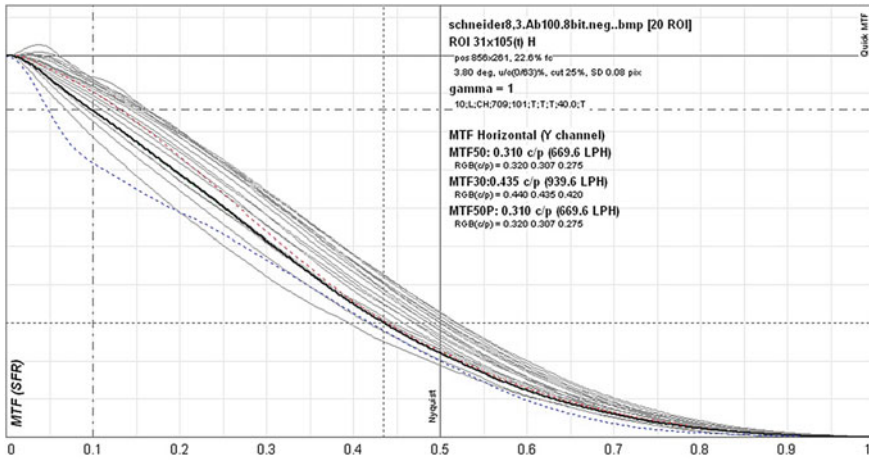
Fig. A.2 Repeated measurements of modulation transfer function of one of the HD cameras used in the prototype setup of the DLR-DFS validation experiment. MTF50 = 0.31 cycles/pixel derived from 939.6 line pairs over the analysed image range

with maximum value $c_{max} = 1$ for $I_{min} = 0$ (ideal black) and $c_{min} = 0$ for equal max/min luminance $I_{max} = I_{min}$. In real systems c may vary across the image diameter (see above), i.e. it is a function of the image coordinates $c = c(x, y)$. A contrast measurement of black-white line targets over a range of spacial frequencies yields the modulation transfer function MTF) that quantifies the optical system with regard to contrast. Because a line pair requires at least two pixels of the image chip to be resolved, the pixel width as minimum resolvable line width defines the Nyquist limit of the system for object detection under ideal luminance conditions.

Figure A.1 depicts an example measurement with one of the HD cameras used in the validation experiments (Mahmoudzadeh Vaziri, 2013). Shown is the horizontal resolution. The MTF-calculation is based on the Fourier transformation of the image into the spatial frequency domain using a commercial software (Quick MTF: (2013)). One recognizes the typical decrease of contrast down to ca. 20% of maximum at the Nyquist limit of 0.5 cycles/pixel (1 pixel/line)

For the characterization of the system typically MTF50 (spacial frequency at 50% of maximum contrast) and MTF30 values are used as indicated in the figure legend: MTF50 at 0.31 c/p, MTF 30 at 0.435 c/p.

A.4 Determining Effective Resolution by Detectability Experiments

The initial experimental DLR-videopanorama system was used for determining the effective video resolution under realistic conditions (see chapter "Remote Tower

Experimental System with Augmented VisionVideopanorama"), i.e. the difference between real detectability of small objects in the airport environment from the ideal Nyquist limit (pixel resolution) of the camera system. Different pre-determined flight situations or events during airport circling of a DLR aircraft were used to measure time-of-detection by human observers (e.g. first A/C-detection during approach, determination of landing gear-up/down). For this purpose, based on Eq. (A.4) for idealized pixel resolution $\alpha^V/\alpha^E \approx 2$, we derived a relationship for the difference of object (or event i) detection distance under video replay and under direct visual observation respectively $\Delta x_{i = =} x_i^E - x_i^V$, as function of object/event distance x_i, with eye resolution/video resolution ratio (a_E/a_V) as constant system parameter. This procedure allows for an averaging of observations with different detectability distances via linear regression. The initial measurement was performed via time differences $\Delta t = t^V - t^E$, with a common time base for observers and aircraft, and aircraft GPS-position x^E correlated with observation time t^E. The corresponding video-observation position $x^V \leq x^E$ was obtained via A/C speed v through $x^V = x^E - v\,\Delta t$, yielding:

$$\alpha^V = G/(x^E - v\Delta t) = \alpha^E(1 - \alpha^E v\Delta t/G)^{-1} \qquad (A.10)$$

With $\Delta x = v\Delta t$ we get for object (event) i at distance x_i

$$\Delta x_i = (1 - \alpha^E/\alpha^V)x_i^E = \beta x_i^E \qquad (A.11)$$

so that an average effective video resolution as obtained from observers detection distances x_i^E, x_i^V during airport circling may be derived from the slope β as

$$\widehat{\alpha}^V = \frac{\alpha^E}{1 - \beta} \qquad (A.12)$$

A.5   Basics of Physiological Optics for Detectability

The subjectively perceived resolution and contrast determining the detectability of static and moving objects depends on the physiological and psychophysical perceptual properties of the human observer. This is taken account of by a transformation of the radiation optics magnitudes (light power, irradiation power, etc.) into corresponding technical optics magnitudes and a number of psychophysical laws. The transformation is based on the spectral sensitivity distribution $V(\lambda)$ of the human eye with a maximum in the green range at $\lambda = 550$ nm, and lower/upper sensitivity boundary between 400 and 800 nm for the bright light adapted eye. It is shifted into the blue range by ca. 50 nm for the darkness adapted eye. (e.g. (Gobrecht, 1978)). This contrasts to the camera's CCD image sensor with a significant sensitivity in the near infrared spectral range which provides an advantage under low brightness conditions.

The proportionality constant for the transformation as derived by integrating $V(\lambda)$ with Planck's famous black body radiation law (spectral radiant density: radiant energy per wavelength or frequency interval (dl), per second (s), per m^2 per spacial angle (sterad, sr)), yielding for the ratio of technical optics magnitude (X_v) units/radiation physics magnitude (X_e) units:

$$K_m = 673 \frac{candela\ steradian\ (cd\ sr)}{Watt\ (W)} = 673 \frac{lumen\ (lm)}{W} \qquad (A.13)$$

for bright light adaptation, e.g. 1 W of radiant power corresponds to 673 lm of perceived power at maximum sensitivity ($V(\lambda) = 1$). For night vision the darkness adapted eye is much more sensitive: $K'_m = 1725$. Several correspondences between the technical optics (physiologically relevant) and radiation physics magnitudes are established via K_m and K'_m:

- radiant intensity I_e / W/sr   luminous intensity I_v / Candela cd

- radiation power Φ / W | luminous flux Φ_v / lumen (lm)

- radiation density L_e / W/(sr m^2) | luminous density L_v / cd/ m^2

- emittance M_e / W/m^2 | radiance M / lm/m^2

- irradiance (intensity) E_e / W/m^2 | illumination E / lm/m^2 = Lux (lx)

- irradiation H / Ws/m^2 | exposure H / lx s

A.6 Luminance Sensitivity and Gamma Correction

Individual γ-values quantify the nonlinear luminance sensitivity characteristic of camera, display and human observer which classically is described by the Weber-Fechner law as logarithmic stimulus (S) – (subjective) response (E) function

$$E = \log S \qquad (A.14)$$

An improved version of the nonlinear functional relationship between stimulus strength S and luminance sensitivity E which better matches the empirical findings is given by the Stevens function (Birbaumer & Schmidt, 2010)

$$E = k\ S^\gamma \qquad (A.15)$$

A most natural video reconstruction of the real scenery should realize a linear relationship between system input and visual output. For the display with digital input signal I_{in} and output intensity I_{out} the corresponding overall relationship is

$$I_{out}^{Display} = kI_{in}^{\gamma(Display)} = kI_{out}^{\gamma(Display)}(Camera) = kI_{in}^{\gamma(Camera)\gamma(Display)} \quad (A.16)$$

The human observer exhibits a typical value of $\gamma = 0.45$ which should be realized by the camera sensor in order to obtain a natural reconstruction. This results in the typical display γ-setting of

$$\gamma(Display) = 1/\gamma(Camera) \quad (A.17)$$

which yields a good matching of display-γ to human perception if γ(Display) is selected in the range $1.8 < \gamma < 3$, with a typical average value of 2.4. Decrease of display-γ increases the contrast for dark objects in low light level areas while at the same time keeping small luminance differences in light areas in an acceptable level. An example is depicted in chapter "Remote Tower Prototype System and Automation Perspectives", Fig. 14.

References

Birbaumer, N., & Schmidt, R. (2010). *Biologische psychologie* (7th edn.). Springer Medizin Verlag.

Gobrecht, H. (1978). *Lehrbuch der Experimentalphysik III: Optik* (7 edn.). In H. Gobrecht (ed.) Walter de Gruyter.

Hecht, E., & Zajac, A. (1974). *Optics.* Addison-Wesley.

Mahmoudzadeh Vaziri, G. (2013). *Projekt RAiCe: Modulationsübertragungs-funktion (MTF) der Panoramakameras.* Institute of Flight Guidance. German Aerospace Center (DLR).

Quick MTF—An image quality test application. (2013). Retrieved from www.quickmtf.com

Appendix B
Signal Detection Theory and Bayes Inference for Data Analysis

Norbert Fürstenau

This Appendix provides some additional basics of signal detection theory and Bayes inference. Based on experimental data with operational experts, these methods are employed in chapters "Model based Analysis of Two-Alternative Decision Errors in aVideopanorama-Based Remote Tower Work Position" and "Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration during Simulated Aircraft Landing" for quantifying requirements and performance characteristics of the RTO system via analysis of two-alternative decision experiments. More details may be obtained from the references, for our purpose in particular from Robert (2001), Green and Swets (1988), MacMillan and Creelman (2005), Zhang and Mueller (2005). Practical examples are taken from the corresponding previous chapters (prediction errors with RTO-video framerate experiments: chapter "Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration during Simulated Aircraft Landing"; discrimination errors with validation trials: chapter "Model Based Analysis of Two-Alternative Decision Errors in aVideopanorama-Based Remote Tower Work Position") and the related original publications respectively, referenced therein.

B.1 Bayes Inference

Bayes inference was used in preceding chapters to quantify the risk of inferring from an observation with limited evidence for one of two possible world states (e.g. a specific observable aircraft maneuver in the control zone taking place or not), on a false cause which does not correspond to the actual situation. Generally it allows for quantifying the probability of a random event A acting as a cause for another, dependent random event B (an observation) by means of inverting the measured conditional

probability using Bayes theorem. According to this fundamental statistical law the compound probability of two interdependent random variables is given by

$$p(A, B) = p(A \wedge B) = p(A|B)p(B) = p(B|A)p(A) \qquad (B.1)$$

which yields for the conditional probability of A given B, i.e. the conclusion from effect B to probability of the cause A, via inversion of probabilities

$$p(A|B) = \frac{p(B|A)}{p(B)} p(A) \qquad (B.2)$$

as classical version of Bayes theorem for two interdependent random variables. For statistically independent variables Bayes formula reduces to the well known product of independent probabilities p(A,B) = p(A) p(B). For our purpose of analysing two-alternative decision experiments this most simple version of Bayes theorem is sufficient. It is worth mentioning, however, that extensions allow for generating complex Bayesian networks, e.g. for analysing the performance of complex sociotechnical systems such as delay propagation in air traffic networks. For this purpose convenient software tools are available (e.g. the Bayesian belief network tool NETICA (https://www.norsys.com/)).

　　For our purpose we are interested in the probability of a certain world state (situation S_i) given the probability of an observers response R_j, based on her perceived evidence in support of a corresponding hypothesis. We start with the measured estimate of an observers decision making within the context of an observation task. For a two-alternative decision task the response alternatives to two world states, prepared for the experiment as independent (random) situations S_1, S_2 (e.g. landing aircraft with gear up or down) are quantified as estimates of conditional probabilities, abbreviated as Hit rate H, rate of Misses M, correct rejections CR, and False Alarms FA. They are typically defined within a response matrix as follows:

$$\begin{aligned} p(R_1|S_1) &:= H \quad p(R_2|S_1) := FA \\ p(R_1|S_2) &:= M \quad p(R_2|S_2) := CR \end{aligned} \qquad (B.3)$$

with normalization H + M = 1 and CR + FA = 1 for two-alternative decision experiments due to the independence of events S_1, S_2 (Green & Swets, 1988; MacMillan & Creelman, 2005). R_1, R_2 are the two possible alternative subjective responses, based on the subjects hypothesis due to the perceived evidence. It is quite convenient to visualize this matrix within a Venn diagram (Fig. B.1).

　　Corresponding to Eq. (B.2) the probability of causating world state S_i for response R_j is obtained as inversion of the conditional response probability $p(R_j|S_i)$ due to an event observation (e.g. observers response probability approximated by the hit rate H of a decision experiment), given a certain precondition, e.g. the a-priori knowledge of one of the two possible world states (situations) S_1, S_2. This a-priori knowledge
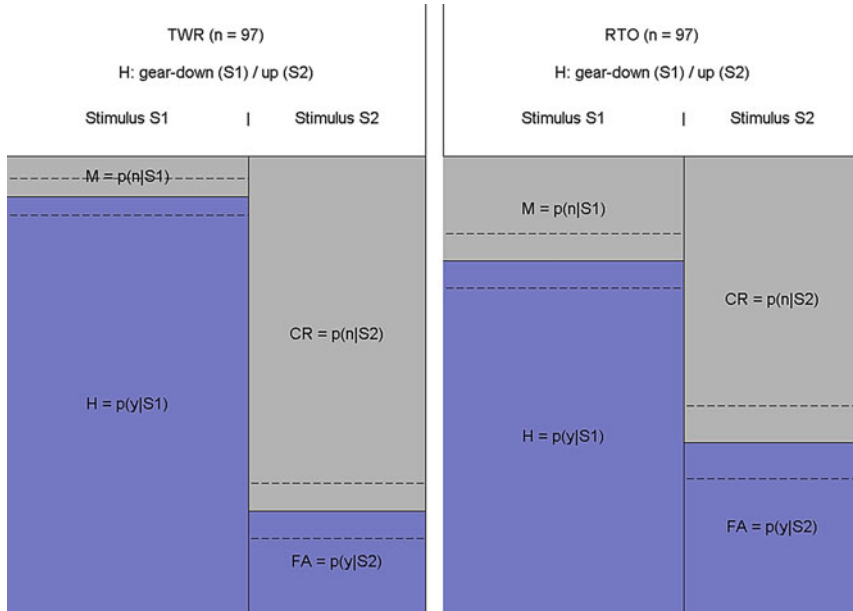
Fig. B.1 Measured response matrices (probabilities) visualized within Venn-diagrams as relative size of respective areas. Here a concrete example is shown, taken from the experimental results of chapter "Multiple Remote Tower Simulation Environment (S. Schier)": Model-based analysis of two-alternative decision making (visual discrimination task: gear-up or -down of approaching aircraft) for quantifying RTO-performance. Areas correspond to probabilities of Eq. (B.3)). Dotted lines indicate standard errors of mean

on S_i is known through the experimental design (relative number N_1, N_2 of situations S_1, S_2). The corresponding a-priori probability $p(S_i)$ is multiplied with the likelihood of the observed evidence $p(R_j|S_i)/p(R_i)$, e.g. for calculating via the Bayes theorem the risk of an unexpected situation S_i (false conclusion on the world state) as cause for the subjective observation (erroneous perception) R_j if $i \neq j$:

$$p(S_i|R_j) = \frac{p(R_j|S_i)}{p(R_j)} p(S_i) \tag{B.4}$$

with situations and responses S_i, R_j; $i, j = 1, 2$. We may choose $R_1 =$ signal detected (or alternative 1), $R_2 =$ noise detected: no signal (or alternative 2). Subjects response probability is $p(R_j) = p(R_j|S_i) p(S_i) + p(R_j|S_j) p(S_j)$, and $p(R_i|S_i) + p(R_j|S_i) = H + M = CR + FA = 1$ (i.e. for a given experimentally determined world state (situation) the subjects decision is either correct or incorrect). From the design of the experiment with $N(S_1) = N_1, N(S_2) = N_2, N = N_1 + N_2$ follows for the prior probabilities $p(S_1) + p(S_2) = 1$.

For practical purpose it is quite often convenient to use Bayes odds as relative measure instead of probabilities. For this purpose Eq. (B.4) may be written with the likelihood ratio (e.g. ratio of Hit-rate to False Alarm-rate) defined by

$$LR_{ji}(R_j) = p(R_j|S_j)/p(R_j|S_i) \tag{B.5}$$

The Bayes inference of Eq. (B.4) can then be expressed using LR

$$p(S_i|R_j) = \frac{p(R_j|S_i)}{p(R_j|S_i)p(S_i) + p(R_j|S_j)p(S_j)}p(S_i) = \frac{1}{1 + LR_{ji}(R_j)\frac{p(S_j)}{p(S_i)}}p(S_i) \tag{B.6}$$

With the prior odds for a two-state world (derived from the known world states with our experimental ratio N1/N2) given by

$$Opr_{ji} = p(S_j)/p(S_i) = p(S_j)/(1-p(S_j)) \tag{B.7}$$

Analogously the posterior odds (ratio of world state probabilities as modified by the hypothesis due to observed evidence based response R_j) is given by

$$OPo(R_j)_{ij} = p(S_i|R_j)/p(S_j|R_j) \tag{B.8}$$

With Eq. (B.6) we obtain from Eq. (B.8) the posterior odds for the world state i contrary to prediction (unexpected situation due to the decision derived from perceived evidence) as:

$$OPo(R_j)_{ij} = \frac{p(R_j|S_i)p(S_i)p(R_j)}{p(R_j)p(R_j|S_j)p(S_j)} = LR_{ij}(R_j)OPr_{ij} \tag{B.9}$$

B.2   Signal Detection Theory

Within psychophysics signal detection theory (SDT) plays an important role in quantifying decision making, in particular for two-alternative experiments. The standard paradigm is to discriminate a signal embedded in a noisy background from the noise without a signal. In the RTO context (see main chapters) we have used this method to discriminate aircraft landing with sufficient braking deceleration (signal case) from landing with too weak braking, leading to runway overrun (noise). Another example was the discriminability of flight maneuvers for quantifying the RTO performance as compared to the standard tower work condition. The unique feature of SDT is its capability to separate the intrinsic "detector" sensitivity or system discriminability of the observer from his subjective preference to judge more conservative (avoiding false alarms at the cost of missing some correct ones, i.e. increasing the number of

misses) or more liberal (preference for identifying as much as possible signals at the cost of increasing the FA-rate).

B.2.1 Parametric Discriminability d' and Subjective Criterion c

For this purpose it has to be assumed that the observers internal response or familiarity with the two alternative signals (S_2 or noise, S_1 or stimulus + noise) is distributed according to a Gaussian density. Discriminability d' and decision criterion c are then defined by means of the z-scores (integration limits) of the inverse cumulative densities. This is visualized with the two density functions in Figure A.2. The subjective criterion at the position c of the familiarity-axis of the two possible random events depicts the integration limits, separating the H and M areas of S1 (right density function) on the one hand, and CR and FA for the S2 density on the other.

The inverse of the normalized cumulative Gaussian densities $f_1(x)$ for S_1 (situation 1 or signal + noise and $f_2(x)$ for S_2 (situation 2 noise), i.e. z-scores of hit (H) and false alarm (FA) rates, define a linear relationship with discriminability $d' = (\mu_1 - \mu_2)/\sigma$ as intersection with the z(H)-axis:

$$z(H) = \frac{\mu_1 - \mu_2}{\sigma} + z(FA) \qquad (B.10)$$

H and FA are taken as estimates of the indicated areas in the density functions of Fig. B.2, with respective integration limits or z-scores (inverse Φ^{-1} of the cumulative normalized probability density f(x)) defined by criterion c. If a sufficient number of data (hit rate H, false alarm rate FA) are given, e.g. between subjects with different confidence ratings, a linear regression may be performed in order to determine the distance between the means μ_1, μ_2 of the two densities f_1, f_2 as intersection with the z(H) axis.
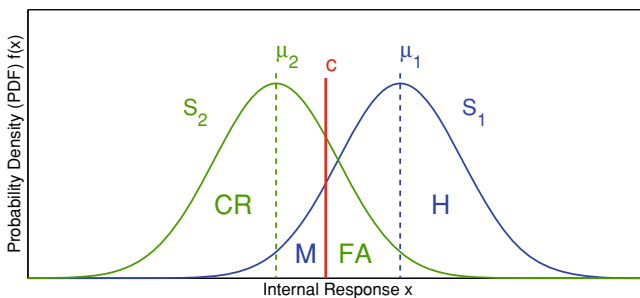
Fig. B.2 Gaussian density assumption of observers internal random response (or familiarity) x to noise (S_2) and noise + signal (S_1) stimulus

If the variances of the two Gaussian densities can not be assumed equal as precondition, e.g. $\sigma_{Signal} = \sigma_1 \neq \sigma_2 = \sigma_{Noise}$, Eq. (B.10) can be shown to change as follows (e.g. Metz et al. (1998)):

$$z(H) = \frac{\mu_1 - \mu_2}{\sigma_1} + \frac{\sigma_2}{\sigma_1} z(FA) \qquad (B.11)$$

From (B.10) it follows that with two equal variance Gaussian densities for the subjective (internal) response or familiarity with situations S_1, S_2 the discriminability d' is defined as difference between normalized mean values

$$d' := \frac{(\mu_1 - \mu_2)}{\sigma} = \Phi^{-1}(H) - \Phi^{-1}(FA) = z(H) - z(FA) \qquad (B.12)$$

measured in units of standard deviations between signal means. Φ is the Gaussian probability integral (cumulative density) of density f(x), (x = subjective response or familiarity with situations S_1 (signal + noise), S_2 (noise)).

Correspondingly the criterion value c is obtained as

$$c := 0.5(z(H) + z(FA)) \qquad (B.13)$$

In Fig. B.2c separates the M from H area and CR from FA area in Fig. B.2. Due to the independence of the two alternative events S_1, S_2 (with independently normalized densities $f(S_1), f(S_2)$) the results of the response matrix are unambiguously represented by the (FA, H)-data pair.

As a standard graph of SDT the so called receiver operating characteristic (ROC) unambiguously characterizes the observer in this experiment via his discriminability d' and decision criterion c. A single data point in (FA, H)-ROC space, typically as average over many runs and/or participants of an experiment (representing the average of e.g. 100 decisions) is unambiguously characterized by a pair of (isosensitivity d', isobias c)-parametrized ROC-curves. In this way the same conditional probabilities $p(R_1|S_1) = H$, $p(R_1|S_2) = FA$ that were used with the Bayes inference for calculation of the risk of world state contrary to expectation $(S_i \neq R_j)$ can be employed for deriving an unbiased discriminability value for the observer/decision maker. Examples of ROC-curves calculated with the above equations for concrete experimental data are presented in chapters "Remote Tower Experimental System with Augmented Vision Videopanorama" and "Multiple Remote Tower Simulation Environment (S. Schier)". Each point (H, FA) on a ROC-curve is unambiguously determined by the criterion c, separating the subjective yes/no, signal/noise, world-state 1/2 decision threshold. It follows that c is unambiguously characterized by the ROC-curve slope that decreases with more liberal decisions, i.e. gathering more hits H at the cost of allowing for more false alarms FA when c shifts to the right (decreases). Because the criterion corresponds to the integration boundary c of the two densities $f(S_1), f(S_2)$ in Fig. B.2, it can be expressed through the likelihood ratio (see Eq. (B.5) for the discrete case) via the probability densities

$$l(c) = \frac{-f(c|S_1)}{-f(c|S_2)} \tag{B.14}$$

that in fact equals the slope of the ROC curve at c. For details see Green and Swets (1988).

If sufficient data are available they may be used for deriving optimum d′ and c via data fitting. Quite often however, the amount of data is limited and a single average pair ($\langle$FA$\rangle$, $\langle$H$\rangle$) is used for deriving an unambiguous pair of d′- and c-parametrized ROC-curve crossing at this (FA, H)-coordinate. Figure B.3 depicts a series of d′-parametrized ROC-curves that shows how different discriminability values can be attributed to three series of measurements (red, green, blue (H, FA)-datapoints from framerate experiments described in chapter "Remote Tower Experimental System with Augmented VisionVideopanorama"), in this case by using the average of each set of four points.

In chapter "Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration During Simulated Aircraft Landing", Fig. 11 the unambiguously ($\langle$d′$\rangle$, $\langle$c$\rangle$) parametrized curve pairs are plotted, intersecting at the single average data point of each data set, represented by the crosses with error bars (standard errors). They correspond to three groups of subjects with three different
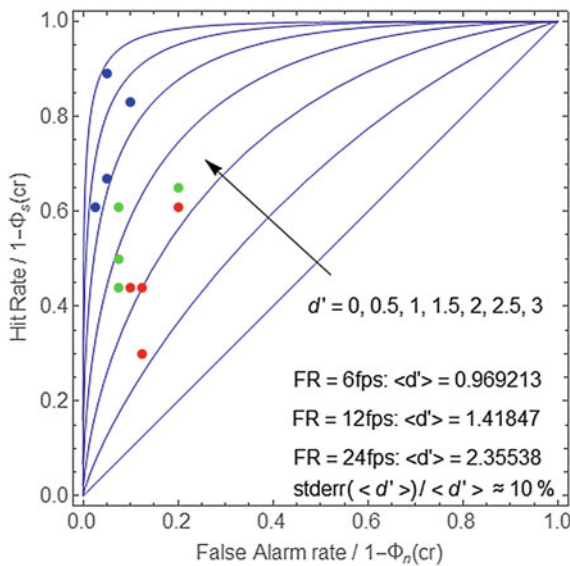
Fig. B.3 Series of d′ parametrized ROC curves with three sets of example datapoints (red: 6, green: 12, blue: 24 Hz) from chapter "Videopanorama Frame Rate Requirements Derived from Visual Discrimination of Deceleration During Simulated Aircraft Landing" (Framerate experiments). Unambiguous discriminability parameter $\langle$d′$\rangle$ for each set via average $\langle$H, FA$\rangle$ for each set. Axes titles indicate calculation of ROC curves via the cumulative probability densities of noise and signal (n = S_2, s = S_1), with criterion cr ($\equiv$ c) as integration boundary

experimental conditions (in that case different frame rates) used for generating three average pairs ($\langle FA \rangle$, $\langle H \rangle$). In this way within the experimental uncertainty three different pairs of isosensitivity/isobias curves are attributed to the measured average responses.

Quite often the limited set of measured data is not sufficient for verifying the Gaussian density precondition with regard to the familiarity or subjective response to the signal & noise vs. noise without signal. In this case a nonparametric variant may be advantageous for calculating discriminability and decision bias. Such a method based on the area under the ROC-curve is described in the next section.

B.2.2   Nonparametric Discriminability A and Subjective Bias b

This method is based on an estimate of the average area under ROC curves. For the estimate the possible areas for the sets of straight lines enclosing all proper ROC curves (with non-negative slope) for any specific (FA, H) point are determined as depicted in Fig. B.4. Proper ROC-curves must lie within areas A_1, A_2. Different formulas for average area A as discriminability index and a corresponding index b as nonparametric subjective bias were derived in the literature, but only recently a final correct version was published by Zhang and Mueller (2005).

The isosensitivity and isobias curves are calculated directly from the measured conditional probabilities H, FA, and are given by the Zhang and Mueller formulas as follows for the A isopleth:

Fig. B.4 Proper ROC curves must lie within areas A_1, A_2. Re-drawn after Zhang and Mueller (2005)
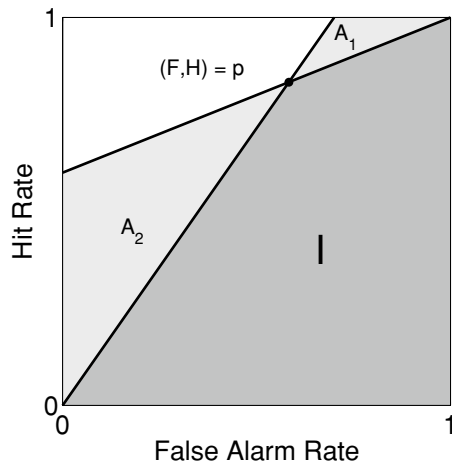
$$A = \begin{cases} \frac{3}{4} + \frac{H-FA}{4} - FA(1-H) & if \ FA \leq 0.5 \leq H \\ \frac{3}{4} + \frac{H-FA}{4} - \frac{FA}{4H} & if \ FA \leq H \leq 0.5 \\ \frac{3}{4} + \frac{H-FA}{4} - \frac{1-H}{4(1-FA)} & if \ 0.5 \leq FA \leq H \end{cases} \qquad (B.15)$$

and for the associated measure of decision bias which is based on the slope of the constant discriminability A-isopleths, and which corresponds to the likelihood ratio [13]:

$$b = \begin{cases} \frac{5-4H}{1+4FA} & if \ FA \leq 0.5 \leq H \\ \frac{H^2+H}{H^2+FA} & if \ FA \leq H < 0.5 \\ \frac{(1-FA)^2+(1-H)}{(1-FA)^2+(1-FA)} & if \ 0.5 < FA \leq H \end{cases} \qquad (B.16)$$

A further advantage of the discriminability index A is its limited range of values $0.5 \leq A \leq 1$ as compared to the parametric index with $0.5 \leq d' \leq \infty$. Figure 14 (chapter "Remote Tower Experimental System with Augmented VisionVideopanorama") and Fig. 8 (chapter "Multiple Remote Tower Simulation Environment (S.Schier)") illustrate the application of this method with the example of increase of discriminability of moving objects on a videopanorama with video framerate. The position of the group average (H, FA)-results (large crosses) on the A isopleth determines the corresponding decision bias b which in this case indicates conservative decision making, i.e. avoiding false alarms.

References

Robert, C. P. (2001). *The Bayesian choice.* Springer.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics.* Peninsula Publishing: Reprint Edition.

MacMillan, N., & Creelman, C. D. (2005). *Detection theories.* Psychology Press, Taylor and Francis.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrica, 70*(1), 203–212.

Metz, C. E., Benjamin, A. H., & Shen, J.-H. (1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine, 17*, 1033–1053.

https://www.norsys.com/ (2014).

Appendix C
Mental Workload and Measures for (Quasi) Real-Time Applications

Norbert Fürstenau & Anne Papenfuss, DLR

In this Appendix we present a brief overview on important aspects of mental workload (MWL) and task load (TL). It may be used as reference for workload related discussions in Part III of this volume. Our main interest is in online and real time applications which are suitable for human-in-the-loop simulation experiments. The definition of MWL (short WL) and TL is followed by subjective (quasi real-time) measures and a selection of objective methods to assess the mental load experienced during execution of tasks that serve for achieving a goal under well defined work conditions. Part of this Appendix (Sects. C.1, C.3, C.4, C.5, C.6.3) is based on Sect. 2 in Fürstenau and Radüntz (2021).

C.1   Mental Workload

Quantification of mental workload constitutes one of the main issues in cognitive ergonomics and human-factors research. Like many concepts in psychology, there is no singular agreed-upon definition or method for measuring mental workload. Much more, it is assumed that successful performance on a task or test requires cognitive resources, which can be seen as mental workload. In other words, mental workload is a theoretical construct referred to as "the cost incurred by the human operator to achieve a particular level of performance (Hart & Staveland, 1988). Similar definitions were given by Kahnemann (1973), Wickens and Hollands (2000), and Xie and Salvendy (2000). Nevertheless, its quantification contributes to the evaluation of human–machine systems, estimation of the appropriateness of automation levels, and enhancement of interface design. A good overview on different theoretical and practical aspects of workload with focus on transportation as our major field of interest is given in (Hancock & Desmond, 2001).

As mentioned above, for measuring mental workload there are several methods available that can be categorized in two groups: the objective and subjective methods.
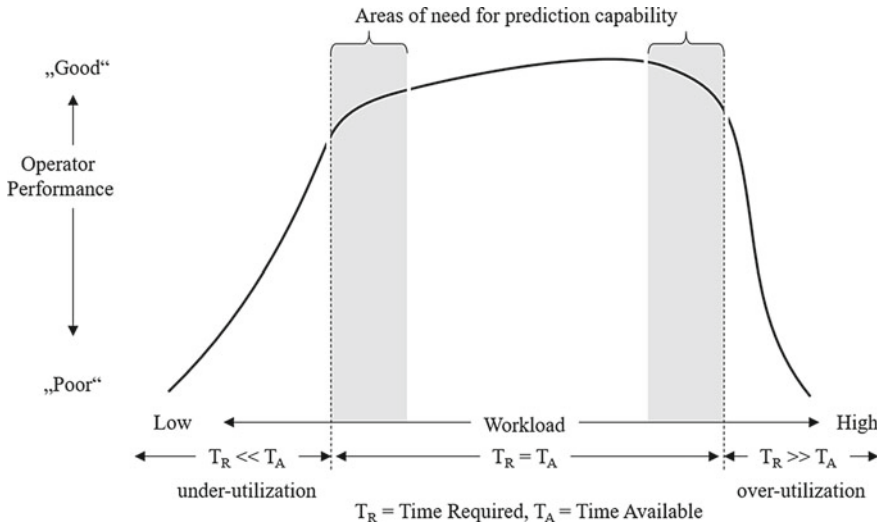
Fig. C.1 Hypothetical relationship between workload and operator performance, after (Johannsen et al., 1979, p. 4)

Objective methods rely upon quantification of performance (e.g. decision errors) or bio-physiological data (e.g. heart rate HR and HR variation) while the subjective methods consider the subjective rating given by the performer. Although all measurement methods aim to describe the relation between task demands and subject's ability to cope with them, several investigations reported dissociations among methods' results. A possible explanation might be that mental workload is a multidimensional concept that cannot be captured in all its facets by a single method. Apart from the task requirements, mental workload variations are caused by individual characteristics such as habituation, actual precondition, and coping styles (ISO-10075, 1991, 1996, 2004).

C.2 Importance of Workload for Design of Operational Concepts and Work Environment

Workload is regarded as one contributing factor to human and overall system performance. It is assumed that workload and performance are related like an inverse u-shaped function, Yerkes-Dodson law (Yerkes & Dodson, 1908) and (Johannsen, et al., 1979). A schematic visualization is shown in Fig. C.1, with workload on the abscissa and operator performance on the ordinate.

It is important to understand this figure as an illustration with the goal to set several relevant factors into context and not as an exact model. Exact proven models explaining the relationship between these factors are helpful but also difficult to

derive as data derived by observing humans is noisy and rarely influenced by a single factor. Furthermore, it needs to be emphasized that there is a broad range of "good" performance where time required to fulfil a task (T_R) ≤ time available (T_A). One might argue that the range $T_R \approx T_A$ should be treated as a threshold, and in fact empirical research shows that operators adapt their working strategies (e.g. (Sperandio, 1978)) and accordingly T_R, in order to avoid high time pressure $T_P = T_R/T_A$ (see Sect. 6.1).

Nevertheless, there are areas where these strategy shifts cannot mitigate any longer the mismatch of T_A and T_R. These areas are labelled as under- and over-utilization. Furthermore, there are transition areas (grey areas in Fig. C.1) where operator's performance starts to decrease but still meets required performance levels. So, both under-utilization and over-utilization of the cognitive resources can explain degraded overall system performance (Johannsen, 1979).

When developing new operational concepts and work environments, the workload experienced by the human operator thus is an important indicator to decide for a specific design alternative (Wickens, 2008). First, the negative effects of over- and under-utilization should be avoided. As an example, for the negative effects of under-utilization the so-called "dead man's switch" is implemented in work environments, where little manual work from the human operator is required or shifts are long or the task is dull. In these phases of under-utilization, humans tend to lose vigilance, get tired and thus are more likely to oversee important events or fall asleep. In order to detect or prevent these states, the operator is supposed to regularly press a button to signal to the system, that s/he is awake. On the contrast, in phases of over-utilization, operators cannot perform as expected when (a) they are simply not able to conduct all tasks as and when required because of physical limitations and (b) the pressure of the situation lead to a narrowing of attention, getting too much focused on one specific task and neglecting important other tasks. Investigations of cockpit accidents revealed that in high workload phases pilots are likely to neglect auditory signals from the system and from team members (Silberstein & Dietrich, 2003).

Nevertheless, operators are not unidirectionally exposed to workload but can apply strategies to first, mitigate effects of over- and under-utilization and second adapt their workload (Lee et al., 2007). In air-traffic control, ATCOs handle traffic adapted to the traffic situation and considering their own (Moehlenbrink et al., 2011). In an MRT simulation study, ATCOs reported that in case of higher workload they avoid to build traffic situations which require additional monitoring from their side. For instance, they avoided conditional clearances which might increase efficiency of the air traffic but also need to be visually double-checked. So, based on previous experience, human operators can adapt to a certain degree to their experienced workload. Nevertheless, any evaluation of a new operational concept and work environment should consider the impact of the experienced workload and avoid constant under- and over-utilization.

Phases of low workload and underutilization can lead to loss of vigilance and boredom. On the long term, motivation and job satisfaction can suffer. All these effects can lead to degradation of job performance. Phases of excessive workload can, when they occur as peaks, lead to sensory overload and errors and insufficient

performance. These errors can be related to the safety of operations. When operators apply control strategies to mitigate excessive workload, the impact will rather be on the efficiency of conducting the given tasks. For instance, in Air Traffic Control (ATC), flights will be delayed in order to keep the level of safety high. In this environment, it is preferred that human operators will violate their efficiency goal but must not violate the safety of operations. Constant phases of overutilization can lead to stress and unhealthy working conditions.

C.3   Subjective Quasi Real-Time WL Measures

Several researchers suggested that the subjectively experienced workload is of particular importance when evaluating subject's state Yeh and Wickens (1984), Sheridan (1980) and Johannsen et al. (1979) stated that "if an operator feels effortful and loaded, he is effortful and loaded". Moreover, questionnaires for registering mental workload possess a high face validity, are cheap, easy to conduct, and non-invasive. The most accepted subjective measure in ATC appears to be the multidimensional NASA task-load-measure (TLX) based on questionnaires for capturing the different aspects constituting the experienced W (Hart & Staveland, 1988). TLX data together with expert ratings and Instantaneous Self Assessment (ISA) self reports (see below) were evaluated in (Radüntz et al., 2019) to study the WL effect of a non-nominal event within HitL approach sector ATC simulations. The main advantages of subjective methods are the relatively low data acquisition effort and the high user acceptance. Their main drawback is that they suffer from subjective distortion. They are influenced by memory lapses as the experienced workload took place at some time in the past (NASA-TLX) and they are subject to social desirability bias as the individuals think that they are expected to provide a certain kind of answer (Lehrer et al., 2010; Radüntz, 2017). The questionnaire's items may not be readily understood or participants may lack the ability to introspect. What is more, they do not allow for fine-grained temporal sampling on the time scale of seconds and can alter the current workload state, e.g., if during a monotonous task the participant becomes activated by answering questions (Radüntz, 2017).

In the present work our interest was focused on WL measures appropriate for (quasi) real time data analysis of realistic simulator experiments. Two established subjective self-report measures suitable for near real-time application are SWAT (Subjective Workload Assessment Technique) and the above mentioned one-dimensional ISA method. SWAT measures the three load dimensions, time, effort, and stress, each with three levels (Reid et al., 1989), ISA monitors online in fixed time intervals of a couple of minutes via self reports the experienced WL on a five level scale. It minimizes possible additional WL (due to the reporting) by not discriminating load dimensions in contrast to SWAT (Brennan, 1992; Jordan, 1992; Kirwan, et al., 1997; Tattersall & Foord, 1996). The latter authors reported significant correlations of ISA ratings with Heart Rate Variation (HRV, see below) and task performance although the primary task performance on a tracking task turned

Table C.1 ISA workload categories after (Kirwan et al., 1997)

Level	WL heading	Spare mental resources	Description
5	Excessive	None	Behind on task; loosing track of the full picture
4	High	Very little	None essential tasks suffering. Could not work at this level very long
3	Comfortable busy pace	Some	All tasks well at hand. Busy but stimulating pace. Could keep going continuously at this level
2	Relaxed	Ample	More than enough time for all tasks. Active on tasks less than 50% of the time available
1	Underutilized	Very much	Nothing to do. Rather boring

out poorer during periods when ISA responses were required. Of course this distortion certainly depends on the details of task and reporting method (verbal, keypad, touchscreen). The characterization of the reported subjective load levels is listed in the following Table C.1.

Because the scale levels represent the subjective decision of participants on the experienced load during task execution the level differences are probably not equidistant. In the theoretical model of chapter "Model Based Analysis of Subjective Mental Workload during Multiple Remote Tower Human-in-the-Loop Simulations" (see also Appendix D) we assume an equidistant ISA scale so that any deviation from linearity is included in the nonlinearities of the model equations.

An early subjective quasi real time WL assessment technique was introduced by (Stein, 1985): the Air Traffic Workload Input Technique (ATWIT) using the 7-level WL Assessment Keypad (WAK). Lee et al. in (2005) reported on analysis of ATC simulation ATWIT-WL data with nonlinear (sigmoid) dependency of WL on traffic count (see Sect. C.6 below).

C.4   Psycho-Physiological Measures Heart Rate and HR-Variation

The analysis of bio signals as objective measures offer the possibility to continuously determine mental workload. They do not interfere with participant's current workload state as they can be obtained on-the-fly during task execution. Their main issue is that user acceptance may be impaired because of the complexity of the registration system. However, recent developments in mobile sensor technology promise small, lightweight, and wireless systems (Radüntz, Dual Frequency Head Maps: A New Method for Indexing Mental Workload Continuously during Execution of

Cognitive Tasks, 2017). Bio-physiological data include, among others, cardiovascular biomarkers which are easy to assess and were frequently used to analyse cardiovascular activity under a wide range of experimental conditions (e.g. Lehrer et al. (2010)). The heart rate (HR) and the heart rate variability (HRV) are the most prominent biomarkers.

In most cases HRV is characterized in the frequency domain by means of various spectral features. According to the definitions by (Mulder et al., 2004), the frequency range can be categorized in three bands: the low-frequency (LF: 0.02–0.06 Hz), mid-frequency (MF: 0.07–0.14 Hz), and high-frequency (HF: 0.15–0.4 Hz) bands. In 1981, (Mulder & Mulder, 1981) found that spectral power of the HRV between 0.02 and 0.20 Hz was in association with non-linear processes of body temperature and blood pressure control while spectral power between 0.20 and 0.40 Hz was related to the respiratory activity (parasympathetic control mechanisms). Under mental load the total spectral power decreased, whereby the spectral power between 0.02 and 0.20 Hz was particularly affected and contributed about 80% to the total spectral energy.

Basic research on HRV as WL measure for adaptive automation was investigated by (Prinzel et al., 2003) with a tracking task, together with EEG (see below) and event related potentials. Lehrer et al. (2010) reported an increase of association between self-report (using NASA task load index (TLX), scale given immediately after each 5 min task) and both expert ratings of task load and task performance in a flight simulator by means of cardiac data. We recently reported on analysis of HR and HRV measures within a simulator experiment where we aimed at clarifying their inherent timescales (Radüntz et al., 2020).

C.5   Neurophysiological (EEG-Based) Measures

The spectral power of EEG oscillations in different frequency bands (specifically α (4–7 Hz), β (8–13 Hz), Θ (14–30 Hz) may be linked to different levels of workload by means of analysis of variance (ANOVA) (e.g. Lei & Roetting (2011) and Aricó et al. (2018)). The potential of an EEG based task engagement-index (based on the power ratio $\beta/(\alpha + \Theta)$ recorded from 4 scalp sites, 40 s moving average, 2 s clock rate) within the context of adaptive automation was demonstrated by Prinzel et al. by means of a laboratory type multi-attribute cockpit-instrument tracking-task simulator experiment, using ANOVA for quantifying the significance of the engagement level (Prinzel et al., 2003). The important artifact rejection was based on a pre-set threshold voltage which for real-world applications of course would not be sufficient. Meanwhile classifiers are increasingly used for the separation of workload levels. In previous publications we have described the development and validation of the new DFHM WL-index using a support vector machine classifier (based on frontal α-band and parietal Θ-band powers), performed under laboratory conditions with standard task load batteries. Once calibrated for discriminating low, medium, and high WL-levels, it was shown to require no retraining of the machine learning

algorithm, neither for new subjects nor for new tasks (Radüntz, Dual Frequency Head Maps: A New Method for Indexing Mental Workload Continuously during Execution of Cognitive Tasks, 2017). For this experiment a commercial 25 active-electrode system (g.tec.Ladybird) with 500 Hz sample rate and 0.5–50 Hz bandpass was used. The corresponding data from the real world simulation experiment with different subjective and objective reference measures that were used also for the present model based data analysis showed the objective DFHM index to provide significant correlation with controller's subjectively experienced self rating ISA-WL measure under traffic load variation (Radüntz et al., 2020). For testing the DFHM-index WL sensitivity, the participants in this analysis were separated into two groups (low and high WL sensitivity) according to their individual linearized WL-sensitivity parameters that were formally derived from the logistic ISA characteristic of the subjective self report measures (Fürstenau et al., 2020).

C.6   Formal Models

In this section we provide a brief description of three formal models used for quantifying objective task load and subjective workload measures for (quasi) real time applications.

C.6.1   Time Pressure and Information Processing TP/IP for Objective Workload

C.6.1.1   Sector Capacity for Networks (CAPAN)

Eurocontrol's network management, responsible for the configuration of the upper airspace sectors, applies the CAPAN methodology for sector capacity assessment, based on fast time simulation. Here, they are calculating a theoretical sector capacity based on predicted working times of ATCos based on a fast time simulation. Basically, these working times are interpreted as workload levels. In their model, five workload levels are differentiated. The idea is that in each hour the predicted working time should not exceed 42 min (T_R = required time). This corresponds to about 70% of the time available (T_A). This threshold considers that ATCOs should have 30% of their time for tasks not bound to a discrete event (a flight). A predicted level of working time above 42 min is considered as an overload situation, which has to be avoided. In the methodology, fast time simulations of the air traffic within the sector is used to predict ATCOs tasks and consequently summing up the time required to fulfil all basic tasks. In the CAPAN model, workload is the sum of the time durations needed for a basic set of ATCO tasks. So, workload models are based on models of the ATC

Table C.2 Overview on ATCO basic tasks and typical percental shares in terms of time spent on tasks (Russo, 2016)

| ATCO task	Average percental share	sd	Average T_R [min]/h
Flight data management	7.9	1.2	4.7
Conflict Search	23.4	2.6	14.1
Coordination	4.1	1.6	2.5
Radio Communication (RC)	**40.0**	**4.3**	**24.0**
Radar	24.3	8.5	14.6

task. Table C.2 summarizes the basic ATCo tasks and percental share which were observed in seven different ATC sectors.

C.6.1.2 Information Processing/Time Pressure (IP/TP) Theory

Time pressure is defined as ratio between time required for a decision TR (selection of one from several alternatives for solution of a task) and available time TA: Time Required/Time Available (TP = TR/TA). Hendy et al. formalized TP within the framework of the information processing hypothesis with exponentially growing error rate with growing TP (IP/TP hypothesis). Validation experiments of the TP model were conducted with a simulated en-route controller work place (Hendy et al., 1997). For explaining experimental results on a formal basis these authors developed a feedback model that describes adaptation of work strategy for reducing task load in order to keep TR below TA. Figure C.2 depicts a modified schematic of their model.

The available airspace for the air traffic (en-route sectors, approach sector with well defined approach routes, and airport with control zone and TMA) provides the boundary conditions for the relationship between traffic and ATCO's workload generated through the controlled development of traffic with anticipated situations. Knowledge of these traffic patterns enables the ATCo to use strategies of information processing which simplify the traffic management. Limitations of possibilities for control result e.g. from activation of "special use" air space or AC in holdings/trombones. Limitation of airspace increases the probability of violating separation rules. HITL validation experiments of the TP/IP model were performed with a simulated radar control workplace.

Within a RTO validation experiment the measured observation error rates during flight maneuvers in a control zone were analyzed by means of a modified TP model (see chapter "Model Based Analysis of Two-Alternative Decision Errors in aVideopanorama-Based Remote Tower Work Position" in Part II of this volume and Fürstenau (2016)).
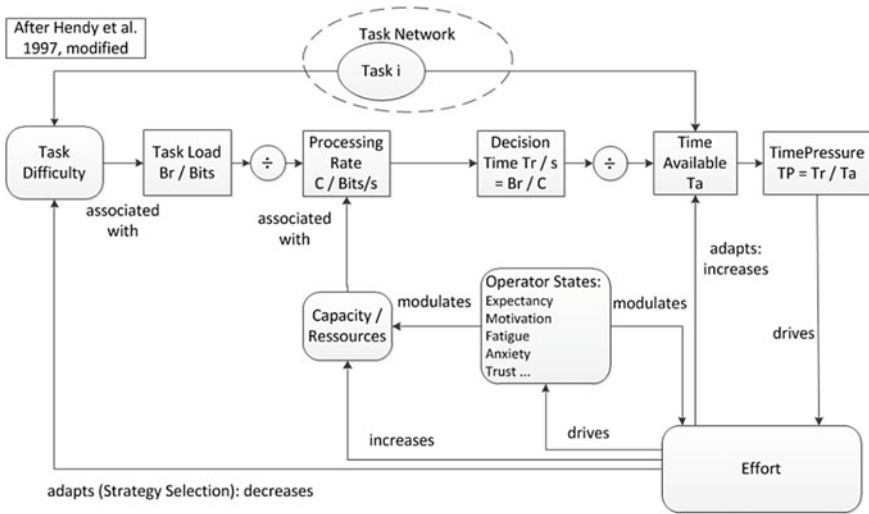
Fig. C.2 TP/IP feedback model for time pressure stabilization through adaptive change of work strategy (after Hendy et al. (1997), modified)

$$p = 0.5\left(1 + exp\left\{-\left(\frac{TP - \mu}{\beta}\right)\right\}\right)^{-1} \tag{C.1}$$

Error probabilities of the two alternative decision experiment (Flight maneuver yes/no) were modeled by means of a logistic TP function. The fit to the data obtained under RTO videopanorama and real tower observation conditions provided reasonable values for TP threshold (TP: $= \mu \approx 1$) and error sensitivity parameter β. At TP $= 1$ the major part of the available information required for a correct decision is no longer processed. $1/\beta =$ error sensitivity ~ slope at transition (inversion) point μ. For TP > 1 the error probability p asymptotically approaches p $= 0.5$, i.e. random decision. The results were consistent with independent data analysis through signal detection theory (see Ch. 10 in Part II).

C.6.2 Sigmoid Model of Subjective WL During En-route Sector Control

Lee (2005) and Lee et al. (2005) suggested that there is no linear relationship between controller workload and traffic count. They compared linear, exponential and sigmoid relationships between empirically observed workload ratings and traffic counts in an experimental simulator study. In their study, four controllers with very high familiarity with their sectors took part in a high-fidelity human-in-the-loop simulation. Subjective workload was collected every 5 min with the Air Traffic Workload Input

Technique (ATWIT, see above (Stein, 1985)) using the 7-level scale of the WL Assessment Keypad (WAK), This discrete WL scale ranges from 1 (low workload) to 7 (extreme subjective workload). They reported results of logistic WL-data fits based on a heuristic sigmoid WL function as dependent on aircraft count n within enroute sectors, with 4-parameter estimates (a, b, c, d):

$$W = \frac{a}{1 + exp\{-bn + c\}} + d \tag{C.2}$$

In this study, the sigmoid-curve fitted the observed empirical workload ratings best, compared to linear and exponential fits. The authors concluded that the subjective workload is categorial, where rapid transitions from low to high subjective workload can occur. The authors emphasize that also from an operational point of view the sudden transition from high to unmanageable workload is of importance.

It has to be noted that the parameter table in Lee (2005) includes a wrong sign for the estimate of c: it has to be "+" (instead of "−" in the table) in order to reproduce their curve (dashed line in Fig. C.3). Moreover the authors provide no confidence intervals or error margins. Parameter estimates are supposed to have large uncertainties due to the low number of data points available for estimating the four parameters (Pearson correlation $R^2 = 0.84$ with N = 23). In Fig. C.3 we compare the heuristic four parameter fit result (dashed curve) with a logistic model based 3-parameter fit using
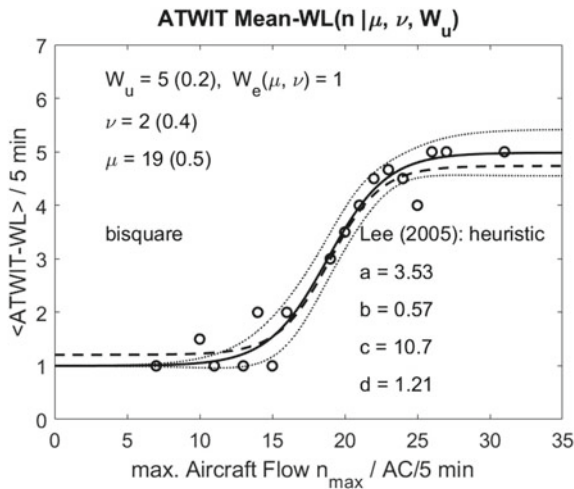
Fig. C.3 Sector control ATWIT-WL measurements of Lee (2005) and Lee et al. (2005) with sigmoid fit depicted by dashed curve: logistic Eq. (C.2) used for 4 parameter fit with (a, b, c, d) = (3.59, 0.57, 10.7, 1.21); specific sector monitor alert value MAP = $18 \approx c/b = 18.8$. Solid sigmoid = theoretical logistic 3-parameter fit with Eq. (D.6) of Appendix D (shift μ = inversion point, scaling ν = sensitivity parameter, W_u = estimate of upper asymptote), using $\Delta := 1$ as prior information for specifying lower asymptote at scale limit $W_d = 1$ (dotted curves for 95%-CI; robust fit procedure with bi-square residual weighting for reducing outlier effects)

Eq. (D.6) in Appendix D. The latter represents a generalisation of a model used in chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-in-the-Loop Simulations" (Eq. (D.8) in Appendix D and Fürstenau and Radüntz (2021)). It was derived from a cognitive resource limitation hypothesis and allowed for predicting parameter Δ (= d in Eq. (C.2)) through use of prior information. Specifically, because the measurements indicate an extended range of lowest WL reports (W = W_d = 1) for low traffic load it appears reasonable to set offset d = Δ: = 1 so that W_d = 1 ≈ lower asymptote = W(n → −∞). The remaining three parameters are: μ = c/b = shift or inversion point (representing an estimate for the sector monitor alert value MAP, i.e. transition to overload), ν = sensitivity index = inverse rate parameter b, and W_u = W(n → ∞) = effective upper asymptote for extreme environmental load. It can be seen that the heuristic 4-parameter fit is enclosed in the confidence interval CI (dotted curves) of the 3-parameter model, with the exception of the low traffic range (<ca. 10 AC/5 min). Here CI converges to 0 due to vanishing difference between logistic intercept W(n = 0): = W_e ≈ 1 and W_d. In Appendix D we prove that also theoretically for Δ = 1 and $\mu/\nu \gg 1$ lim(W_e – W_d) → 0. As expected, comparison of the corresponding 4- and 3-parameter fit estimates shows agreement within the given uncertainties (stderr) if we use the correspondences: a + d = 4.74 ≈ W_u = 5 (0.2), 1/b = 1.8 ≈ ν = 2 (0.4), c/b = 18.8 ≈ μ = 19 (0.5). The cognitive resource limitation hypothesis would predict for subjective WL self-assessment convergence of the sigmoid to the nominal upper scale limit, i.e. W_u = 7. However, a corresponding test with a 2-parameter fit using W_u = 7 as additional prior information exhibits significantly increased std errors and CI of parameter estimates. Although this discrepancy certainly requires further model based analysis of additional experimental data for improved statistics there is no reasonable doubt on the nonlinear logistic dependency of subjective workload reports as dependent on the environmental traffic load, when taking into account also recent results from two independent ATC HitL simulation experiments reported in chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations" of this volume and in Fürstenau and Radüntz (2020, 2021).

C.6.3 Psychophysics of Subjective Mental Workload

Despite the fact that subjective WL-measures are widely accepted and used, there have been very few studies examining their methodological viewpoint. Based on laboratory experiments with standardised cognitive tasks Gopher and Braune (1984, 1985) proposed a scaling approach that can be traced back to the psychophysical measurement theory of Stevens (1975).

Psychophysical research aims to describe the relationship between changes in the amplitude of a physical stimulus (e.g. brightness, loudness) and the subjective perception of these variations. First relations were found experimentally by Weber in 1864. Thereby, the so called just-noticeable difference JND was defined that

describes the smallest change ΔS that can be perceived between two stimuli. Weber noticed that the greater the initial stimulus S, the larger the difference ΔS needed to distinguish between the two stimuli, and that the relation is a constant k (Weber's law): $k \sim \Delta S/S$.

Based upon Weber's law, Fechner suggested that for any sense modality there is a constant just-noticeable difference in subjective perception related to the constant k (Masin et al., 2009). Furthermore, Fechner indicated that the relation between stimulus S and perception P is logarithmic comprising an experimentally determined constant c and a stimulus threshold S_t. The latter denotes the intensity of the stimulus at a state with no perception (Buntain, 2012). This relationship is known as the Weber-Fechner law $P = c\,ln(S/S_t)$. An improvement of the Weber-Fechner law was introduced by Stevens (1957). In Stevens' power law the sensation magnitude is a power function of stimulus intensity and the corresponding generalized linear curve (double logarithmic scale) is described by the constant b and Steven's exponent γ (slope or sensitivity) that is characteristic for the type of stimulus.

$$ln(P) = ln(b) + \gamma\,ln(S) \tag{C.3}$$

It is valid also for the stimulus–response transfer between sensor input (stimulus amplitude) and sensor neurons firing rate (action potential) (e.g. Birbaumer & Schmidt (2010)). Both parameters were determined for a large number of different modalities (e.g. brightness, loudness, apparent length) in order to adjust the curve to the different psycho physical functions. The magnitude of the sensitivity parameter γ typically is of the order of 1. Steven's power law may be derived from an information theoretic approach with P ~ perceived sensory (Shannon) information. Within this context it represents an approximation for lower amplitude stimuli with prolonged sampling time as compared to the Fechner law representing the large amplitude brief stimulus duration approximation (Norwich, 1987). In fact Gopher et al. argued that … "if the human information processing system can be assumed to invest processing facilities to enable the performance of tasks then subjective measures can be thought to represent the perceived magnitude of this investment, in much the same way that the perception of …" a physical stimulus is changed with variation of its magnitude".

Gopher et al. based their formal power law relationship on the measured average values across the sample of 55 participants of perceived load for each of 21 single and dual task conditions of a task load battery, with tasks guided by Wickens' multiple resources paradigm (Wickens, 2002). Scores of individual subjects for the different task conditions were defined with respect to a single dimension tracking task as reference and transformed such that $0 <$ load score ≤ 1. In contrast to a regular psychophysical experiment, in their WL-experiment there existed no physical quantity (e.g. brightness or sound pressure) that generated the experience and its perceived value, and that would allow to derive the two Stevens parameters (γ, b) of the power law function (3) through regression analysis. As a loophole the Sternberg auditory-verbal memory search tasks with set size 2 and 4 were assumed to be one unit apart on the hypothetical resource investment (power law) function and were given the actual values of S = 1 and 2. These values were based on the information theoretic analysis that would assign to these tasks 1 and 2 bits of information respectively (e.g.

(Norwich, 1987)). Together with the corresponding average scores the two resulting equations allowed to quantify (γ, b), with exponent γ of the order of 1 in agreement with classical psychophysics values. The 21 measured score values in turn allowed to define the units of the associated processing resources axis for generating power law graphs of score means for further analysis of repeated test runs. In this way it was possible, e.g. to successfully predict dual-task perceived loads from the derived scores of the single tasks (using Eq. (C.3)) by applying a simple additive processing-resource rule.

Recently Bachelder and Godfroy-Cooper (2019) reported on the application of the psychophysics power law to the analysis of a pilot workload estimation experiment. For their theoretical model they referred to the information theoretic approach (Norwich, 1987). The authors used a flight compensatory tracking task with Bedford hierarchical unidimensional WL scale (a modified Cooper-Harper rating scale) designed to identify operators spare mental capacity while completing the task. The physical stimulus S determining WL in Eq. (C.3) was derived from the measured standard deviations of control error rates based on standard manual control theory. Theoretically predicted Stevens exponents of different tasks were in the range $0.24 \leq \gamma \leq 0.41$ and compared favourably with those obtained from regressions of the data using the power law (3): $0.21 \leq \gamma \leq 0.37$, i.e. order of magnitude was comparable with those of the classical psychophysics experiments.

Of interest appears also the work of Link (1992) who presented a stochastic (brain) wave discrimination theory allowing for formal derivation of psychophysical laws, in particular the Stevens power law. Starting point was the probability for reaching a decision threshold through random sampling of the difference between stimulus and referent waves that defined a logistic response function with exponential dependence on wave amplitude difference and threshold. The Stevens power law was derived from sensation matching by combining the two corresponding logistic functions. The ratio of two normalized subjective response thresholds A_S/A_P relate two simultaneously measured sensations with logistic response probability functions (Link 1992). The product of this ratio with the log(normalized sensation of physical stimulus S/S_0) equals the log(normalized subjective response P/P_0) in the generalized linear form of Stevens law (Eq. (3)). This result of Link's stochastic wave discrimination theory explains classical stimulus–response results of Stevens (e.g. (Stevens, 1975). In these experiments the subjective response of subjects to different physical criterion stimuli S (e.g. loudness, brightness, pressure) was simultaneously matched, e.g. by means of squeezing a dynamometer as subjective measure (P) of the stimulus. Based on the cognitive resource limitation hypothesis as our starting point in chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations" and Appendix D we use a comparable formal procedure for the derivation of the ISA(RC) power law (with RC = frequency (rate) of radio calls between ATCo and pilot), however in the present approach through combination of the deterministic logistic work load ISA(n) (subjective response) and task load RC(n) (= physical communication stimulus) characteristics.

In Appendix D it is shown that a psychophysics power law characteristic can be derived for any pair of subjective and objective measures with a logistic dependence on some kind of external objective load variable, comparable to our ISA(n), RC(n) measures (see chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations"). A limitation of the Gopher et al. studies may be seen in the fact that it was conducted under highly controlled experimental laboratory conditions with basic cognitive tasks that have no evident relation to an (independent) physical stimulus variable. To our best knowledge, only very few experiments like Bachelder and Godfroy-Cooper (2019) and Lee (2005) use a more realistic human-in-the-loop simulation (HitLS) with online measures for a test of the psychophysics hypothesis of mental workload. One advantage of a HitL simulation experiment is the possibility of parametric interdependence, such as ISA-WL(n) and RC-TL(n) measures on the common external traffic load variable n (traffic flow as controllable simulation scenario parameter, (see chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations")) that in our case mediates the functional power law relationship. After suitable normalization and nonlinear transformation $(P(ISA), S(RC))$ of ISA and RC variables the $y_P(y_S)$ power law relationship in log–log coordinates corresponds to the generalized linear form of Stevens law (Eq. (3)) with subjective perception $y_p \sim \ln(P)$ and physical stimulus $y_s \sim \ln(S)$ (see Appendix D and chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations").

References

Aricó, P., Borghini, G., Flumeri, G. D., Sciaraffa, N., & Babiloni, F. (2018). Passive BCI beyond the lab: Current trends and future directions. *Physiological Measurement, 39*, 08TR02. https://doi.org/10.1088/1361-6579/aad57e

Bachelder, E., & Godfroy-Cooper, M. (2019). Pilot workload esimation: Synthesis of spectral requirements analysis and Weber's law. 322514/620191228

Birbaumer, N., & Schmidt, R. F. (2010). *Biologische Psychologie* (7th ed.). Springer Medizin Verlag.

Brennan, S. (1992). An experimental report on rating-scale descriptor set for the instanteneous self assessment (ISA) recorder. Tech. rep., DRA Maritime Command and Control Devision, Portsmouth.

Buntain, C. (2012). Psychophysics and Just Noticeable Difference. (D. o. University of Maryland, Ed.) *Report 4, CMSC828D*.

Fürstenau, N. (2016). Model-based analysis of two-alternative decision errors in a videopanorama-based remote tower work position. In N. Fürstenau (Hrsg.), *Virtual and Remote Control Tower* (S. 241–260). Springer.

Fürstenau, N. (2016). *Virtual and remote control tower.* In N. Fürstenau (Hrsg.) Springer.

Fürstenau, N., & Ellis, S. R. (2016). Videpanorama frame rate requirements derived from visual discrimination during simulated aircraft landing. In N. Fürstenau (Hrsg.), *Virtual and Remote Control Tower* (S. 114–137). Switzerland: Springer International Publishing.

Fürstenau, N., & Radüntz, T. (2021). Power law model for subjective mental workload and validation through air-traffic control human-in-the-loop simulation. *Cognition, Technology, and Work*. Retrieved May 30, 2021, from https://doi.org/10.1007/s101011-021-00681-0

Fürstenau, N., Radüntz, T., & Mühlhausen, T. (2020). Model based development of a mental workload sensitivity index for subject clustering. *Theoretical Issues in Ergonomics Science.* https://doi.org/10.1080/1463922X.2020.1711990

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures. *Human Factors, 26*(5), 519–532.

Gopher, D., Chillag, N., & Arzi, N. (1985). The psychophysics of workload—A second look at the relationship between subjective measures and performance. In *Proceedings of the Human Factors Society* (29th Annual Meeting) (pp. 640–644).

Hancock, P., & Desmond, P. A. (Eds.). (2001). *Stress, workload, and fatigue.* Lawrence Erlbaum Publishers.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Result of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183).

Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information processing load. *Human Factors, 39*, 30–47.

ISO-10075. (1991, 1996, 2004). Ergonomic principles related to mental workload—general terms and definitions.

Johannsen, G., Morey, N., Pew, R., Rasmussen, J., Sanders, A., & Wickens, C. (1979). Final report of experimental psychology group. In N. Morey (Ed.), *Mental workload. Its theory and measurement* (NATO Conference Series ed., Vol. 8, pp. 101–114). Springer.

Jordan, C. (1992). Experimental study of the effect of an instantaneous self assessment workload recorder. Tech. rep., DRA Maritime Command and Control Division, Portsmouth.

Kahnemann, D. (1973). *Attention and effort.* Prentice Hall.

Kirwan, B., Evans, A., Donohoe, L., Kilner, A., Lamoureux, Atkinson, T., & MacKendrick, H. (1997). Human factors in the ATM system design life cycle. In *FAA/Eurocontrol ATM R&D Seminar.* Paris, France: Eurocontrol.

Lee, P. U. (2005). A non-linear relationship between controller workload and traffic count. In *Proc. Human Factors and Ergonomics Society, 49th meeting 2005*, *49*, pp. 1129–1133. https://doi.org/10.1177/154193120504901206

Lee, P. U., Mercer, J., & Callantine, T. (2007). Examining the moderating effect of workload on controller task distribution. In *Proceedings HCII 2007*.

Lee, P. U., Mercer, J., Smith, N., & Palmer, E. (2005). A non-linear relationship between controller workload, task load, and traffic density: The straw that broke the camel's back. In *Proc. 2005 Int. Symp. Aviation Psychology*, (pp. 438-444). Retrieved March 11, 2021, from https://corescholar.libraries.wright.edu/isap_2005/66

Lehrer, P., Karavidas, M., Lu, S.-E., Vaschillo, E., Vaschillo, B., & Cheng, A. (2010). Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory

study. *International Journal of Psychophysiology, 76*, 80–87. https://doi.org/10.1016/j.ijpsycho.2010.02.006

Lei, S., & Roetting, M. (2011). Influence of task combination on EEG spectrum modulation for driver workload estimation. *Human Factors, 53*, 168–179. https://doi.org/10.1177/0018720811400601

Link, S. W. (1992). *The wave theory of difference and similarity.* Lawrence Erlbaum Associates and Routledge.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical rewiew and broader implication. In H. F. Society (Hrsg.) *Human Factors, 49*, 376–399.

Masin, S. C., Zudini, V., & Antonelli, M. (2009). Early alternative derivations of Fechner's law. *Journal of the History of Behavioral Sciences, 45*, 56–65.

Moehlenbrink, C., Papenfuss, A., & Jakobi, J. (2011). The role of workload for work organisation in a Remote Tower Control Center. In *Proc. 91h USA/Europe Air Traffic Management Research and Development Seminar (ATM2011).*

Mulder, G., & Mulder, L. J. (1981, 7). Information processing and cardiovascular control. *Psychophysiology, 18*, 392–402. https://doi.org/10.1111/j.1469-8986.1981.tb02470.x

Mulder, L., Waard, D., & Brookhuis, K. (2004). Estimating mental effort using heart rate variability. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 20.1–20.8). CRC Press. https://doi.org/10.1201/9780203489925.ch20

Norwich, K. H. (1987). On the theory of Weber fractions. *Perception and Psychophysics, 42*(3), 286–298.

Prinzel III, L. J., Parasuraman, R., Freeman, F. G., Scerbo, M. W., & Mikulka, P. J. (2003). Three experiments examining the use of electroencephalogram, event-related potentials, and heart-rate variability for real-time human-centered adaptive automation. NASA/TP-2003-212442, NASA Langley Research Center, Hampton, VA.

Radüntz, T. (2017). Dual frequency head maps: A new method for indexing mental workload continuously during execution of cognitive tasks. *Froniers in Physiology, 8*, 1–15. https://doi.org/10.3389/fphys.2017.01019

Radüntz, T., Fürstenau, N., Mühlhausen, T., & Meffert, B. (21. 4 2020). Indexing mental workload during simulated air traffic control tasks by means of dual frequency head maps. *Frontiers in Physiology, 11*, 300. https://doi.org/10.3389/fphys.2020.00300

Radüntz, T., Fürstenau, N., Tews, A., Rabe, L., & Meffert, B. (2019). The effect of an exceptional event on the subjectively experienced workload of air-traffic controllers? In L. Longo, & M. C. Leva (Eds.), *Human mental workload: Models and applications* (Communications in Computer and Information Sciences ed., Vol. 1012, pp. 239–257). Springer International Publishing. https://doi.org/10.1007/978-3-030-14273-5

Radüntz, T., Mühlhausen, T., Freyer, M., Fürstenau, N., & Meffert, B. (2020). Cardiovascular biomarkers' inherent timescales in mental workload assessment

during simulated air traffic control. *Applied Psychophysiology and Biofeedback.* https://doi.org/10.1007/s10484-020-09490-z

Reid, G., Potter, S., & Bressler, J. (1989). *Subjective workload assessment technique (SWAT): A user's guide.* AAMRL. American National Standards Institute.

Russo, R. (2016). CAPAN methodology: Sector capacity assessment. In *Air traffic services system capacity seminar.* ICAO.

Sheridan, T. (1980). Mental workload—What is it? Why bother with it? *Human Factors Society Bulletin, 23*, 1–2.

Silberstein, D., & Dietrich, R. (2003). *Cockpit communication under high cognitive workload.* Buske.

Sperandio, J. C. (1978). The regulation of working methods as function of workload among air traffic controllers. *Ergonomics, 21*(3), 195–202.

Stein, E. (1985). *Air traffic controller workload: An examination of workload probe.* DOT/FAA/CT-TN84/24, DOT/FAA, Atlantic City, NJ.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153–181.

Stevens, S. S. (1975). Psychophysics: Introduction to its perceptual, neural and social prospects. Wiley.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as ameasure of workload. *Ergonomics, 39*(5), 740–748.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. http://www.tandfonline.com/doi/abs/10.1080/14639220210123806

Wickens, C. D., & Hollands, J. G. (2000). Attention, time sharing, and workload. In *Engineering psychology and human performance* (3rd ed., pp. 439–479). Prentice-Hall.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single and multitask environments. *Work and Stress, 14*(1), 74–99.

Yeh, Y.-Y., & Wickens, C. (1984). *The dissociation of subjective measures of mental workload and performance.* Techreport, Engineering-psychology Research Laboratory, University of Illinois at Urbana-Champaign.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparitive Neurology and Psychology, 18*, 459–482. http://psychclassics.yorku.ca/Yerkes/Law

Appendix D
Psychophysics of Mental Workload: Derivation of Model Equations

Norbert Fürstenau

The present appendix provides the derivation of theoretical work and task load model equations for parameter predictions and regression-based parameter estimates as used for human-in-the-loop simulation data analysis in chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations". The initial derivation was published in Fürstenau and Radüntz (2021). The concept of cognitive resource limitation as starting point for the formal derivation of the power law of subjective workload e. g. goes back to Kahnemann (e.g. Kahnemann, 1973, 2011). It is described in Wickens and Hollands (2000) as a concept to relate task performance to invested resources via a performance—resource function (a monotonously increasing function with decreasing slope). General references for Stephens' stimulus—response power law are Stevens (1957, 1975) and Link (1992). Initial work on the psychophysics approach to workload using Stevens law was published by Gopher and Braune (1984) and Gopher et al. (1985). A recent application was published by Bachelder and Godfroy-Cooper (2019). For further references see chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations".

D.1 Resource Limitation Model

The most simple formal approach for growth dynamics under limited resources of a characteristic system variable is given by the nonlinear (1st order, 2nd degree) logistic or Verhulst differential equation:

$$df/dt = \dot{f}(t) = c\, f(t)(1 - f(t)) \tag{D.1}$$

with normalized function $f(t) = F(t)/\Delta F$, growth interval $\Delta F = F_u - F_d$ between upper F_u and lower limit F_d. Parameter c is determined by boundary conditions and rate constant κ.

In the present context our interest is in energy consuming cognitive activity under limited cognitive resources such as attention, processing of information, and memory capacity. In what follows we assume the subjectively reported (1-dimensional) workload under (quasi) real time conditions to be a true measure of the cognitive resources used for executing a mental task. For derivation of specific workload and task load functions as dependent on the independent environmental (traffic) load variable n (aircraft/hour) we replace the time variable t by n and the function f as solution to Eq. (D.1) by the (unknown) work or task load function W(n), i.e.

$$dw/dn = w'(n) = c\,w(n)(1 - w(n)) \tag{D.2}$$

with normalized variable $w = W/W_u$, and W_u = maximum WL. Clearly increasing amount of used mental resources required for completing increasing cognitive task demand corresponds to vanishing free cognitive resources $\Omega = W_u - W$ (with W_u = asymptotic resource limit), or $\omega = 1 - w$ that is modelled by the same differential equation, although with negative constant c.

The interpretation of the dynamical Eq. (D.2) is as follows: mental workload W is defined as the fraction of required cognitive resources (perception, processing, attention, memory) from available (normalized) resources W_u (normalized by upper limit: $w = W/W_u$, $w_u = 1$) and by definition it is a continuous measure. We assume that subjective (quasi real time) WL self-reports within a limited scale (and for practical purpose with discrete levels $W = 1 (= W_d)$, 2, ...,W_u) represent a direct measure of the actual fraction of available cognitive resources used to complete a task generated by an external stimulus n (representing the environmental load that may be the trigger for different specific tasks for achieving a goal, in our case for the controllers to establish safe and efficient air traffic). We look for the exact functional dependence w(n) that according to Eq. (D.2) for positive c increases monotonously with n. For low task requirements we allow for a region of underload where increase of task requirements exhibits weak or even vanishing effect on WL. For low w (and n), we assume a linear dependency of change with w, i.e. $dw/dn = \kappa w$ (i.e. an exponential increase $w = \exp(\kappa n)$ with rate constant κ). When approaching the resource limit W_u however, we expect a corresponding decrease of the WL change with w that is modelled by the factor $(1 - w)$, approaching 0 (i.e. $dw/dn = 0$) with $w \rightarrow w_u$.

Equation (D.1) is a special version of the Bernoulli equation with time dependent coefficients, $c_1(t)$, $c_2(t)$ for the linear and quadratic terms that for the former simplify to ($c_1 = c_2 = c$). The well known general solution of Eq. (D.2) is the logistic or sigmoid function with lower and upper asymptotic limits $w(n \rightarrow -\infty) = 0$ and $w(n \rightarrow +\infty) = 1$ respectively:

$$w(n) = \frac{1}{1 + k\,exp\{-\kappa n\}} \tag{D.3}$$

With rate parameter κ, shift factor $k = \exp\{\kappa m\}$, $m =$ translation (shift parameter). Differentiation of Eq. (D.3) yields (D.2) with $c = \kappa$. Slope at the sigmoid inversion point $n = m$ is $w'(n = m) = \kappa/4$. Boundary condition $w(n = 0) = w_d = W_d/W_u$ (i.e. minimum WL level for zero environmental load) yields:

$$k = \frac{W_u}{W_d} - 1 = \exp\{\kappa m\} \tag{D.4}$$

It is easily verified that for large negative and positive arguments n the sigmoid approaches asymptotes $w(n \to -\infty) = 0$ ($W(n) \to 0$) and $w(n \to +\infty) = 1$ ($W(n) \to W_u$) respectively. For zero shift $m = 0$ the resulting sigmoid curve intersects the ordinate at $w = 0.5$. Because for online (near real time) WL measures like ISA ($W_u = 1$) and ATWIT ($W_u = 7$; see Appendix C) the fixed scale range $W_d = 1 \le W \le W_u$ represents prior knowledge (limits of possible WL reporting), the two logistic parameters (k, κ) or $(m, \tau = 1/\kappa$ shift, sensitivity) are related via:

$$m = \tau ln\left(\frac{W_u}{W_d} - 1\right) = \tau ln(k) \tag{D.5}$$

so that model Eq. (D.3) is characterized by a single free parameter m or τ.

In order to allow adjustment of the model by a vertical offset (keeping the scale range $[W_d, W_u]$) we introduce (for normalized WL $w = W/W_u$) the offset parameter Δ:

$$w(n) = \frac{1 + \Delta}{1 + exp\{-\frac{n-m}{\tau}\}} - \Delta = \frac{1 + \Delta}{1 + kexp\{-\frac{n}{\tau}\}} - \Delta \tag{D.6}$$

For $\Delta = 0$ we obtain model Eq. (D.3) with $w(n \to -\infty) = 0$, $w(n \to +\infty) = 1$. For $\Delta = 1$ Eq. (D.6) may be expressed by $w(n) = \tanh\{(n-m)/2\tau\}$ and the sigmoid extends between $-1 \le w \le +1$ (only $w \ge 0$ to be used for load reporting). The positive valued load measure $w(n)$ starts with linear slope w' (at inversion point $\{n, w\} = \{m, 0\}$) $= 1/2\tau$. Via differentiation of Eq. (D.6) we again obtain the dynamical equation (D.2) with different coefficients for the linear and quadratic term: $c_1 = \kappa(1 - \Delta)/(1 + \Delta)$ and $c_2 = \kappa/(1 + \Delta)$. Figure D.1 depicts an example representing the ISA scale with WL-scale limits $W_u = 5$, $W_d = 1$. Setting intercept $W(0) := W_d$, yields $\tau = m/ln(k)$, and $m := 10$, $\Delta := 0$.

In contrast to the assumptions in Fürstenau and Radüntz (2021) and chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations" concerning ISA-WL intercept $W(0) := 1$ with lower asymptote $W(n \to -\infty) = 0$, experimental results may suggest for low environmental load (traffic $n < n_c$) an extended WL-reporting range with $W(n < n_c) \approx W_d$, so that intercept $\approx$ asymptote has to be assumed as $W(n \to -\infty) = 1$. This is realized in Eq. (D.6) with $\Delta = -W(-\infty)/W_u = -1/W_u$ and sufficiently large shift parameter $m \gg \tau$. The actual (effective) intercept is derived as $W(n = 0) := W_e = (W_u + k)/(1 + k)$, with $k = \exp(m/\tau)$. This proves that for $\tau/m \ll 1$ the exponential $k \gg$

Fig. D.1 Logistic (sigmoid) curve matched to ISA workload scale with intersection at WL(n = 0) = 1 and inversion point (shift parameter) m: = n = 10. $\tau(m) = m/ln(4)$
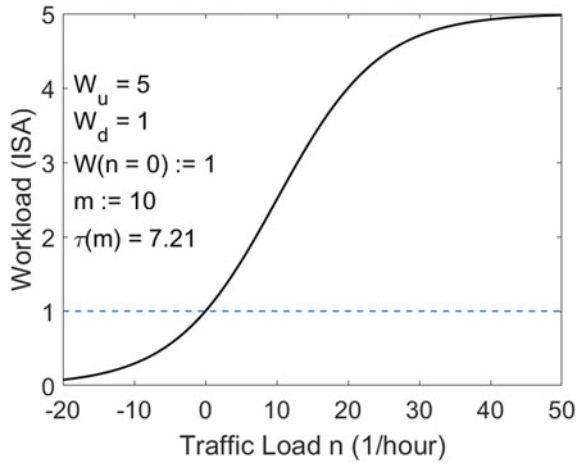
W_u so that in fact $W(n = 0) \approx W(-\infty)$. Corresponding results can be shown for WL data reported by Lee (2005) and Lee et al. (2005) who to our best knowledge were the first to suggest on a heuristic basis a nonlinear sigmoid characteristic for WL(n). This is illustrated in Appendix C, Sect. 6.2 (Fig. C.3).

A generalized liner version of the logistic model Eq. (D.6) is obtained by some basic algebraic operations, after taking the logarithm:

$$y_w(n) = ln\left(\frac{\Delta + w(n)}{1 - w(n)}\right) = ln(T_w(n)) = \frac{1}{\tau}n - \frac{m}{\tau} \qquad (D.7)$$

Where $T_w(n)$ defines the nonlinear transformation $(= \exp\{\kappa(n - m)\})$ of the normalized workload w(n).

A stochastic Bernoulli-Langevin equation is obtained by adding a random noise term to Eq. (D.1) or (D.2). By transformation into the equivalent Fokker–Planck stochastic partial differential equation (Risken, 1988) with deterministic (Bernoulli) drift and stochastic (e.g. Gaussian) diffusion term it allows for modelling the time dependent dynamics of the probability density with mean f(t). From an experimental point of view the theoretically predicted parameter values obtained with the model equations and prior knowledge (scale range, domain experts information) should correspond to regression based parameter estimates of measured work and task load data averaged across a sufficiently large random sample of well trained participants. In what follows we apply the logistic model to the case of the subjective one dimensional quasi real time five level Instantaneous Self Assessment (ISA) WL-reports and to the objective communication task load measure represented by ATCO's frequency of radio calls R(n) (calls/h).

D.2 Logistic ISA(n)-WL Model

For the ISA-WL scale ($W_d: = I_d$, $W_u: = I_u$, scaling parameter $\tau: = \nu$, shift parameter $m: = \mu$) with $I_d = 1 \leq I \leq I_u = 5$ we assume that vanishing environmental load n implies ISA-WL self report $I(n = 0) = I_d = 1$, corresponding to the lower scale limit. This implies $I_d >$ lower asymptote $\lim I(n \rightarrow -\infty) = 0$ which is fulfilled by Eq. (D.6) with $\Delta = 0$. From Eq. (D.5) we get $k = \exp\{\mu/\nu\} = I_u/I_d - 1 = 4$ so that Eq. (D.6) turns into:

$$p(n) = \frac{1}{1 + 4exp\left\{-\frac{n}{\nu}\right\}} \tag{D.8}$$

i.e. a 1-parametric sigmoid characteristic as theoretical model for the ISA self reports on subjectively perceived workload in normalized form $p(n) = I(n)/I_u$. Maximum slope is obtained at inversion point ($n = \mu$, ISA $= I_u/2$): $dI/dn = I'(n = \mu) = I_u/4\nu$. This means that the inverse of the logistic scaling parameter ν represents the slope or WL sensitivity in the nearly linear range in the vicinity of the sigmoid inversion point.

An exponential version of the logistic characteristic is obtained from Eq. (D.8) according to Eq. (D.7) after nonlinear transformation P(ISA): $= ISA/(I_u - ISA) = p(ISA)/(1 - p(ISA)$ yielding the transformed ISA-WL-variable

$$P(n) = \frac{1}{4}exp\left(\frac{n}{\nu}\right) \tag{D.9}$$

Logarithmation yields a (generalized linear) semilogarithmic characteristic according to Eq. (D.7) with linear dependence of the transformed ISA self report measures P(n).

$$y_p(n) = \ln(P) = \frac{1}{\nu}n - ln(4) = a_g n + b_g \tag{D.10}$$

Consequently a linear 1-parameter regression of the transformed WL data would provide a good test for the theoretical resource limitation model of the ISA-WL measure by means of the confidence of the a_g estimate, with theoretical prediction $b_g = -1.3863$ as prior information.

D.3 Linearized Logistic ISA Model

It appears plausible that the nearly linear WL range in the vicinity of the logistic inversion point $n = \mu$ corresponds to the operational traffic flow (environmental load) interval, because it avoids nonlinear saturation effects when approaching the upper asymptote. The results in chapter "Model Based Analysis of Subjective Mental

Workload During Multiple Remote Tower Human-In-The-Loop Simulations" show that the respective interval of the logistic characteristic may be sufficiently represented by a linearized model for the selected traffic flow range $0 \leq n \leq 10$ (AC/2 min). In the vicinity of the logistic shift parameter μ with $n = \mu + \Delta n$, $|\Delta n| = |n - \mu| \ll \mu$, $I(n = \mu) = I_u/2$ the logistic characteristic is approximated by

$$I(n) \approx b_{lt} + a_{lt}n \approx \frac{I_u}{2}\left(1 - \frac{\mu}{2\nu}\right) + \frac{I_u}{4\nu}n = \frac{I_u}{2}\left(1 - \frac{\ln(4)}{2}\right) + \frac{I_u\ln(4)}{4\mu}n \quad \text{(D.11)}$$

where the slope in the vicinity of μ, $dI/dn \approx a_{lt} = I_u/(4\nu) = I_u \ln(4)/(4\mu)$ is obtained through neglection of quadratic and higher exponents $(n - \mu)/\nu$ of the linearization.

Within the linear approximation a dimensionless index $s_{ak}(a_k, b_k)$ can be defined for individual participants $k = 1\ldots21$ (not to be confused with exponential $k(m)$ (Eq. (D.4)) via slope and intercept (a_k, b_k) by normalisation of the independent and dependent variables $(n, I(n))_k$ through division by the mean value of the traffic variable n $(n_1 + n_4)/2$ and the individual ISA intervals $(I_{max} + I_{min})/2$ respectively. In what follows we skip the participant index k. Normalized variables are written with small letters or index n, yielding $n_n = \frac{n}{(n_1 + n_4)/2} = \frac{n}{40}$, and for the individual ISA scenario means $\langle ISA \rangle_{jk}$:

$$\langle ISA \rangle = \frac{\langle ISA \rangle}{\frac{\langle ISA \rangle_{max} + \langle ISA \rangle_{min}}{2}} = \frac{\langle ISA \rangle}{40a + b} = s_b + s_a n_n \quad \text{(D.12)}$$

where the individual maximum and minimum ISA values are taken from the linear regression predictions at the maximum n_4 and minimum n_1 of the predictor variable. This yields the dimensionless linear sensitivity index

$$s_a = \frac{\langle isa \rangle_{max} - \langle isa \rangle_{min}}{n_{nmax} - n_{nmin}} = \frac{1}{1 + \frac{b}{40a}} \quad \text{(D.13)}$$

with anticorrelation $s_b = 1 - s_a$. s_a allows for clustering of participants k with individual sensitivities a_k as used in Radüntz et al. (2020) for discriminating between subjects of low and high WL sensitivity.

D.4   Logistic Radio Calls Frequency Model RC(n)

It appears plausible to assume the radio calls frequency as task load to increase linearly with traffic load from $n = 0$, and at very high load levels like ISA(n) to approach asymptotically the absolute maximum R_u, i.e. an arbitrary time interval (e.g. 2 min for the experiment of chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations") filled

by cumulative ATCO—pilot radio communications. The hypothesized logistic radio calls characteristic R(n) with linear increase and maximum slope at n = 0 is obtained from Eq. (D.6) by setting m = 0 and offset parameter Δ: = 1. With normalized call rate s: = $R(n)/R_u$ and x: = $n/(2\rho)$ Eq. (D.6) can be written in the form s(x) = (1 − exp(−2x))/(1 + exp(−2x)) = (exp(x) − exp(−x))/(exp(x) + exp(−x)) which is the definition of the hyperbolic function tanh(x)

$$s(n) = \left[\frac{2}{1 + \exp\{-n/\rho\}} - 1 \right] = tanh\left(\frac{n}{2\rho} \right) \tag{D.14}$$

With logistic inversion point at the origin (n, s) = (0, 0) it starts with slope s'(n) = ds/dn given by

$$s' = \frac{1}{2\rho cosh^2(n/2\rho)} \tag{D.15}$$

Maximum slope (at n = 0) is $s'(n = 0) \approx \Delta s/\Delta n = 1/2\rho$.

Like for ISA-WL the generalized linear form for R(n) is obtained via nonlinear transformation S(s) of the normalized TL-variable s(n) that is obtained from Eq. (D.14) by some basic algebraic operations: S(s) = (1 + s(n))/(1 − s(n)) = exp{n/ρ}, followed by taking the logarithm:

$$y_s(n) = ln(S) = \frac{1}{\rho}n \tag{D.16}$$

with slope $a_{gs} = 1/\rho$.

D.5 Psychophysics Power Law Model ISA(RC)

In the simulation experiment of chapter "Model Based Analysis of Subjective Mental Workload During Multiple Remote Tower Human-In-The-Loop Simulations" with environmental load variable n [AC/h] the function pair (R(n), I(n)) defines the parametric inter-dependence of the subjectively perceived and reported ISA-WL and objectively measured frequency of radio calls R(n) as communication taskload (TL). By formal combination of the logistic WL- and TL-characteristics Eqs. (D.8) and (D.14) respectively, we can derive the psychophysical power law relationship between subjective response P(ISA(n)) (representing fraction of used cognitive processing resources) and objective stimulus TL-measure p(s(n). This is easily achieved by means of the generalized linear versions of the logistic functions Eqs. (D.10) and (D.16). A more general expression for the logistic characteristics is obtained by transformation of Eq. (D.6) into the generalized linear form that includes the general bias term Δ and the general scaling and shift parameters τ and

m respectively:

$$y(n) = ln\left(\frac{\Delta + p(n)}{1 - p(n)}\right) = ln(P(n)) = \frac{1}{\tau}n - \frac{m}{\tau} \qquad (D.17)$$

Based on this general expression we define separate equations for objective (TL) stimulus $y_s(n)$ and perception (of mental load) $y_p(n)$

$$y_s(n) = ln\left(\frac{\Delta_s + s(n)}{1 - s(n)}\right) = ln(S(n)) = \frac{1}{\varrho}n - \frac{m_s}{\varrho} \qquad (D.18)$$

$$y_p(n) = ln\left(\frac{\Delta_p + p(n)}{1 - p(n)}\right) = ln(P(n)) = \frac{1}{\nu}n - \frac{m_p}{\nu} \qquad (D.19)$$

By solving Eq. (D.18) for n ($\Delta_s = 1$) and introducing the expression into (D.19) ($\Delta_p = 0$) we obtain the generalized linear form of the power law:

$$y_p(y_s) = ln(P(n)) = \gamma y_s + \frac{m_s - m_p}{\nu} = \gamma y_s + b_s \qquad (D.20)$$

with exponent γ of the power law for ISA(RC) defined by the ratio of the scaling coefficients of stimulus ρ and perception $\nu : \rho/\nu := \gamma$. With $m_s = 0$ and $m_p = \mu$ we have $b_s = b_{gt} = -ln(4)$. Equation (D.20) defines the WL–TL relationship in terms of a 1-parameter power law model with exponent γ in correspondence to Stevens' well known generalized linear relationship between physical stimulus (e.g. brightness, loudness) and subjective response. Of course Eq. (D.20) may be written also as explicit power law in linear coordinates

$$P(S) = \exp\left\{-\frac{\mu}{\nu}\right\}S^\gamma = \frac{1}{k}S^\gamma \qquad (D.21)$$

where P, S are the nonlinear transformations of I(n), R(n) as defined in Eq. (D.7) by $T_w(n)$, with $\Delta = 0$ and 1 for I(n) and R(n) respectively.

References

Bachelder, E., & Godfroy-Cooper, M. (2019). Pilot workload esimation: Synthesis of spectral requirements analysis and Weber's law. 322514/620191228

Fürstenau, N., & Radüntz, T. (2021). Power law model for subjective mental workload and validation through air-traffic control human-in-the-loop simulation. *Cognition, Technology, and Work*. Retrieved May 30, 2021, from https://doi.org/10.1007/s101011-021-00681-0

Fürstenau, N., Mittendorf, M., & Kamo, S. (2016). Nonlinear dynamics approach for modeling of air traffic performance disruption and recovery. In Eurocontrol (Ed.), *Proc. 7th Int.Conf. Res. in Air Transportation (ICRAT 16)*. Eurocontrol.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures. *Human Factors, 26*(5), 519–532.

Gopher, D., Chillag, N., & Arzi, N. (1985). The psychophysics of workload—A second look at the relationship between subjective measures and performance. In *Proceedings of the Human Factors Society* (29th Annual Meeting) (pp. 640–644).

Kahnemann, D. (1973). *Attention and Effort.* Prentice Hall.

Kahnemann, D. (2011). *Thinking, fast and slow* (German edition, 2012, Siedler Verlag ed.). Farrar, Straus, Giraux.

Lee, P. U. (2005). A non-linear relationship between controller workload and traffic count. In *Proc. Human Factors and Ergonomics Society, 49th meeting 2005* (Vol. 49, pp. 1129–1133). https://doi.org/10.1177/154193120504901206

Lee, P. U., Mercer, J., Smith, N., & Palmer, E. (2005). A non-linear relationship between controller workload, task load, and traffic density: The straw that broke the camel's back. In *Proc. 2005 Int. Symp. Aviation Psychology* (pp. 438–444). Retrieved March 11, 2021, from https://corescholar.libraries.wright.edu/isap_2005/66

Link, S. W. (1992). *The wave theory of difference and similarity.* Lawrence Erlbaum Associates and Routledge.

Radüntz, T., Fürstenau, N., Mühlhausen, T., & Meffert, B. (2020). Indexing mental workload during simulated air traffic control tasks by means of dual frequency head maps. *Frontiers in Physiology, 11*, 300. https://doi.org/10.3389/fphys.2020.00300

Risken, H. (1988). *The Fokker-Planck equation* (2nd ed.). Springer.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153–181.

Stevens, S. S. (1975). *Psychophysics.* Wiley.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/14639220210123806

Wickens, C. D., & Hollands, J. G. (2000). Attention, time sharing, and workload. In *Engineering Psychology and Human Performance* (3rd ed., pp. 439–479). Prentice-Hall.

Wickens, C., & Hollands, J. (2000). Engineering psychology and human performance.