Don Harris (Ed.)

LNAI 10275

Engineering Psychology and Cognitive Ergonomics

Performance, Emotion and Situation Awareness

14th International Conference, EPCE 2017
Held as Part of HCI International 2017
Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part I

1 Part I

HCI 2017
INTERNATIONAL

Springer

Lecture Notes in Artificial Intelligence 10275

Subseries of Lecture Notes in Computer Science

More information about this series at http://www.springer.com/series/1244

Don Harris (Ed.)

Engineering Psychology and Cognitive Ergonomics

Performance, Emotion and Situation Awareness

14th International Conference, EPCE 2017
Held as Part of HCI International 2017
Vancouver, BC, Canada, July 9–14, 2017
Proceedings, Part I

Springer

Editor
Don Harris
Faculty of Engineering and Computing
Coventry University
Coventry
UK

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-58471-3 ISBN 978-3-319-58472-0 (eBook)
DOI 10.1007/978-3-319-58472-0

Library of Congress Control Number: 2017939722

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

The 19th International Conference on Human–Computer Interaction, HCI International 2017, was held in Vancouver, Canada, during July 9–14, 2017. The event incorporated the 15 conferences/thematic areas listed on the following page.

A total of 4,340 individuals from academia, research institutes, industry, and governmental agencies from 70 countries submitted contributions, and 1,228 papers have been included in the proceedings. These papers address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The papers thoroughly cover the entire field of human–computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas. The volumes constituting the full set of the conference proceedings are listed on the following pages.

I would like to thank the program board chairs and the members of the program boards of all thematic areas and affiliated conferences for their contribution to the highest scientific quality and the overall success of the HCI International 2017 conference.

This conference would not have been possible without the continuous and unwavering support and advice of the founder, Conference General Chair Emeritus and Conference Scientific Advisor Prof. Gavriel Salvendy. For his outstanding efforts, I would like to express my appreciation to the communications chair and editor of *HCI International News*, Dr. Abbas Moallem.

April 2017 Constantine Stephanidis

HCI International 2017 Thematic Areas and Affiliated Conferences

Thematic areas:

- Human–Computer Interaction (HCI 2017)
- Human Interface and the Management of Information (HIMI 2017)

Affiliated conferences:

- 17th International Conference on Engineering Psychology and Cognitive Ergonomics (EPCE 2017)
- 11th International Conference on Universal Access in Human–Computer Interaction (UAHCI 2017)
- 9th International Conference on Virtual, Augmented and Mixed Reality (VAMR 2017)
- 9th International Conference on Cross-Cultural Design (CCD 2017)
- 9th International Conference on Social Computing and Social Media (SCSM 2017)
- 11th International Conference on Augmented Cognition (AC 2017)
- 8th International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management (DHM 2017)
- 6th International Conference on Design, User Experience and Usability (DUXU 2017)
- 5th International Conference on Distributed, Ambient and Pervasive Interactions (DAPI 2017)
- 5th International Conference on Human Aspects of Information Security, Privacy and Trust (HAS 2017)
- 4th International Conference on HCI in Business, Government and Organizations (HCIBGO 2017)
- 4th International Conference on Learning and Collaboration Technologies (LCT 2017)
- Third International Conference on Human Aspects of IT for the Aged Population (ITAP 2017)

Conference Proceedings Volumes Full List

1. LNCS 10271, Human–Computer Interaction: User Interface Design, Development and Multimodality (Part I), edited by Masaaki Kurosu
2. LNCS 10272 Human–Computer Interaction: Interaction Contexts (Part II), edited by Masaaki Kurosu
3. LNCS 10273, Human Interface and the Management of Information: Information, Knowledge and Interaction Design (Part I), edited by Sakae Yamamoto
4. LNCS 10274, Human Interface and the Management of Information: Supporting Learning, Decision-Making and Collaboration (Part II), edited by Sakae Yamamoto
5. LNAI 10275, Engineering Psychology and Cognitive Ergonomics: Performance, Emotion and Situation Awareness (Part I), edited by Don Harris
6. LNAI 10276, Engineering Psychology and Cognitive Ergonomics: Cognition and Design (Part II), edited by Don Harris
7. LNCS 10277, Universal Access in Human–Computer Interaction: Design and Development Approaches and Methods (Part I), edited by Margherita Antona and Constantine Stephanidis
8. LNCS 10278, Universal Access in Human–Computer Interaction: Designing Novel Interactions (Part II), edited by Margherita Antona and Constantine Stephanidis
9. LNCS 10279, Universal Access in Human–Computer Interaction: Human and Technological Environments (Part III), edited by Margherita Antona and Constantine Stephanidis
10. LNCS 10280, Virtual, Augmented and Mixed Reality, edited by Stephanie Lackey and Jessie Y.C. Chen
11. LNCS 10281, Cross-Cultural Design, edited by Pei-Luen Patrick Rau
12. LNCS 10282, Social Computing and Social Media: Human Behavior (Part I), edited by Gabriele Meiselwitz
13. LNCS 10283, Social Computing and Social Media: Applications and Analytics (Part II), edited by Gabriele Meiselwitz
14. LNAI 10284, Augmented Cognition: Neurocognition and Machine Learning (Part I), edited by Dylan D. Schmorrow and Cali M. Fidopiastis
15. LNAI 10285, Augmented Cognition: Enhancing Cognition and Behavior in Complex Human Environments (Part II), edited by Dylan D. Schmorrow and Cali M. Fidopiastis
16. LNCS 10286, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Ergonomics and Design (Part I), edited by Vincent G. Duffy
17. LNCS 10287, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Health and Safety (Part II), edited by Vincent G. Duffy
18. LNCS 10288, Design, User Experience, and Usability: Theory, Methodology and Management (Part I), edited by Aaron Marcus and Wentao Wang

19. LNCS 10289, Design, User Experience, and Usability: Designing Pleasurable Experiences (Part II), edited by Aaron Marcus and Wentao Wang
20. LNCS 10290, Design, User Experience, and Usability: Understanding Users and Contexts (Part III), edited by Aaron Marcus and Wentao Wang
21. LNCS 10291, Distributed, Ambient and Pervasive Interactions, edited by Norbert Streitz and Panos Markopoulos
22. LNCS 10292, Human Aspects of Information Security, Privacy and Trust, edited by Theo Tryfonas
23. LNCS 10293, HCI in Business, Government and Organizations: Interacting with Information Systems (Part I), edited by Fiona Fui-Hoon Nah and Chuan-Hoo Tan
24. LNCS 10294, HCI in Business, Government and Organizations: Supporting Business (Part II), edited by Fiona Fui-Hoon Nah and Chuan-Hoo Tan
25. LNCS 10295, Learning and Collaboration Technologies: Novel Learning Ecosystems (Part I), edited by Panayiotis Zaphiris and Andri Ioannou
26. LNCS 10296, Learning and Collaboration Technologies: Technology in Education (Part II), edited by Panayiotis Zaphiris and Andri Ioannou
27. LNCS 10297, Human Aspects of IT for the Aged Population: Aging, Design and User Experience (Part I), edited by Jia Zhou and Gavriel Salvendy
28. LNCS 10298, Human Aspects of IT for the Aged Population: Applications, Services and Contexts (Part II), edited by Jia Zhou and Gavriel Salvendy
29. CCIS 713, HCI International 2017 Posters Proceedings (Part I), edited by Constantine Stephanidis
30. CCIS 714, HCI International 2017 Posters Proceedings (Part II), edited by Constantine Stephanidis

Engineering Psychology and Cognitive Ergonomics

Program Board Chair(s): **Don Harris, UK**

- Henning Boje Andersen, Denmark
- Martin Baumann, Germany
- Nicklas Dahlstrom, United Arab Emirates
- Shan Fu, P.R. China
- John Huddlestone, UK
- Kyeong-ah Kate Jeong, USA
- Wen-Chin Li, UK
- Andreas Luedtke, Germany
- Randy Mumaw, USA
- Jan Noyes, UK
- Paul Salmon, Australia
- Axel Schulte, Germany
- Patrick Waterson, UK
- Alf Zimmer, Germany

The full list with the Program Board Chairs and the members of the Program Boards of all thematic areas and affiliated conferences is available online at:

http://www.hci.international/board-members-2017.php

HCI International 2018

The 20th International Conference on Human–Computer Interaction, HCI International 2018, will be held jointly with the affiliated conferences in Las Vegas, NV, USA, at Caesars Palace, July 15–20, 2018. It will cover a broad spectrum of themes related to human–computer interaction, including theoretical issues, methods, tools, processes, and case studies in HCI design, as well as novel interaction techniques, interfaces, and applications. The proceedings will be published by Springer. More information is available on the conference website: http://2018.hci.international/.

General Chair
Prof. Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
E-mail: general_chair@hcii2018.org

http://2018.hci.international/

Contents – Part I

Mental Workload and Performance

Psychological and Emotional Issues in Interaction

Situation Awareness and Control

Contents – Part II

Cognition and Design

Cognition in Aviation and Space

Cognition and Driving

Mental workload and performance

A Method to Estimate Operator's Mental Workload in Multiple Information Presentation Environment of Agricultural Vehicles

Xiaoping Jin(✉), Bowen Zheng, Yeqing Pei, and Haoyang Li

China Agricultural University, Beijing 100083, China
`jinxp@cau.edu.cn`

Abstract. The development of mechanized agriculture has brought many new features to agricultural vehicles today. Considering more human limitations, such as situational awareness and the potential for mental overload, operator's mental workload seems to be a critical issue. This study presents a quantitative method for determining the mental workload imposed on agricultural vehicle operators and to express their mental workload by means of a model that was objective and could stand up to the test of validity. The proposed model consisted of three task elements and six weighted moderating factors. And a modified task workload analysis measurement was used to validate the model by analyzing video recordings of eighteen subjects working in three combine harvester vehicles over an eight-hour shift. The comparison results implied the model's validity. Subjective workload was mildly correlated with the three task elements in the model. It is shown to provide an objective method for assessment and prediction of operator's mental workload in the multiple information presentation environment of farming machines.

Keywords: Agricultural vehicle · Cognitive ergonomics · Mental workload · Information presentation · Task analysis

1 Introduction

Driving is a task that requires comprehensive capacities of drivers, not only the ability of action, but also the ability of perception and decision making. Moreover, the driving needs driver to highly coordinate these different dimensions of requirements and quickly respond under time pressure [1, 2]. For operators of agricultural vehicles, they are faced with greater challenges and more severe mental workload by completing additional field tasks. In more complex situations, they are struggling to cope with a variety of unexpected situations and have to deal with a variety of information simultaneously. In recent years, with more birth of vehicle information and entertainment systems (such as GPS navigation systems, DVD players, etc.), it is more common for operators to work in multiple task driving conditions, thus making mental workload increasing [3]. While human operator is regarded as communication channels with limited and fixed capacity, they are facing higher risks in face of an excess of

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 3–20, 2017.
DOI: 10.1007/978-3-319-58472-0_1

information. Based on this, researches have been done to find out mechanisms of operator mental workload, so that a theoretical basis can be provided for safer and more comfortable operating-related design.

With the development of cognitive theory, cognitive models are widely used in analysis of mental workload. The term workload covers a broad spectrum of human activity. Mental workload focuses on activities that are primarily mental or cognitive in nature and may involve physical coordination, but excludes activity resulting only in muscular fatigue. Not only is there no single, commonly accepted definition of workload but there are many conflicting concepts of what workload comprises [4]. Mental workload has ever been defined as a function of capacity, in terms of experienced load or time load. In Wickens' theory, workload can be defined in terms of relationship between resource supply and task demand. It is argued that operator workload is directly related to the extent to which the tasks performed by the operator utilizes the limited resources. Wickens' multiple resource theory proposes that the human operator does not have one single information processing source that can be tapped, but several different pools of resources that can be tapped simultaneously, such as visual, auditory, cognitive and psychomotor [5]. Depending on the nature of the task, these resources may have to process information sequentially if the different tasks require the same pool of resources, or can be processed in parallel if the task requires different resources. In analysis of operator workload, because of low utility of auditory resource, often visual and auditory resources are combined as perceptual resource.

The assessment and prediction of operator workload has a vital impact on the design of new human-machine systems. By evaluating operator workload during design of a new system, or iteration of an existing system, problems such as workload bottlenecks and overload can be identified [6]. As the human operator is a central part of a human-machine system, the correction of these problems is necessary for the operation of safe and efficient systems.

Generally, there are four kinds of mental workload measurements as follows, measures of primary task performance, secondary measures of spare mental capacity, subjective rating techniques, physiological indicators of mental workload. These four methods have advantages and disadvantages in application. The primary task measure is task-oriented method to compute the mental workload imposed by the task through subject's performance. Its weakness is that the result is varied with the nature of work so that it is hard to compare two different tasks. Further, the performance measurement indicators did not directly tell about mental workload.

The secondary measures is applied to a wide range of multiple task driving. Subjects are engaged in two tasks simultaneously, the primary task to consume most energy and the secondary task to occupy spare mental capacity. Its disadvantage is that it is hard to find out a standard and proper secondary task and often causes interference on primary tasks, making measurements not reflecting main mental workload.

The physiological measures assume mental workload can be measured through physiological changes [7–10]. Indicators such as the heart rate and P300 were found to be correlated with the degree of novelty of the information people received [11]. This method can continuously measure data and does not affect the performance of primary tasks, simple and natural. However, it requires special equipment. Many other factors that have nothing to do with mental workload might have influences on the changes of physiological indicators. Different brain resources may lead to different physiological responses, affecting the universality of this approach.

Subjective evaluation is an important tool in assessing mental workload. This is by far the easiest and most popular method of measuring mental workload [12]. Subjects were asked to determine and report their mental workload some task caused on them. It is more practical since it is easy to implement without instrument and has higher sensitivity. The theoretical basis for this method is that the subject's ability and their personal effort to accurately report are related. Some researchers claim that subjective measurement can best reflect the nature of the mental workload [13]. NASA Task Load Index (NASA-TLX), SWAT are commonly-used subjective rating scales. However, it may be impacted by subject's education background since mental workload seems a little metaphysical for people who stay far away from research area. Mental fatigue might be a good way to explain.

Human operators perceive information constantly from the environment. Day-to-day interaction with a vehicle is definitely a processing task. They receive information input (visually or auditory or tactile) and process the information to make decisions on how to perform primary tasks (action). When multiple tasks begin to compete for the operator's attention, his/her mental workload possibly approaches the capacity limits, resulting in lower efficiency or accidents. This cognitive overload phenomenon was examined by Wickens and is described by his general model of human information processing [5]. Basically people have a fixed amount of attention (resources) to give and when that is exceeded, their performance will decline.

The development of mechanized agriculture has brought many new features to agricultural vehicles today, making the human-machine interface more and more complicated and functional variety. Generally intended to improve productivity and user satisfaction, poor implementation of these features without due consideration of operator requirements and/or limitations can have negative consequences. The more complex a system is, the more complex human interaction with it becomes. Human limitations have to be taken into considerations, such as situational awareness and the potential for operator mental overload created by these ancillary technologies. Sometimes electronic control mediation and automation can help mitigate the impacts to operators, but if the human capabilities are not addressed or understood properly, the opposite can happen: loss of awareness, cognitive overload, and the resulting decreased human-machine performance. How people use machines is just as important as how people fit in machines [3]. Past estimation techniques have focused on measuring workload directly from the drivers or inferring it from traffic factors. The limitations of these techniques are interfering into the operating job and not being able to capture the differences amongst individuals. And few attention has been paid to operator's mental workload of farming vehicles though they are working in a really intense environment.

The objective of this study is to develop a quantitative method for determining the mental workload imposed on agricultural vehicle operators and to express their mental workload by means of an index that is objective and can stand up to the tests of validity. Two distinct approaches are used to measure and predict the mental workload of agricultural vehicle operators: objective workload prediction and task analysis workload measurement, which would be a validation for the prediction model.

2 Method

The guiding principle in this approach was to develop a methodology that would provide an objective measurement of the task-related mental workload factors imposed on agricultural vehicle operators. The criteria that the mental workload indicators should meet would be as follows:

Validity. Validity of conducting was very important since the process had to measure the factor of mental workload only without any other unrelated factors.

Non-intrusiveness. The agricultural vehicle operators perform field tasks and the method should not distract them or affect their performance.

Sensitivity. The method should detect changes in the mental workload.

The aim of the method is to create a prediction tool that can be applied to classify different information presentations and control strategy at human-machine interface of agricultural vehicles in terms of the mental workload imposed on operators. This classification will in return facilitate designing and optimization of the human-machine interface resources to ensure high-efficiency and safe field operations.

2.1 Development Process

Besides physical level, people also interact with their products and environments on a cognitive level. From simple opinion to situational awareness of a complex system, the understanding of human perception, mental models, and human limitations is essential to a successful outcome [14]. Dooley [3] proposed there are a number ways that people relate to products on a mental level from operation aspects, such as control identification, ease of use, perception of complexity, fitness for task, information presentation and situational awareness.

Based on the above guiding principles, the mental workload prediction index system was developed over a number of discussions from an expert group, including 18 experienced operators, two agricultural vehicle owners and four technicians in university. The outcome was the following factors to be included. The first refers to the three task elements, the sum of which comprise the quantity of work. They are the number of joystick and pedal operations, the number of display observation and function actuations, and the number of mistake-correcting actions. The second includes the moderating factors that reflect the nature (complexity) of the work, including interface complexity (25%), shift (15%), operator's experience (20%), crop growing conditions (20%), climate conditions (10%), operating environment (10%).

Due to the differences in the importance of each task element and moderating factor to the overall mental workload, they were weighted (as the number in the braces), the weighting factors of which were also determined by the expert group. The mental workload model was in the following form:

$$\mathrm{MW} = \sum_{i=1}^{3} (TE_i \times W_i) \times \prod_{j=1}^{6} MF_j \tag{1}$$

where MW is the mental workload, TE_i are the three task elements, W_i are the weights of the three task elements, MF_j are the value of each weighted moderating factor.

2.2 Definition of the Elements

Number of Joystick and Pedal Operations. Taking the combine harvester for example. Joystick, handles and pedals are the most frequently used controllers by operators in field work, such as the primary clutch handle, reel height adjustment joystick, header height adjustment joystick, shift handle, unloading clutch handle, braking handle, hand throttle, throttle pedal, braking pedal and steering wheel, etc. These operations are regarded as the apparent reactions of mental workload since they are performed after operator's perceptive and cognitive activities. Any activity that would increase the workload had to be included in the indicators. The activities of joysticks and pedals increase the time demands on operators and require decision-making and accuracy and therefore concentration. The number of joystick and pedal operations could be obtained from video recordings.

Number of Display Observation and Function Actuations. It was considered for the purpose of reflecting the complexity of executing the tasks. For combine harvester operators, they pay more attention to the warning or alarms, such as low fuel pressure alarm, fuel indicator, water temperature warning, axle roller jam alarm, full grain tank warning, and various sensor parameters, etc. since they are closely related to safety and operability. The number of display observations and function actuations (such as button, switch, and touch screen settings) could be obtained also from video recordings.

Number of Mistake-Correcting Actions. It was considered to be another response that could reflect the state and changes of mental workload. If the number of joystick operations seemed to reflect the cumulative effect of mental workload, the number of mistake-correcting actions was expected to show the correlated jump effect, which represented an abrupt change of mental workload level. For example, the continuous gear shift more than twice in less than one minute might imply the operator could not judge the right speed requirement immediately and temporarily entered into mental overload state.

2.3 Definition of the Moderating Factors

Definition of the six moderating factors was listed in Table 1.

Interface Complexity. In the human-machine interface of combine harvester, interfaces can be single or multiple interface, in form of instrumentations or touch screen displays. Reconfigurable displays provide greater flexibility for operators to interact

Table 1. Weighted moderating factors

Description	Original weight	Scaled moderator
MF_1 - Interface complexity (25%)		
Low: single interface, view readings	1.3	1.150
Medium: multiple interfaces, review parameters	1.0	1.000
High: multiple interfaces, review and set parameters	1.5	1.250
MF_2 - Shift (15%)		
08:00–12:00	1.0	1.000
08:00–18:00	1.1	1.075
18:00–06:00	1.2	1.150
MF_3 - Experience (20%)		
0–3 years	2.5	1.200
3–8 years	1.0	1.000
8–15 years	1.25	1.033
More than 15 years	1.5	1.067
MF_4 - Crop growing conditions (20%)		
Low: low density (<5000 kg/ha) or low moisture (<25%) or even field	1.0	1.000
Medium: medium density (5000–7000 kg/ha) or medium moisture (25–45%) or relatively even field	1.2	1.080
High: high density (>7000 kg/ha) or high moisture (>45%) or uneven field	1.5	1.200
MF_5 - Climate conditions (10%)		
Low: no wind or small wind (e.g. <8–10 m/s)	1.0	1.000
High: strong wind (e.g. >=10 m/s)	1.2	1.100
MF_6 - Operating environment (10%)		
Low: low noise and vibration, air-conditioning	1.0	1.000
High: relatively higher noise and vibration	1.2	1.100

with machine functions and monitor performance. From one single location, operators now have the ability to set and review multiple machine parameters, as well as view readings of machine and/or implement performance, which implies an increasing of operator's perceptive activities.

Shift. It is the time of day and length of shift. The combine operators are usually very busy in harvesting season, even working at night. So all possible shifts were captured in the scales for people to select the time period with the greatest overlap to the actual shift time.

Experiences. It is the level of practical experiences in number of years, related to the fitness for task.

Crop Growing Conditions. The presence of various crop density and moisture would carry different difficulties, impacting combine working speed and efficiency. And if the crops were laid (to the ground) or grown with weeds or grown on uneven land, the operators would be faced with more challenges. They needed to know the difficulty classification (high, medium, low) of crop growing conditions so as to determine control strategy. It was classified into three levels: high, medium, and low.

Climate Conditions. Climate conditions like strong wind would have considerable impact on operation complexity and control identification. It was classified into two levels: high and low.

Operating Environment. Environment factors such as noise and vibration have great influence on operator's task performance. It was also classified into two levels: high and low.

2.4 Weights

For the task elements, the expert group came to consensus that the number of display observations was much more than that of the other two elements, since the display observations were performed to monitor the running state of every part of agricultural vehicle and implied more mental work. The weights used in the model were as follows:

TE_1 - the number of mistake-correcting actions: 1
TE_2 - the number of joystick and pedal operations: 6
TE_3 - the number of display observation and function actuations: 20.

Since the factors did not contribute to the mental workload on an equivalent basis, original weights were given to represent the proportionality of these factors. When the original weights for all moderators were determined by the expert group, all the ratings were mathematically scaled in such a way that the highest moderating value in each case corresponded to the weight allocated to the moderator. For example, the maximum moderating factor for interface complexity (weight 25%) is 1.25 (which means that the mental workload value was moderated by 25%). The same logic was applied to all the six moderators in Table 1.

2.5 Mental Workload Estimation

Cameras and video recording devices were mounted on three different combine harvesters to capture the operators' activities while operating. An eight-hour shift (very short time for lunch) was recorded for each operator and analyzed by the author and the expert group to identify the tasks. And with the Eq. (1) and the weights of the task elements and moderating factors, the mental workload were worked out for nine agricultural vehicle operators, shown in Table 2.

Table 2. Predicted mental workload results

Subjects	TE_1	TE_2	TE_3	MF_1	MF_2	MF_3	MF_4	MF_5	MF_6	MW
	(1)	(6)	(20)	(25%)	(15%)	(20%)	(20%)	(10%)	(10%)	
1	15	160	300	1.0	1.075	1.2	1.0	1.0	1.0	8998
2	8	200	327	1.0	1.075	1.0	1.0	1.0	1.0	8329
3	12	180	291	1.0	1.075	1.0	1.0	1.0	1.0	7430
4	6	220	321	1.25	1.075	1.2	1.0	1.0	1.1	13739
5	10	245	338	1.25	1.075	1.0	1.0	1.0	1.1	12180
6	8	150	286	1.25	1.075	1.0	1.0	1.0	1.1	9797
7	11	213	332	1.15	1.075	1.2	1.0	1.0	1.0	11763
8	6	192	278	1.15	1.075	1.0	1.0	1.0	1.0	8305
9	4	176	319	1.15	1.075	1.0	1.0	1.0	1.0	9198

3 Verification Study: Mental Workload Measurement by Task Analysis

The prediction model was developed but its validity was still remained to be proved. The purpose of the verification study was to use another independent measuring approach of mental workload applied in agricultural vehicle operations to find support.

Many factors make the task of assessing operator's workload difficult. First, many of the tasks involve cognitive activity that is not directly observable. Second, the timeframe of task performance is often unpredictable due to the delays that may exist between event onset and the initiation of a response. Finally, assessment is aggravated by the fact that several tasks are often performed concurrently during a given time period [15]. Given these problems, a modified version of an observational methodology and analytical technique based on the Task Analysis Workload [16, 17], developed originally for the assessment of military helicopter flight crews, was used effectively in agricultural vehicle environment and is referred to as the mTA.

3.1 Method Description

The mTA is a task-oriented approach that assumes multiple and simultaneous tasks occur through time and that these tasks might vary in their demand on the operator's performance resources. The mTA treats workload as the sum of the difficulty of all concurrent tasks for each minute of an observation. Two operators may handle an equal number of field crops, yet significant differences may exist in mental workload across a shift if one of the operator's activities take place within a short period of time while the other operator's load is spread evenly over time. The mTA also takes task difficulty into account. Two tasks may be of equal time duration, yet one task may call on more resources of listening, watching, thinking, or overtly acting than the other. The mTA used in this study refers to these resources as the perceptive (combination of auditory with visual), cognitive and psychomotor channels. Any task performed by an operator can be broken down into these components. This method calculates workload by summing the loads for each of the individual channels across all tasks.

The agricultural vehicle operator's activities in an eight-hour shift were recorded by video. Then the expert-developed criteria were applied to create mTA scores by making judgments about the level of effort required from each of the three resource channels for each task.

Table 3 listed the tasks hierarchy defined to describe the agricultural vehicle operator's job according to discussions with an expert group (experienced operators, machine owners and technicians in university). There are mainly two kinds of tasks. The primary are the fundamental tasks, which are either continuous or scheduled at a time decided by the operator, e.g. visual monitoring, watching for unanticipated events on the informational displays, speed adjustment, header height adjustment, etc. This kind of tasks do not require an immediate response. The other is the emergency task, which is unexpected and probably require an immediate response (e.g. grain outcome blocked, high water temperature, abnormal noise, etc.)

Table 3. Task analysis hierarchy of combine harvesting operation

Tasks	Perceptive behavior	mw_{perc}	Cognitive behavior	mw_{cog}	Psychomotor behavior	mw_{psym}
1.1 Overall check	Appearance detection	2	check if machine is in good maintenance	2	Head and eye movement	1
1.2 Preparations before start	Observe circumstance and check if there is any risk	2	Make sure it is right and safe	2	Head and eye movement, steering wheel adjustment, alarm	2
1.3 Initialization of working device	Inspect device position	1	Make sure the header and reel in right position	2	Operate handles to adjust device height	3
1.4 Gear selection	Detect current gear position	1	Determine required action	1	Finish gear shift	2
1.5 Assessment of requirements	Inspect field situations	3	Judge soil moisture and crop density to make decision	4	Head and eye movement	1
1.6 Header adjustment	Monitor crop situations	7	Determine required action	4	Operate header handle for adjustment	3
1.7 Reel adjustment	Monitor crop situations	7	Determine required action	4	Operate reel control handle for adjustment	3
1.8 Route planning	Observe field conditions	4	Determine working route	3	Steering and braking, lift device when turning	4

(*continued*)

Table 3. (*continued*)

Tasks	Perceptive behavior	mw_{perc}	Cognitive behavior	mw_{cog}	Psychomotor behavior	mw_{psym}
1.9 Running pattern selection	Observe field conditions	4	Determine which one to choose, manual mode, automatic control mode or autopiloting mode	2	Press corresponding buttons	2
1.10 Monitor running state	Monitor performance indicators on the tablet PC (e.g. water temperature, transmission), discriminate abnormal noises	6	Determine if there is failure, what and where it is	6	Stop operation	5
1.11 Change feeding speed	Monitor the output state	5	Determine required action	4	Operate touch screen and buttons	5
1.12 Speed adjustment	Monitor crop growing conditions	5	Determine required action	4	Control throttle pedal	3
1.13 Inform fellow workers	Observe circumstances	2	Determine if there is any risk	3	Alarm, buttons and handles to start	3
1.14 Unloading	Inspect unloading process	2	Determine required action	4	Unloading, head and eye movement	4

3.2 Rating Scales

Rating scales have been developed for each component, which could provide a relative rating of the degree to which each resource component is used [18–20]. Table 4 includes the rating scales and each scale value has a text description as an anchoring statement. The higher the scale value the greater the degree of use of the resource component. A little modification has been made to the original rating scales. Equal intervals from 0 to 7 are used for easy understanding. Based on this, each task was scored by expert group judgments on a scale from zero to seven according to its contribution to workload in each of the three information-processing channels, perceptive (mw_{perc}), cognitive (mw_{cog}) and psychomotor (mw_{psym}), shown in Table 3. The mTA finally works out a graph that provides a way of identifying peak workload in one

Table 4. Mental workload rating scales

Scale value	Descriptor	Example
Perceptive (visual and auditory)		
0	No visual/auditory activity	
1	Visually register/detect (detect occurrence of image)	Detect alarm light, sound horn
	Detect/register sound (detect occurrence of sound)	
2	Visually discriminate (detect visual differences)	Discriminate signals, alarm
	Orient to sound (general orientation/attention)	
3	Visually inspect/check (discrete inspection/static condition)	Check fuel gauge, music listening
	Orient to sound (selective orientation/attention)	
4	Visually locate/align (selective orientation)	Locate target on the fields, orientation of alarm
	Verify auditory feedback (detect anticipated sound)	
5	Visually track/follow (maintain orientation)	Track potential risks
	Interpret semantic content (speech)	
6	Visually read (symbol)	Read unfamiliar signs or displays
	Discriminate sound features (detect auditory differences)	
7	Visually scan/search/monitor (continuous/serial inspection, multiple conditions)	Monitor the quality of field operation
	Interpret sound patterns (pulse rates, etc.)	
Cognitive		
0	No cognitive activity	
1	Automatic (simple association)	Need braking, keep in line
2	Alternative selection	Decide if stop
3	Sign/signal recognition	Recognize field conditions, make if-then decision
4	Evaluation/judgment (consider single aspect)	Determine header height or gear shift
5	Encoding/decoding, recall	Recall unfamiliar displays
6	Evaluation/judgment (consider several aspects)	Determine required operation according to field and crop conditions
7	Estimation. calculation, conversion	Estimate area or yields

(*continued*)

Table 4. (*continued*)

Scale value	Descriptor	Example
Psychomotor		
0	No psychomotor activity	
1	Simple and automatic response	Wheel steering and head movement to keep straight
2	Discrete Actuation (button, toggle, trigger)	Press button, braking
3	Continuous adjustive (flight control, sensor control)	Adjust acceleration, steering to finish turning
4	Manipulative	Adjust header height
5	Discrete adjustive (rotary, vertical thumbwheel, lever position)	Gear shift, use GPS
6	Symbolic production (writing)	Record
7	Serial discrete manipulation	keyboard entries

or more of the information-processing channels. Where no more value existed, the median value from the range of responses was used.

3.3 Task Data Analysis

The task analysis of activities performed by agricultural vehicle operators was conducted so as to identify the task related mental workload. The hierarchy of tasks that an agricultural vehicle operator usually performs is listed in Table 3. Information processing might have two ways, parallel and serial, resulting in different numerical analysis of mental workload. If taking parallel processing as premise, the mental workload over some period is the cumulative value of all the mental workload that all the tasks have caused on this period. An accumulated value is used to represent this situation. Although this approach makes numerical values no sense of meanings, it better shows the information processing density and strength under multiple tasks. If it is premised as serial processing, the cumulative value can be averaged according to the number of tasks on any period so as to obtain the average mental workload. This approach retains the meaning of the original numerical value, but it is insensitive to the attributes of multiple tasks. Generally, the actual information processing would probably stay between the above two approaches, and the mental workload would also be within the interval defined by the accumulated value and mean value.

The cumulative value of the mental workload of all the tasks were obtained. First define MW_{perc} as the cumulative value of the mental workload caused by perceptive behaviors at the same time, MW_{cog} as the one caused by cognitive behaviors and MW_{psym} as the one caused by psychomotor behaviors, then the mental workload at different time period can be put into comparison. Define the whole mental workload vector modulus $|W|$ as

$$|W| = \sqrt{(W_{perc}^2 + W_{cog}^2 + W_{psym}^2} \\ = \sqrt{\left(\sum_{i=1}^{n} mw_{perci}\right)^2 + \left(\sum_{i=1}^{n} mw_{cogi}\right)^2 + \left(\sum_{i=1}^{n} mw_{psymi}\right)^2} \tag{2}$$

where mw_{perc}, mw_{cog}, mw_{psym} is the rating value of the mental workload of perceptive, cognitive or psychomotor channel respectively. And n is the recording times.

Figure 1 illustrates an example of the source channels workload. In this example, it showed a low level of activity at the beginning and a period of high activity appeared about two hours later. The expert group revealed that during the busy period they had to deal with a large number of observations, handle operations, failure recognition and judgments since the machine entered into a “heated” state. Especially for the perceptive (visual and auditory) and cognitive channels, the workload is higher. A significant amount of time spent on displays monitoring, frequent adjustments of working devices and processing presented information would account for the heavy workload.

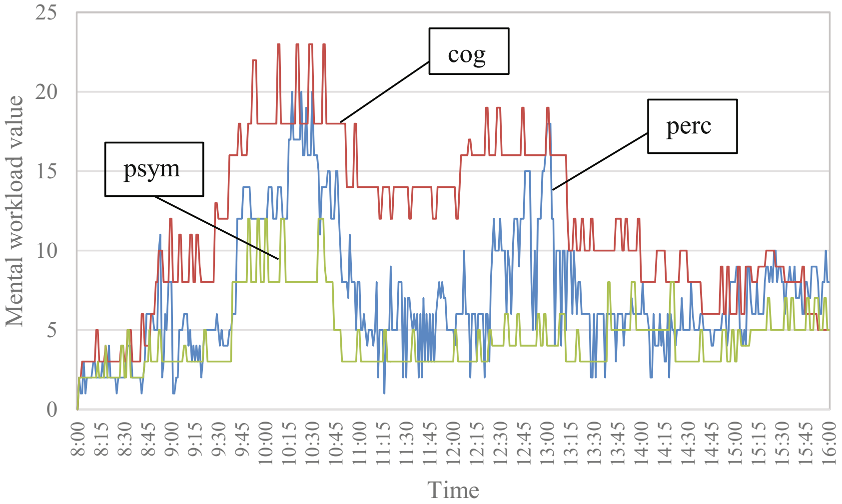


Fig. 1. Workload of three source channels

Finally, subjective workload measurement was conducted as supplements, referring to the agricultural vehicle operator’s perception of both: (1) the demands of their work in terms of difficulty, complexity and time pressure, and (2) the effort they need to spend to meet those demands. It is often assessed using self-report rating scales during or immediately following task performance. Originally, the NASA-TLX was the first thought for this study. However, considering both the practical intense working environment and less education background of agricultural operators in China, it is time consuming and too hard for the operators to fully understand (most of whom are middle-school level). The idea of NASA-TLX was therefore abandoned. And an overall workload scale ranging from 0 to 100 was finally utilized to evaluate their perceived level of workload once every two hours.

4 Result and Discussion

Agricultural vehicle operator's mental workload was examined using two separate sources of data. These data were collected through naturalistic observation of the operators performing their job, including three different combine harvesters and 18 operators, one half for the model development and the other half for the validation study. The test agricultural vehicles are all large ones but equipped with different human-machine interaction configurations, that is, different patterns of information presentation. For example, combine No. 3 has been equipped with a moveable multi-function joystick with multiple primary controls, making it easier and flexible to operate. Combine No. 1 is equipped with fixed-type multi-function handle with primary controls while combine No. 2 only has traditional separate handles. And their secondary control panel positions are also different.

The mental workload results of the three test agricultural vehicles obtained by the task analysis was listed in Table 5. In comparison with the predicted value, the ratio between them is 1.19, 1.39 and 1.33 respectively. The close proportion implies the validity of the prediction model of mental workload.

Table 5. Mental workload rating values

	Agricultural vehicles No.		
Mental workload	No. 1	No. 2	No. 3
Perceptive	3453	4221	3647
Cognitive	5577	6439	5792
Psychomotor	2174	3684	2596
mTA workload	6910	8535	7320
Predicted value (mean)	8252	11905	9755
Predicted/mTA	1.19	1.39	1.33

An ANOVA analysis was computed comparing separately the numbers of joystick and pedal operations (TE_2) and the number of display observation and function actuations (TE_3) among the study farming vehicles (with all the operators). Significant differences were found among these vehicles by interface complexity ($p < 0.01$). Independent t-tests were used for comparison of experienced operators separated by interface complexity. Table 6 examined the comparison of average number of TE_2 and TE_3 every two hours, showing the task elements distribution and frequency. A more conservative significance level, $p < 0.01$, is used to control for type error. It was combine No. 2 that was reported to have greatest number of operations and observations. The number of operations and observations handled at the combine No. 1 and No. 3 every two hours across an eight-hour period, are more evenly distributed than combine No. 2. This was expected as the two vehicles have a higher level of information automation, which save operator's sources. The number of joystick and pedal operations also fluctuated since crops on the fields were somewhere in complicated situations such as being laid and weeds. And the comparison between the operators of different years of experiences showed significant differences in number of operations and observations. This might be because the operators with 3–8 years of experience are more skilled in time-saving performance and can quickly react to stress and emergency situations, while the operators with 0–3 years of experience seemed not to be used to frequent switch of actions.

Table 6. Comparison of average number of TE2 and TE3

Interface complexity	Experience	2 h		4 h		6 h		8 h	
		TE_2	TE_3	TE_2	TE_3	TE_2	TE_3	TE_2	TE_3
Low(ComNo. 2)	0–3 years	58.0	90.4	67.1	96.5	**52.1***	88.3	40.0	61.2
	3–8 years	**67.3***	97.2	77.2	99.3	56.8	91.2	45.7	**68.9***
Medium(ComNo. 1)	0–3 years	40.9	75.2	52.3	83.2	38.3	72.5	**30.8***	64.3
	3–8 years	45.3	81.1	59.6	**88.3***	44.1	76.2	38.6	68.2
High(ComNo. 3)	0–3 years	34.0	63.6	**45.7***	70.5	32.3	61.6	28.3	52.5
	3–8years	46.2	72.4	51.9	76.7	39.5	66.3	**34.5***	**62.2***

*Indicated $p < 0.01$ level of significance.

ANOVA was used to assess the effects of interface complexity on subjective workload ($p < 0.01$). The interface complexity of low level (combine No. 2) began with higher subjective ratings of workload that the other two. This might be due to the anticipation of the hard work they were going to handle with a relative low level of machine, heavy noise and vibration, and displays or controls uneasy to use, etc. They fell in the feelings all day long until they came into state of mental fatigue after about five hours. This fitted with the workload data based on the number of operations and observations, which started to decrease after about the same amount of time. Generally, subjective workload seems to closely follow the operation times and intensity.

Correlations were conducted between the number of operations and observations and perceived workload. Subjective mental workload was mildly correlated with the statistics of operations and observations ($R2 = 0.31$) except TE_1. These associations were found to be significant and thus reliable. Figure 2 presented the task elements data along with perceived workload over eight hours averaging the three combines. The graph illustrated a decline in the number of task elements while the subjective values gradually increased. This seems following the nature of people work feelings after a long time.

Although the task elements seemed to be in close relationship with the subjective ratings, they were not particularly good indicators of mental workload when presented individually. And when they were combined and weighted, it would lead to better prediction. The task analysis method did provide some useful information, evaluating and demonstrating the workload distribution over time. However, its shortness was the dependence on the task hierarchy and time estimation (especially time for cognitive activities). Nevertheless, the extent of the task decomposition in the task analysis is

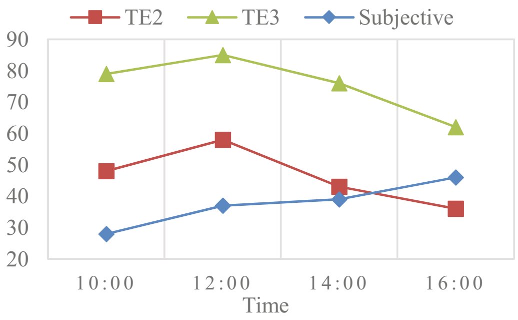


Fig. 2. Relation of task elements and subjective ratings

more appropriate to mental workload analysis, since it is neither too macro to reflect operator's cognitive process, nor too microscopic to simplify the task description. It was also important to notice that the subjective workload scores did not reach excessive high level. A possible reason could be that there were only a few instances like blocked failure or abnormal noise phenomenon that made mental workload went far beyond average. Or it was also possible that the operators did not fully understand the meaning of the investigation since they generally received less education.

There are a number of ways that operators divide their attention to manage their information processing besides temporal division. A popular method of information division is by input modality, such as visually, auditory or tactile. Because the amount of attention that an operator can use to focus on a given input modality is limited, it is possible to allocate greater amounts of attention if you split the presentation over different sensory inputs. For example, warnings that require immediate attention also have associated auditory feedback in the form of buzzers, bells, or beeps. There are also some samples of current technologies that allow operators to performing multitasks more efficiently, such as GPS auto guidance, automatic height control, and Automatic crop settings, etc.3, which seem to be so good ways to free the mental resources of agricultural vehicle operators. Currently, there are still a lot of potential demands for information automation and information management regarding farming operations.

5 Conclusions

Agricultural vehicle operators are presented with even more information regarding system performance, faced with more challenges and risks of overload. The human factor expert should regulate the extent of information presentation to an appropriate level. Regarding this, prediction of mental workload became a critical issue. In this study, a mental workload prediction model was proposed to provide an objective method for the assessment and prediction of operator's mental workload in multiple information presentation environment of agricultural vehicles. An expert group comprising experienced operators and technical staff was involved in identifying task factors and assigning weights for task and moderating factors. The model consisted of three task elements and six moderating factors, each with a different weight in terms of its contribution to overall mental workload. A modified task workload analysis measurement was used to validate the proposed model, and it also proved valuable for identifying and evaluating the mental resources components of workload for a given task over time. The task elements were moderately correlated with the subjective workload but combined to lead to better prediction. Next, improved information automation and information management will be considered as to lower operator's mental workload level.

Acknowledgements. This study was financially supported by the National Key Technology Support Program "Large Agricultural Machinery Digital Manufacturing and Automatic Production" (2012BAF07B01) and Chinese Universities Scientific Fund (2013XJ002).

References

1. Borghini, G., Astolfi, L., Vecchiato, G., et al.: Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. **44**, 58–75 (2014)
2. Di Stasi, L.L., Álvarez-Valbuena, V., Cañas, J.J., et al.: Risk behaviour and mental workload: multimodal assessment techniques applied to motorbike riding simulation. Transp. Res. Part F Traffic Psychol. Behav. **12**(5), 361–370 (2009)
3. ASABE Distinguished Lecture Series, vol. 36, pp. 1–8 (2012)
4. Tobaruela, G., Schuster, W., Majumdar, A., et al.: A method to estimate air traffic controller mental workload based on traffic clearances. J. Air Transp. Manag. **39**, 59–71 (2014)
5. Wickens, C.D.: Multiple resources and mental workload. Hum. Factors J. Hum. Factors Ergon. Soc. **50**(3), 449–458 (2008)
6. Liang, S.F.M., Rau, C.L., Tsai, P.F., et al.: Validation of a task demand measure for predicting mental workloads of physical therapists. Int. J. Indus. Ergon. **44**(5), 747–752 (2014)
7. Stuiver, A., Brookhuis, K.A., De, W.D., et al.: Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. Int. J. Psychophysiol. Official J. Int. Organ. Psychophysiol. **92**(1), 35–41 (2014)
8. Ayaz, H., Shewokis, P.A., Bunce, S., et al.: Optical brain monitoring for operator training and mental workload assessment. Neuroimage **59**(1), 36–47 (2012)
9. Miller, M.W., Rietschel, J.C., Mcdonald, C.G., et al.: A novel approach to the physiological measurement of mental workload. Int. J. Psychophysiol. Official J. Int. Organ. Psychophysiol. **80**(1), 75–78 (2011)
10. Didomenico, A., Nussbaum, M.A.: Effects of different physical workload parameters on mental workload and performance. Int. J. Indus. Ergon. **41**(3), 255–260 (2011)
11. Kumashiro, M.: Evaluation of Human Work. Taylor and Francis, London (2005)
12. Shelton-Rayner, G.K., Mian, R., Chandler, S., et al.: Leukocyte responsiveness, a quantitative assay for subjective mental workload. Int. J. Indus. Ergon. **42**(1), 25–33 (2012)
13. Averty, P., Collet, C., Dittmar, A., et al.: Mental workload in air traffic control: an index constructed from field tests. Aviat. Space Environ. Med. **75**(4), 333–341 (2004)
14. Nunes, A., Mogford, R.H.: Identifying controller strategies that support the 'Picture'. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **47**, 71–75 (2003)
15. Bin, X., Salvendy, G.: Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. Work Stress **14**(1), 74–99 (2000)

16. Popkin, S., Gertler, J., Reinach, S.: Technical report No. DOT/FRA/ORD-01-08, Department of Transportation, Washington, DC, U.S.A. (2001)
17. Keller, J.: Human performance modeling for discrete-event simulation: workload. In: Winter Simulation Conference: Exploring New Frontiers, pp. 157–162. IEEE Xplore, San Diego (2002)
18. Colle, H.A., Reid, G.B.: Estimating a Mental workload redline in a simulated air-to-ground combat mission. Int. J. Aviat. Psychol. **15**(4), 303–319 (2005)
19. Jo, S., Myung, R., Yoon, D.: Quantitative prediction of mental workload with the ACT-R cognitive architecture. Int. J. Indus. Ergon. **42**(4), 359–370 (2012)
20. Azadeh, A., Rouzbahman, M., Saberi, M., et al.: Improved prediction of mental workload versus HSE and ergonomics factors by an adaptive intelligent algorithm. Saf. Sci. **58**(10), 59–75 (2013)

The Evaluation of Pilot's First Fixation and Response Time to Different Design of Alerting Messages

Wen-Chin Li[1(✉)], Jiaqi Cao[2], Jr-Hung Lin[3], Graham Braithwaite[1], and Matthew Greaves[1]

[1] Safety and Accident Investigation Center, Cranfield University, Cranfield, UK
wenchin.li@cranfield.ac.uk
[2] Commercial Aircraft Corporation of China (COMAC), Shanghai, China
[3] Office of Institutional Research, National Central University, Taoyuan, Taiwan

Abstract. Current research investigates the limitation of current cockpit design for Crew Alerting System (CAS) and Quick Reference Handbook (QRH), and proposes a potential design solution that might enhance crew performance. Using eye tracking device, an experiment was conducted based on Flight Simulator Software. Objective eye movement data as long as subjective feedback from participants were collected to evaluate the design. 24 participants experienced 4 scenarios with both traditional design and integrated design. Results demonstrated that participants never made error in locating the instructions by integrated design. It is due to the obvious fact that the instructions of integrated design will present itself, hence decrease the chance of executing the wrong NNC by eliminating the degree of choices. On the other hand, 75% of the participants experienced at least one error in finding the correct instruction, which is probably due to the limitation of QRH itself due to options too close in formatting which causes clicking the wrong page accidently. Furthermore, a trend was identified by eye movement patterns for longer fixation duration, smaller saccade amplitude, and less time fixated on instruction area occur in integrated design. Current research found that integrated design of CAS and QRH is superior in acquiring accurate solutions for emergent situations and processing the information presented compared with traditional design.

Keywords: Cockpit design · Crew Alerting System · Eye movement · Human-computer interaction · Quick Reference Handbook

1 Introduction

Commercial aircrafts equipped with Engine indication and Crew Alerting System (EICAS) are able to monitor real-time aircraft health and provide Crew Alerting System (CAS) messages which indicate abnormal situations to flight crews. With the support of Quick Reference Handbook (QRH), in which contains Non-Normal Checklist (NNC), flight crews will manage to find solutions for associated situations. However, there is no link or coordination between these two systems. The integrated CAS and QRH is very intuitive to fit the principle of human-centered design. It is

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 21–31, 2017.
DOI: 10.1007/978-3-319-58472-0_2

obvious that once flight crews are alerted by CAS of an emergency, they would like to know how to deal with it immediately; hence the systems present the checklists in QRH to provide pilot relevant information. Kirsh (1995) proposed that contracting the number of decisions users must make is an indication of a good design. The task of handling emergency situations can be simplified as a single cognitive process, using cues from external environment to search for a specific procedure. If new design is able to reduce search space and time, crews will have more time solving the problems in external world. It is possible to make representations of QRH more active in order to help crews see what is most relevant to deciding what to do next (Hollan et al. 2000). One rule in Proximity Compatibility Principle (Wickens and Carswell 1995) in terms of spatial proximity can be used for this integration, which suggests two pieces of information that need to be integrated on a cluttered display should be placed in close spatial proximity. Electronic Checklist (ECL) developed for Boeing 777 is a close to ideal model of this integrated design.

QRH is designed with the intention of allowing flight crews to minimize the need for a lot of effortful analysis when time may be limited and workload is high (Burian 2004), as long as the expectation that they will correctly interpret the cues available to them and find the appropriate checklist to solve problems (Burian 2006). However, it is very important for QRH designers to be aware of how human will behave in emergent situations. Currently there are two types of QRH, paper or electronic form. Nowadays, paper QRH is less frequently used due to the large amount of effort needed to retrieve the correct NNC and the placement of electronic QRH in modern glass cockpit. However, Civil Aviation Authority (2006) requires that paper QRH should be still onboard of aircraft in case of system failure in CAP 676, which is the guideline on design, presentation and use of paper form NNC. NNC are used when the aircraft is experiencing one or more system failures (Boorman 2001), which will be displayed automatically upon detection of the related alert, so ECL's provide a means to accomplish checklists with a reduction of searching time and a reduction in the possibility of crew error (Federal Aviation Administration 1996). A slight downside of Boeing's ECL is that the NNC is presented in another display beneath EICAS display and a few steps still need to be done before reaching the checklist (Fig. 1).

Fig. 1. Position of EICAS display and NNC

The application of eye-tracking in the study of flight simulation is promising as it provides direct feedback, which could diagnose potential factors that impact upon pilot attention and situation awareness on the flight deck (Robinski and Stein 2013). As suggested by Jones and Endsley (1996), over 75% of pilot errors are caused by perceptual failures, so it is interesting to study visual information processing in the form of eye movements and gain insight into the perceptual qualities. By applying eye tracking device to evaluate the relation between human and design, that eye movement, including gaze, fixation and saccade, is controlled by ongoing cognitive processing, so that it is possible to analyze human behavior by examining eye movement. This assumption has been validated by many previous researchers in reading (Carroll and Slowiaczek 1986; Just and Carpenter 1980; Rayner and Pollatsek 1992; Rayner et al. 1989), cognitive tasks (Ahlstrom and Friedman-Berg 2006; Salvucci and Goldberg 2000), information processing tasks (Rayner 1998), scanning behavior (Allsop and Gray 2014), interface evaluation (Goldberg and Kotval 1999), and Human Computer Interaction (Yu et al. 2014). Pilots have to manually go through the QRH and identify the relevant checklists for current situation then act accordingly. It might increase respond time for searching suitable information hence degrade crew performance for emergent events. By examining the visual characteristic of the integration CAS and QRH, it might speed up pilots' reaction time under urgent situations. Therefore, this research will evaluate the effectiveness of proposed new cockpit design and whether or not it will impact to pilots' performance accordingly. However, Yu et al. (2014) argued that most eye tracking experiments are performed in the laboratory and restrict subjects' head and body motion, which differs from the naturalistic setting and limits the application. In this research, a portable glass-like eye tracker is introduced to counter these effects, which allows subjects perform as usual.

2 Method

2.1 Participants

The study involved twenty-four participants (ages range from 21 to 50 years old, M = 27.5, SD = 6.9) consisting of airlines pilots, PPL pilots, Engineers in aviation industry and Professionals in aviation domain (flight experience between 0 and 3,000 h, M = 154.1, SD = 608.3). All participants are reported normal or corrected vision with contact lens. As data was gathered from human participants, a research proposal was submitted to the Cranfield University Research Ethics System (CURES) for ethical approval. Ethical approval was granted for the research prior to starting the experiment by the CURES (CURES ID: 1773).

2.2 Apparatus

2.2.1 Simulator Set Up

The experiment was conducted in a controlled environment (lighting condition and disturb-avoidance), which contained one laptop with Microsoft Windows OS (Operating System) for Microsoft Flight Simulator X (FSX), one joystick and one control lever to assist FSX, one laptop with MAC OS to record eye movement data, one display screen and one laptop for playing the videos of emergent scenarios, and one IPAD to serve as electronic version of Boeing QRH (Fig. 2a).

Fig. 2. (a) Image of Flight Simulator, (b) Pupil Pro eye tracker

2.2.2 Eye Tracking Device

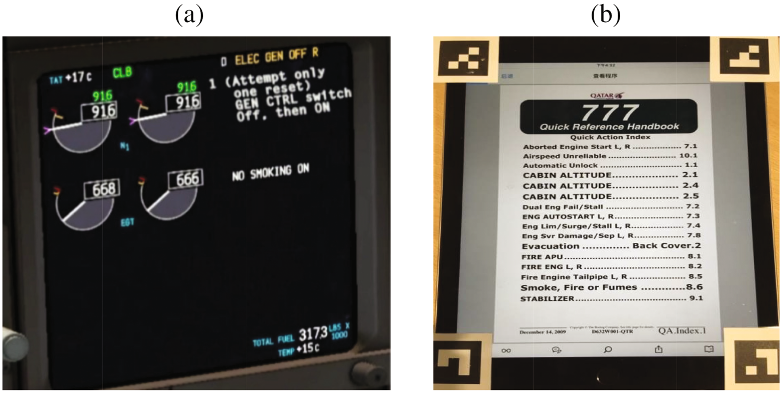
To capture eye movement data while participants were watching the videos of emergent scenarios, a Pupil Labs "Pupil Pro" eye tracking device was used, which carry one Eye Camera with 640 × 480 at 60 fps resolution and one World Camera with 1280 × 720 at 60 fps resolution. Unlike conventional eye tracking devices, the Pupil Pro is a glasses-like eye tracker which is very portable and easy to be configured to multiple test environments. The data can be record and process in a computer (an Apple Mac Book Pro Laptop was used for this experiment) via USB interface. This device has two cameras, which will be synchronized after calibration. The 'World Camera' is mounted on the right top of glasses showing the orientation and view of the wearer's head. The World Camera used in this experiment is a high speed camera with 100-degree diagonal lens. A second camera, the Eye Camera, is mounted offset right and low which is adjustable to suit different wearer's facial layout and track their pupil data accordingly (Fig. 2b).

2.3 Research Design

To design the emergent scenarios for experiment, FSX software was used to record 4 flight segments of the same flight path with fixed cockpit view by 2 different design layouts for each scenario which are current design and integrated design. In current design, instructions for each emergent event are located in QRH by electronic form of IPAD. In integrated design, the same instructions are listed directly on EICAS display under the CAS message (Fig. 3a). The instructions in both electronic versions of QRH and EICAS display are modified to the same context in this experiment. These 4 scenarios as following.

2.3.1 Left Engine Fire at 46 s

After the aircraft takes off from Heathrow airport, it takes a gentle turn heading 137 and initiates a climb to follow the flight path to the waypoint of GURLU with autopilot



D ELEC GEN OFF R

1 (Attempt only
one reset)
GEN CTRL switch
Off, then ON

NO SMOKING ON

QATAR

777

Quick Reference Handbook

Quick Action Index

Aborted Engine Start L, R	7.1
Airspeed Unreliable	10.1
Automatic Unlock	1.1
CABIN ALTITUDE	2.1
CABIN ALTITUDE	2.4
CABIN ALTITUDE	2.5
Dual Eng Fail/Stall	7.2
ENG AUTOSTART L, R	7.3
Eng Lim/Surge/Stall L, R	7.4
Eng Svr Damage/Sep L, R	7.8
Evacuation	Back Cover.2
FIRE APU	8.1
FIRE ENG L, R	8.2
Fire Engine Tailpipe L, R	8.5
Smoke, Fire or Fumes	8.6
STABILIZER	9.1

December 14, 2009 D632W001-QTR QA.Index.1

Fig. 3. (a) Integrated design of CAS and QRH, (b) Electric form of QRH

engaged. When it reaches 4,200 feet at 254 knots, red warning CAS message, FIRE ENG L, appears on the EICAS display with an audio alert and a red warning light.

2.3.2 Cabin Altitude Warning Occurs at 61 s

The aircraft is in climb phase, heading 137, on the flight path to the waypoint of GURLU with autopilot engaged. When it reaches 14,720 feet with airspeed of 320 knots, red warning CAS message, CABIN ALTITUDE, appears on the EICAS display with an audio alert and a red warning light. The emergency event is caused by left cabin door malfunction.

2.3.3 Right Engine is Shut Down at 26 s

The aircraft is in cruise phase, heading 137, airspeed 318 knots, altitude 10,000 feet, on the flight path to the waypoint of GURLU with autopilot engaged. After a sudden engine noise decreased, engine indication of N1 and EGT drop dramatically, then amber caution CAS message, ENG FAIL R, appears on the EICAS display with an audio alert and an amber caution light.

2.3.4 Right Engine Generator Fails at 40 s

The aircraft is in climb phase, heading 137, on the flight path to the waypoint of GURLU with autopilot engaged. When it reaches 3,800 feet with airspeed of 251 knots, firstly, amber caution CAS message, FUEL PUMP CENTER R, appears on the EICAS display, shortly after, amber caution CAS message, ELEC GEN OFF R, appears on the EICAS display with NO audio alert and NO caution light.

2.4 Data Collection and Analysis Processes

All participants were provided an instruction sheet of the experiment design for briefing in the same way. Then a detailed demonstration of the task with 2 screenshot pictures of examples of emergent scenarios were performed, which would assist participants acquire the basic knowledge regarding current design and integrated design, the color of the CAS message, what to do after see the CAS message by call out, then find the instruction for the CAS message in QRH or under the CAS message; read the instructions, then go back monitoring PFD until the scenario is ended. Before calibrate eye tracker, participant was asked to sit comfortably in order to maintain the same seating position during the whole experiment process. After that, the world camera view was checked via asking the participant to look at the screen and look at the IPAD, to ensure the world camera would capture all the components. Afterwards, the eye camera view was also manipulated through adjusting the eye camera mount to keep the right pupil of the participant within the correct view (Fig. 4). Eleven points of calibration which cover different space of the outside views were conducted. Following the calibration, the participant performed the experiment by monitoring 8 developing scenarios in the random sequence. Random table was used to reduce practise effect and random error due to individual differences (Mitchell and Jolley 2012).

Interface Size 1.00

Mode ROI

Click and drag on the blue circles to adjust the region of interest. The region should be as small as possible, but large enough to capture all pupil movements.

Detection method C++ 2d

Fig. 4. Eye camera adjustment view and manual calibration process

All participants' eye movement data are analyzed in the same 30-seconds of time period, which starts from the emergency event occurs and ends 30 s after. This time period contains the most significant eye movement data which can reflect to cognitive processes of participants interacted with non-normal flight operations. The data analysis process is divided into 4 sectors: (1) define Area of Interests (AOIs); (2) select time period; (3) export and organize data; and (4) analyze data. AOIs are the areas which contain critical eye movement information for analysis. In this case, there are three defined AOIs including, (1) CAS message area on EICAS; (2) Instructions on QRH; and (3) Instructions under CAS. There are two independent variables in this research. The first one is the design layout of EICAS, whether it is traditional design which instructions for emergent situations should be found in QRH, or integrated design which instructions are integrated with the CAS message. The second independent variable is four different scenarios. The dependent variables are time of first fixation placed on CAS message (T1), time of first fixation placed on instructions (T2), Task Completion Time (T3), percentage of fixation on instruction AOI, mean fixation duration (MFD) on instruction AOI, mean saccade amplitude (MSA) on instruction AOI, number of fixation on different displays and perceived mental Workload (PMW). For each data, test of normality is conducted, which determine whether the sample can represent general population. T1, T2, T3, percentage of fixation, MFD, MSA and PMW will be analyzed against two design layouts of EICAS and four flight scenarios, using two-way repeated measure Analysis of Variance (ANOVA).

3 Results and Discussions

3.1 Response Time of First Fixation on CAS Message

Response time is analyzed based on the information process of participants during the task performance, which starts from noticing the emergent event (time of first fixation on CAS message, T1) to locating solution for the event (time of first fixation on instructions, T2), then finally time of completing the task (T3). Mauchly's test shows that the assumption of sphericity is met for the effect of scenario, $\chi^2(5) = 4.34$, $p > 0.05$. For the interaction between design and scenario, the assumption of sphericity is violated, $\chi2(5) = 11.06$, $p < 0.05$, therefore degrees of freedom were corrected by using Huynh-Feldt estimated of sphericity ($\varepsilon = 0.82$) (Field 2013). The result indicated no significant main effect of design layouts, F (1, 23) = 0.105, p = 0.748, $\eta_p^2 = 0.005$, suggesting no significant difference between two design layouts on the time of first fixation on CAS message. However, there is a significant main effect of scenario, F (3, 69) = 1.639, $p < 0.001$, $\eta_p^2 = 0.214$. Further post-hoc comparisons showed that participant's first fixation of scenario 3 (M = 0.57) is significant faster than that of scenario 2 (M = 0.93) and scenario 4 (M = 0.96). There is no significant interaction between the design and scenario, F (2.76, 63.40) = 0.192, p = 0.887.

3.2 Response Time of First Fixation on Instruction

There is a significant main effect of design, F (1, 23) = 143.00, $p < 0.001$, $\eta_p^2 = 0.861$. The time of first fixation on instruction AOI of integrated design (M = 6.02) is significant less than that of the traditional design (M = 17.14). There is a significant main effect of scenario, F (2.16, 49.63) = 47.81, p < 0.001, $\eta_p^2 = 0.491$. Further post-hoc comparisons showed that participant's first fixation of scenario 1 (M = 7.18) is significant less than scenario 3 (M = 14.64) and scenario 4 (M = 15.87). Participant's first fixation of scenario 2 (M = 8.62, SD = 7.55) is significant less than scenario 3 (M = 14.64) and scenario 4 (M = 15.87). The effect suggests that time of first fixation on instruction is different among 4 scenarios. There is a significant interaction between designs and scenarios, F (3, 69) = 22.11, $p < 0.001$, $\eta_p^2 = 0.243$. The Fig. 5a shows the average T2 of different scenarios for each design. Generally speaking, participants spent less time finding the instructions for emergent events (T2) in integrated design than in traditional design. As the level of scenarios increases, average T2 increases as well, which might suggest that the instructions for emergent events in latter scenarios are more difficult to be found than former ones. The reason for this might be that Engine Generator Fail and Engine Shut Down are less experienced or trained by crews than Engine Fire and Cabin Altitude, hence it takes flight crews more time to locate the instruction and find the solutions to solve the emergent situations.

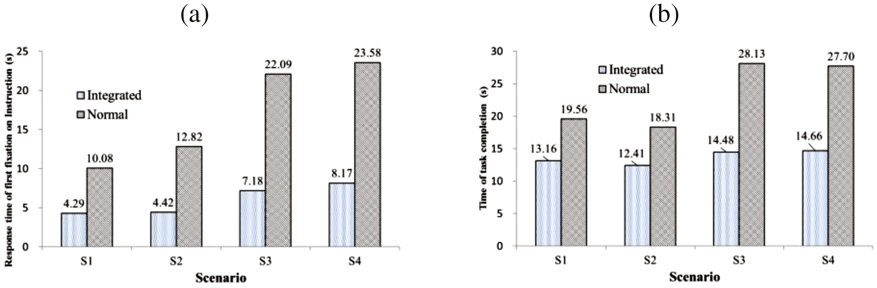


Fig. 5. (a) T2 of 4 scenarios in 2 design layouts, (b) T3 of 4 scenarios in 2 design layouts

3.3 Time of Task Completion

The result of within-subject AVOVA for T3 suggests that there is a significant main effect of design, F (1, 23) = 175.40, $p < 0.001$, $\eta_p^2 = 0.884$. The completion time on integrated design (M = 13.68) is significant less than the traditional design (M = 23.42). There is a significant main effect of scenario, F (2.17, 50.00) = 19.66, $p < 0.001$, $\eta_p^2 = 0.461$. The effect suggests that time of completing the task is different among 4 scenarios. Further post-hoc comparisons showed that participant's first fixation of scenario 1 (M = 16.36) is significant less than scenario 3 (M = 21.30) and scenario 4 (M = 21.18). Participant's first fixation of scenario 2 (M = 15.36) is significant less than scenario 3 (M = 21.30) and scenario 4 (M = 21.18). Furthermore, there is a significant interaction between designs and scenarios, F (3, 69) = 8.97, $p < 0.001$, $\eta_p^2 = 0.281$. It seems that participants spent less time completing the task (T3) in integrated design than in traditional one. It might provide preliminary evidence that integrated design is more efficient for participants to deal with emergency. As the level of scenarios increases, average T3 increases in general, however, there is a slight decrease of average T3 in Scenario 2. These results suggest that the tasks in latter scenarios (4 & 3) are more complicated than former ones (2 & 1), with Scenario 2 to be the easiest one (Fig. 5b). The reason for this might be that the solutions for Cabin Altitude and Engine Fire are curter than others, as the situation is very urgent and demand quick response from flight crews.

The time of first fixation on instruction is a key aspect for the proposed integrated design, because it defines how efficient it is for users to locate the solution, which would provide more time for them to execute the instructions or gain extra time to conduct decision-making and problem-solving processes, in terms of improving SA. From the results of Within-Subjects ANOVA and T-test for T2 and, Fig. 5a, it is obvious that no matter what emergent scenario it is, participants spent less time finding the intended instruction and the difference between them can extend to more than 10 s. In emergent situations, every second is valuable, as it might lead to very different consequences, hence integrated design could be considered as a better design than traditional one. The significant main effect for scenarios might be contributed to the differences in contents of each instruction, in terms of facilitating perception, comprehension and projection for future status. The integrated design in this research is based on Proximity Compatibility Principle (Wickens and Carswell 1995) to design a warning system. There is of little use of alerting design if it is not efficient at disengaging attention from current task (Dehais et al. 2011). Thus, time of first fixation to be placed on CAS message is a very important factor to examine whether the alerts in the design is salient enough. Human operators' visual attention would be influenced by features in the operational environment (Carmichael et al. 2010), hence the result of their visual behavior will become more valid in evaluating the effectiveness of the design. As for the significant main effect of different scenarios, it might be the reason that different scenario has different form in presenting the alert, from color of the text, sound effect, to the content of the alert itself, so it might affect the time for participants to identify. Future study could examine this effect.

Task Completion Time is another key factor for evaluating a design, as the reason similar to T2, it defines how quickly users can achieve their goals. The results derived from T3 reveal similar effect, as participants took longer time finishing the task in current design. The significant main effect for scenario of completing the task suggests that the scenario settings are different, which is an indication that different emergent events pre-programmed are not common, thus provide different levels for this experiment. In integrated design, participants never made error in locating the instructions. It is due to the obvious fact that the instructions of integrated design will present itself, hence decrease the chance of executing the wrong NNC by eliminating the degree of choices.

4 Conclusion

Restructuring CAS message and QRH can facilitate cognitive function which will be able to reduce the cost of visual search by integrated them as an oracle to advise crews on what they must do next, or where particular information is to be found. Icons, objects, texts and emergent structure in CAS and QRH are not incidental to crews' cognition, but part of their thinking process to achieve cognitive goals, and space is an invaluable resource that can be exploited and managed (Hollan et al. 2000). A fundamental way to achieve this is by re-organizing and restructuring messages presentation to facilitate perception, hence making it easier to find relevant items (Kirsh 1995). The integrated design in this research is based on Proximity Compatibility Principle (Wickens and Carswell 1995) to design a warning system. Based on the analysis of collected eye movement data, it can be concluded that integrated design is no difference than current design in alert detecting time, which means the integrated design is as good as the current design as the warning system in current one meets regulations. Furthermore, the integrated design is significant quicker than current design on both finding the solutions and task completion time.

References

Ahlstrom, U., Friedman-Berg, F.J.: Using eye movement activity as a correlate of cognitive workload. Int. J. Ind. Ergon. **36**(7), 623–636 (2006). doi:10.1016/j.ergon.2006.04.002

Allsop, J., Gray, R.: Flying under pressure: effects of anxiety on attention and gaze behavior in aviation. J. Appl. Res. Memory Cogn. **3**(2), 63–71 (2014). doi:10.1016/j.jarmac.2014.04.010

Boorman, D.: Today's electronic checklists reduce likelihood of crew errors and help prevent mishaps. ICAO J. **56**(1), 17–20 (2001)

Burian, B.K.: Emergency and abnormal checklist design factors influencing flight crew response: a case study. In: Paper presented at the Proceedings of the International Conference on Human-Computer Interaction in Aeronautics 2004 (2004)

Burian, B.K.: Aeronautical emergency and abnormal checklists: expectations and realities. in: Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2006)

Carmichael, A., Larson, A., Gire, E., Loschky, L., Rebello, N.S., Singh, C., Rebello, S.: How does visual attention differ between experts and novices on physics problems? In: Paper Presented at the AIP Conference Proceedings (2010)

Carroll, P., Slowiaczek, M.L.: Constraints on semantic priming in reading: a fixation time analysis. Mem. Cogn. **14**(6), 509–522 (1986)

Dehais, F., Causse, M., Tremblay, S.: Mitigation of conflicts with automation: use of cognitive countermeasures. Hum. Factors **53**(5), 448–460 (2011). doi:10.1177/0018720811418635

Federal Aviation Administration: Operational use & modification of electronic checklists. Fed. Aviat. Adm. Advisory Circular **120**(64), 1–9 (1996)

Field, A.: Discovering Statistics Using IBM SPSS Statistics: and Sex and Drugs and Rock 'n' Roll, 4th edn. Sage Publications Ltd (2013)

Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. Int. J. Indus. Ergon. **24**(6), 631–645 (1999)

Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput. Hum. Interact. (TOCHI) **7**(2), 174–196 (2000)

Jones, D.G., Endsley, M.R.: Sources of situation awareness errors in aviation. Aviat. Space Environ. Med. **67**(6), 507–512 (1996)

Just, M.A., Carpenter, P.A.: A theory of reading: from eye fixations to comprehension. Psychol. Rev. **87**(4), 329 (1980)

Kirsh, D.: The intelligent use of space. Artif. Intell. **73**(1–2), 31–68 (1995)

Mitchell, M.L., Jolley, J.M.: Research Design Explained. Cengage Learning, Boston (2012)

Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychol. Bull. **124**(3), 372 (1998)

Rayner, K., Pollatsek, A.: Eye movements and scene perception. Can. J. Psychol. (Revue canadienne de psychologie) **46**(3), 342 (1992)

Rayner, K., Sereno, S.C., Morris, R.K., Schmauder, A.R., Clifton Jr., C.: Eye movements and on-line language comprehension processes. Lang. Cogn. Process. **4**(3–4), SI21–SI49 (1989)

Robinski, M., Stein, M.: Tracking visual scanning techniques in training simulation for helicopter landing. J. Eye Mov. Res. **6**(2), 1–17 (2013)

Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: Paper Presented at the Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (2000)

UK Civil Aviation.: AP 676: A Guidance on the Design, Presentation and Use of Emergency and Abnormal Checklists. CAA, London (2006)

Wickens, C.D., Carswell, C.M.: The proximity compatibility principle: its psychological foundation and relevance to display design. Hum. Factors **37**(3), 473–494 (1995)

Yu, C., Wang, E.M., Li, W.-C., Braithwaite, G.: Pilots' visual scan patterns and situation awareness in flight operations. Aviat. Space Environ. Med. **85**(7), 708–714 (2014)

An Analysis of Pilot's Workload Evaluation Based on Time Pressure and Effort

Wenmeng Liu(✉), Yanyu Lu, Dan Huang, and Shan Fu

School of Aeronautics and Astronautics, Shanghai Jiao Tong University,
Shanghai 200240, People's Republic of China
crytallwm@126.com

Abstract. Human factors has been playing more and more predominant role in aviation. Pilot workload evaluation is one of the most important fields in human factors to make thorough research. Among the various workload measure techniques, it is generally recognized that workload evaluation measurement is hard to elaborate. In this paper, the definition of workload is based on time pressure and effort. The relationship between time pressure and effort is the main focus to help the selection of workload evaluation approach. The models of workload evaluation are respectively set up in time pressure and effort. Heart rate, respiration rate, respiration depth, eye tracker data and control data were obtained in the flight experiment called double hydraulic failure. By comparison with the result of workload evaluation in time pressure and effort, it can indicate that the workload evaluation should take time pressure and effort altogether into account under double hydraulic failure condition. Eventually, the analysis on the relationship between time pressure and effort enables to proceed the research of workload evaluation.

Keywords: Workload evaluation · Time pressure · Effort · Physiological measurement

1 Introduction

Safety puts first priority in civil aviation. With the development of science and technology, the rates of flight accidents caused by has significantly decreased from 80% in the early 20th century to 3% at present [1]. However, there are still more than 60% flight accidents due to human factors. Under this circumstance, it is worthwhile to take scientific researches on human factors in civil aviation. Specifically, human factors play profound roles in the safety, comfort and efficiency of the aircrafts [2].

Pilot's workload can reflect the interaction of task factors, operator responses, operator performances and additional stressors [3]. Thus, evaluating pilot's workload can exert positive effect on the human factors research. It is hard to give a clear and comprehensive definition of pilot's workload. In general, workload measurement techniques are mainly divided into three categories: subjective rating scales technique, task performance measurement and physiological measurement [4–6].

Subjective rating scales technique mainly regards pilot's subjective experiences as their workload. NASA-TLX (Task Load Index), SWAT (Subjective Workload

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 32–41, 2017.
DOI: 10.1007/978-3-319-58472-0_3

Assessment Technique) and CH (Cooper-Harper Technique) are widely used in the field of subjective rating scales. Although it easily gets access to this approach, the result of it may not be reasonable and objective because of individual differences.

Task performance measurement is based on the completion of tasks and then work out pilot's workload. There are two aspects of this way, namely primary task performance measurement and secondary task performance measurement. Primary task performance measurement only values the pilot's completion of primary tasks while it neglects the pilot's efforts and initiatives. The pilot's completion of secondary task performance can reflect on the inability of primary task performance. Under this case, it may be time-consuming and involve in great efforts.

Physiological measurements rely on the changes of physiological signals to evaluate pilot's workload. There are mainly three varieties of physiological signals: ECG, EOG and EEG. Based on the previous studies, it was generally recognized that physiological signals enable to reflect on the changes of pilot's workload. Meanwhile the results of this approach can be objective and reasonable to some extent. Due to the multidimensional aspects of pilot's workload, single physiological signal cannot take all factors into account. Thus, it needs the multidimensional selections of physiological signals.

Based on the previous studies, it is generally recognized that the definition is hard to elaborate, which makes it difficult to select the proper approach to evaluate workload. In this paper, the definition of pilot's workload is based on time pressure and effort when the pilot proceeds a task, which can simultaneously take the pilot's physiological demands and psychological demands into account. Specifically, time pressure is whether there is sufficient time for the pilot to complete tasks under all expected conditions or not. The quantification and assessment of pilot's workload should take pilot's physical channels such as hand into account when the pilot takes actions. In time pressure aspect, it can reflect the combination of pilot's body movements, reactions and necessary perceptions during the flight. This approach involves in different pilot's body channels such as vision, hands, movements, sound and cognition. As for effort, it is whether the tasks can be accomplished without causing excessively physical load or not. The extent of pilot's effort cannot be directly observed during the flight based on the definition of effort. In this paper, physiological signals are used for reflecting the changes of pilot's workload because of its objectivity and capability. Under this case, it can successfully establish the model of pilot's workload based on effort.

With the help of these two models, the relationship between time pressure and effort is the main focus in this paper in order to evaluate workload. The reason for analyzing the relationship between time pressure and effort is aimed to find a proper approach to evaluate pilot's workload under the specific situation.

2 Method

In order to fulfill the pilot's workload, real flight experiments were undertaken in the flight simulator ARJ21-700. During this flight experiment, Heart rate, respiration rate, respiration depth, eye tracker data and control data were obtained in the flight experiment.

2.1 Participants

There were two pilots as subjects who have rich experience in flight between 30 – 60 years old. These participants were required to accomplish flight tasks according to the procedures.

2.2 Apparatus

There are two parts of the flight simulator: the outside view and the flight deck. Inside the flight deck, the arrangements of it consists of control instruments and display instruments. Thus, the flight data such as speed of the aircraft could be obtained from these instruments with the sample rate of 30 Hz.

The Zephry Bioharness device [7] with the sample rate of 1 Hz was used for the collection of heart rate, respiration rate and respiration depth. Pupil diameter, fixation duration and saccade frequency were recorded by Smart Eye Pro [8], the eye tracker with the sample rate of 30 Hz (Figs. 1 and 5).

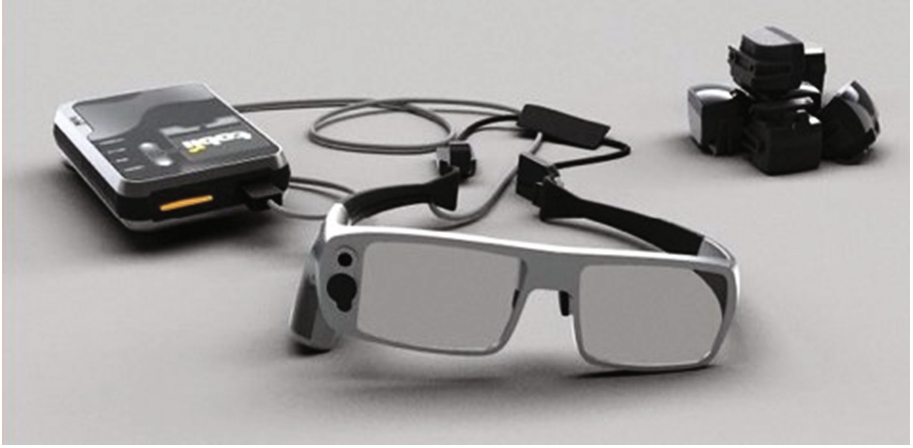


Fig. 1. Tobbi glasses eye tracker record device

2.3 Procedure

The day before flight experiment, all subjects were required to keep in good spirt so that they could perform the flight experiment properly. Meanwhile, light condition on the inner cockpit kept stable and all simulation tasks were taken under daytime. The temperature in cockpit stayed the same as real cockpit temperature. The participants were required to accomplish the flight task named double hydraulic failure.

During the experiment, the aircraft firstly kept in the initial state and it was controlled by the flight control system. While the aircraft was in the night cruise flight state, the teacher station set No. 2 hydraulic system failure and hydraulic pressure reduced from 3000 psi to 1800 psi below. Under this case, the pilot found out this fault and he was in accordance with the prescribed procedures for disposal. With the help of the instructor, the pilot guided the aircraft to the scheduled airport. Meanwhile, the pressure alarm of No. 2 hydraulic system lasted for 15 min, two electric pumps of No. 3 hydraulic system malfunctioned. In order to fix this problem, the pilot disconnected the autopilot function and the aircraft continued to fly to the scheduled airport. Under the instructor's command, the aircraft succeeded to navigate to the end of the designated airport landing phase.

In the experiment, the data were collected by the devices stated above. When the experiment carried out, the experimenter paid attention to the devices because the devices had to work properly. At the end of the experiment, the participant had a short break and the data was carefully saved.

2.4 Data Process

A three-way set of system was used to obtain the data in this experiment, namely the flight recorder, Tobbi Glasses eye tracker and BioHarness Wireless Physiology device. However, the data obtained by these devices was under different sample rate and the start-up time of these devices was not the same. In this paper, the method called first timeline synchronization calibration [9] was performed before the multi-dimensional integrated data analysis. Under this way, a multi-channel data on the time can not only achieve multi-dimensional integrated data acquisition system within different terminals, but also it would be easy to add or remove arbitrary data acquisition terminal.

There were noise and lost-detection occurred during the experiment. The noise in the experiment may result from various reasons such as the environmental factor and the reason for lost-detection may be because of the poor sensor. In order to handle out this problem, Hanning window [10] was applied to filter and smooth these physiological signal data. The reason for Hanning window is capable to deal with this problem. If the physiological signal feature is over 15% data lost in a trial, this trial data should be discarded.

The big issue of data process is about the selection of parameters for time pressure aspect and effort aspect. For effort aspect, the parameters of effort are mainly from physiological signals based on the definition of effort in this paper. Based on the previous studies, heart rate, respiration rate, respiration depth and pupil diameter can take physiological reactions when the pilot take actions [9]. Thus, these factors were selected as the parameters of effort aspect. Under this case, physiological parameters can objectively represent the changes of pilot's workload based on the effort definition stated above. The data of these four physiological signals should take preliminary process before establishing the model of pilot's workload evaluation based on effort. During the data process, the absolute values of these physiological data were meaningless due to the differences among participants. Under this case, the method of z-scores were applied to cope with the above problem.

As for time pressure aspect, the selection of parameters were the combination of cognition and physical motion because time pressure for pilot workload evaluation are divided into two categories: physical motion and cognition. The parameters of physical motion are from flight data, namely control time of instruments and control speed of instruments. The reason for selecting these two factors is that the control time of instruments and control speed of instruments can have the ability to reflect the physical aspect of the pilot to control the flight. For cognition aspect, the parameters of cognition are from eye movement data, namely fixation duration and saccade frequency. In the experiment, the data of fixation duration and saccade frequency cannot be directly obtained. The definition of these two factors are based on the cognition process. In this paper, it mainly focuses on the information process not the whole process of cognition.

Based on the above mentioned, the definition of all parameters were given clear translation and considered specific details in experiment. With the help of MATLAB programming, the data process can be successfully finished.

3 Result

From the Fig. 2 stated above, the physiological parameters of pilot can vividly represent the changes of pilot's workload. The trend of heart rate can show that the pilot took actions to cope with the emergency situations in the experiment. Meanwhile, the respiration rate and respiration depth are associated with activities of sympathetic nervous system [11]. Based on the previous studies, the respiration depth is sensitive to some typical events. It can represent the changes of pilot's workload. Under this case, multi-dimensional physiological parameters are regarded as the input of effort pilot workload model. Pupil diameter can show that the changes of pilot's workload when pilots performed monitor tasks.

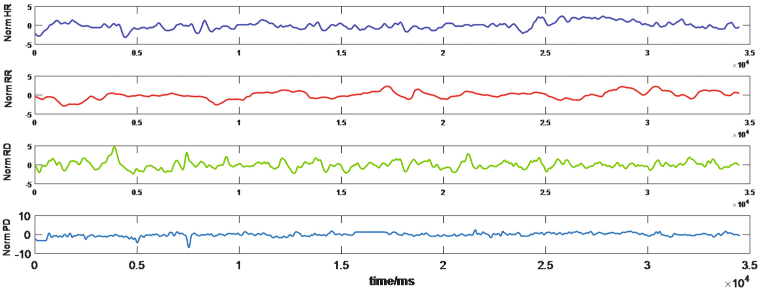


Fig. 2. Results of effort parameters

From the Fig. 3 stated above, control time and control speed of the devices can represent the pilot abilities of physical motion. The pilot followed the procedures of the experiment and took specific actions of the devices. In this paper, the definition of the control speed is the difference between the start-point and the end-point of each motion while the definition of the control time is the difference between the start-point of nonzero control speed and the end-point of zero control speed at one time. Control time can indicate the time of the pilot to take physical actions while control speed can represent the speed of the instrument when the pilot control the instrument. With the help of control time and control speed, the physical motion of pilots can be

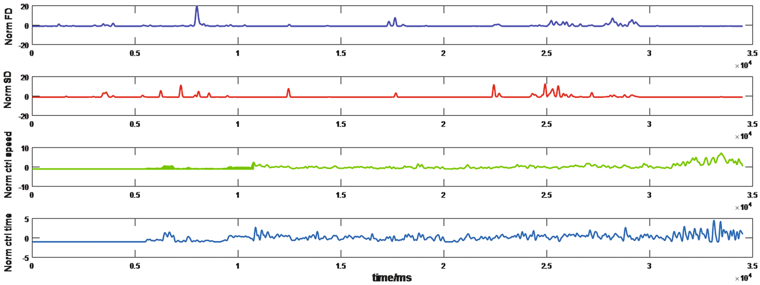


Fig. 3. Results of time pressure parameters

mathematically modeled. In cognitive aspect, the parameters of fixation duration and saccade frequency can represent the time of a pilot to acquire information [12]. The longer fixation duration is; the more cognition activities may be performed by the pilot.

After the above data process results, the model of pilot's workload evaluation in effort aspect can be mathematically set up while the model of pilot's workload evaluation in time pressure can also be mathematically established in the below.

In effort aspect, workload evaluation model can be integrated by the following equation:

$$Workload = a_1HR + a_2RR + a_3RD\ + a_4PD \tag{1}$$

Where:

- Heart Rate, Respiration Rate, Respiration Depth and Pupil Diameter are the results of physiological parameters stated above.
- a_1,a_2, a_3 and a_4 are the weights to represent the contributions of heart rate, respiration rate, respiration depth and pupil diameter. They are set by the algorithm called PCA (Principal Component Analysis).

In this paper, PCA is used to quantity the contributions of these factors (heart rate, respiration rate, respiration depth and pupil diameter). The essence of PCA is the process of transforming the high-dimensional space into low-dimensional space, which makes the problem become more intuitionistic and simple [13]. Each principal component obtained by this method is independent of each other and is a linear combination of original variables. These principal components can reflect most of the information of the original variables, and there is no overlap between them.

Based on the above analysis, the model of pilot's workload in effort can be worked out in Fig. 4. From the Fig. 4, it can indicate that pilot's workload has changed a lot under double hydraulic failure. The peak of pilot's workload happened when the complexity of procedures increased. Specifically, double hydraulic failure involved complex procedures to cope with emergencies. Under this case, pilot's physiological parameters can show this trend. The overall trend of pilot's workload under double hydraulic failure condition is changeable with the time domain. Analyzing the bottoms of the pilot's workload arrives at simple situations or familiar situations. For example, when the first alarm of double hydraulic failure occurred, the pilot had to take

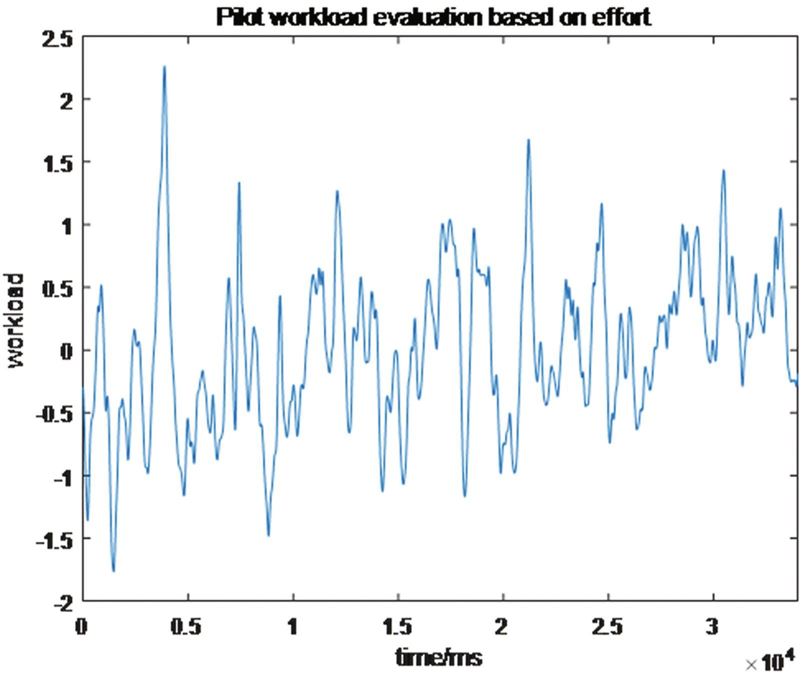


Fig. 4. Result of workload evaluation in effort

immediate action to deal with this emergency and adapted himself into this situation. Under this case, it would result in the changes of pilot's psychological parameters ups and downs.

In time pressure aspect, workload evaluation can also be integrated by the following equation

$$Workload = b_1 \frac{t_{control_{time}}}{T_{control_{time}}} + b_2 \frac{t_{control_{speed}}}{T_{control_{speed}}} + b_3 \frac{t_{fixation_{duration}}}{T_{fixation_{duration}}} + b_4 \frac{t_{saccade\_frequency}}{T_{saccade\_frequency}} \quad (2)$$

Where:

- $t_{control\_time}$, $t_{control\_speed}$, $t_{fixation_{duration}}$ and $t_{saccade\_frequency}$ are the results of parameters in time pressure stated above. $T_{control\_time}$, $T_{control\_speed}$, $T_{fixation_{duration}}$ and $T_{saccade\_frequency}$ are the baseline of four activities. The baseline data comes from the training data.
- b_1,b_2,b_3 and b_4 are the weighs to represent the contributions of physical motion and cognition activity. The PCA algorithm calculates them.

Based on the above analysis, the model of workload evaluation in time pressure can be mathematically worked out. The trend of workload evaluation in time pressure is associated with the tasks. The peak of workload in time pressure arrives at the latter of the experiment because the pilot had to accomplish more than one task simultaneously, which may largely increase the time pressure on the pilot. Meanwhile, the latter of workload evaluation in time pressure has changed a lot than the former one, which can also indicate that the pilot may take more actions during that time. The essence of time pressure is aimed to analyze the changes of pilot's workload associated with pilot's physical motion activity and cognition activity. Under this case, physical motion and cognition can indicate the actions taken by the pilot and then figure out the changes of pilot's workload. The overall trend of pilot's workload in time pressure can be divided into three parts: the stable segment, ups and downs segment, and the ascent segment. According to the procedures of the experiment, it can easily reflect on the complexity of tasks under double hydraulic failure condition. At the beginning of the experiment, the pilot took simple actions before the alarm of the emergency. Thus, the pilot's workload arrives at stable state. When the pilot coped with the emergency, he should take cognition as well as physical motion measures. When the complexity of task increased, it might increase the time pressure on the pilot. Under this case, the pilot's workload in time pressure increased as well.

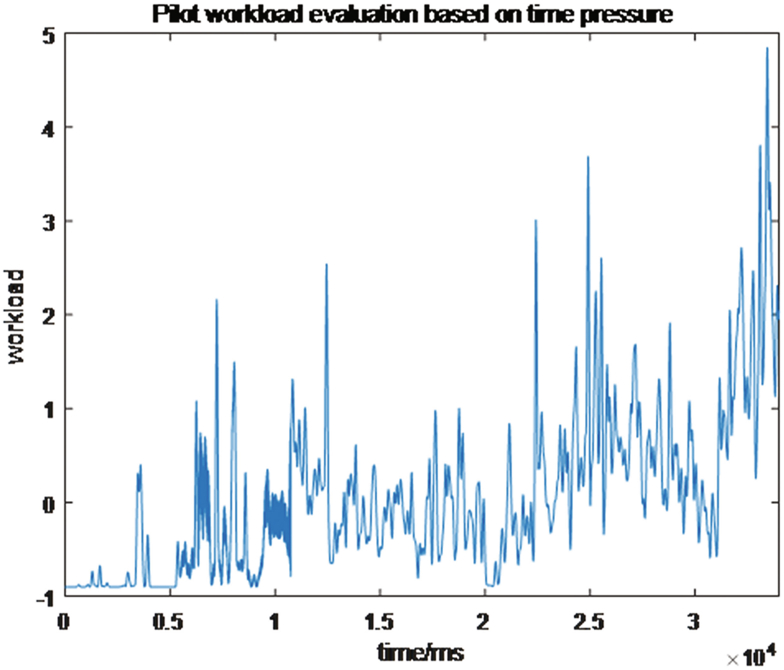


Fig. 5. Result of workload evaluation in time pressure

4 Discussion

By comparison with the result of pilot’s workload in effort and the result of pilot’s workload in time pressure, the trend of them were not the same. During the experiment of double hydraulic failure, the changes of pilot’s workload in effort occurred only when the physiological parameters changed. However, the changes of pilot’s workload in time pressure occurred when the cognition parameters changed or the physical motion parameters changed. Due to the complexity of the double hydraulic failure, the pilot had to pay diligent attention on the following emergencies of the experiment. Under this case, the mechanism of pilot’s workload evaluation should take time pressure and effort altogether into account. In this paper, the reason for doing the double hydraulic failure experiment is that this experiment is a typical flight mission and involved a lot of actions taken by the pilot. Thus, analyzing this typical flight mission is easily to figure out the relationship between the time pressure and effort when evaluating the pilot’s workload.

The relationship between the time pressure and effort under the double hydraulic failure condition is irrelevant. The pilot’s workload evaluation should take these two aspects into account. However, the relationship between the time pressure and effort in other cases may be relevant. Under this case, it should focus on the analysis of tendency on time pressure and effort. After this necessary analysis, the pilot’s workload evaluation should establish specific model under the specific case.

5 Conclusion

In this paper, workload evaluation has been measured into two aspects namely effort and time pressure, which is aimed to research the essence of workload. From the experimental results, it can indicate that the results of workload evaluation in effort are not the same as that in time pressure. Under this case, the workload evaluation of the experimental situation has to take the time pressure together with effort into account. Yet, the result of workload evaluation in effort may be similar with that in time pressure. Under this condition, it is not necessary to consider the overlap sections when calculating the result of workload evaluation. Overall, the essence of workload is hard to understand. The proposed method of workload evaluation is aimed to proceed the workload evaluation based on real pilot experiments. The reliability of this approach will be tested in the future with more thorough experiments.

References

1. Wiegmann, D.A., Shappell, S.A.: Human error and crew resource management failures in Naval aviation mishaps: a review of U.S. naval safety center data. Aviat. Space Environ. Med. **70**, 1147–1151 (1990)
2. FAA System Safety Handbook, Chap. 17: Human Factors Principles & Practices (2003)

3. Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P.: Human Factors Methods: A Practical Guide for Engineering and Design. Ashgate, Aldershot (2005)
4. Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Sage Publications (2006)
5. Kim, J.H., et al.: Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr BioHarness. Int. J. Sports Med. **34**, 497–501 (2013)
6. Wilson, G.F.: An Analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int. J. Aviat. Psychol. **12**, 3–18 (2012)
7. Zephyr Technology, BioHarnessTM User Guide (2010)
8. Smart Eye, A.B.: Smart-Eye Pro 5.6 User manual. Sweden-, Gothenburg, Smart Eye AB, Sweden Smart Eye AB (2009)
9. Wang, Z., Fu, S.: An analysis of pilot's physiological reactions in different flight phases. In: Harris, D. (ed.) EPCE 2014. LNCS (LNAI), vol. 8532, pp. 94–103. Springer, Cham (2014). doi:10.1007/978-3-319-07515-0_10
10. Tuovinen, J.E., Paas, F.: Exploring multidimensional approaches to the efficiency of instructional conditions. Instr. Sci. **32**, 133–152 (2004)
11. Jainta, S., Baccino, T.: Analyzing the pupil response due to increased cognitive demand: an independent component analysis study. Int. J. Psychophysiol. **77**, 1–7 (2010)
12. Wang, Z., Fu, S.: A layered multi-dimensional description of pilot's workload based on objective measures. In: Harris, D. (ed.) EPCE 2013. LNCS (LNAI), vol. 8020, pp. 203–211. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39354-9_23
13. Niu, Z, Qiu, X.: Facial expression recognition based on weighted principal component analysis and support vector machines. In: 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE). IEEE (2010)

The Effects of Task Complexity and Spatial Ability on Teleoperation Performance

Dan Pan, Yijing Zhang(✉), and Zhizhong Li

Department of Industrial Engineering, Tsinghua University,
South 524E, Shunde Hall, Haidian District,
Beijing 100084, People's Republic of China
{pandan16, zhangyijing, zzli}@tsinghua.edu.cn

Abstract. This study aims to explore how task complexity and spatial ability on teleoperation performance, especially the interaction effects of task complexity and spatial ability. Three kinds of robotic arm teleoperation task were designed, namely point aiming, line alignment, and cross alignment. They were respectively treated as teleoperation task with low, middle and high complexity. Teleoperation performance were measured from task completion time, rate of extra distance moved, operation slip and collision. Forty subjects were recruited. They were divided into two groups (with high spatial ability and with low spatial ability) based on their scores of the Vandenberg test and the Guay test.

Repeated measures' analyses of variance was carried out to examine the main effects and interaction effects of task complexity and spatial ability on teleoperation performance. The results shown that spatial ability significantly or marginally significantly influenced task completion time ($p = 0.037$), collision ($p = 0.003$), and operation slip ($p = 0.07$). The subjects with high spatial ability performed better than those with low spatial ability. Task complexity significantly affected completion time ($p < 0.001$), rate of extra distance moved ($p < 0.001$), operation slip ($p = 0.028$), and collision ($p < 0.001$). It was also found that the interaction effect of spatial ability and task complexity on collision was marginally significant ($p = 0.069$). Those results implied that spatial ability plays a key role in teleoperation, especially for high complexity tasks. Spatial ability should be considered as an important criterion for tele-operator selection.

Keywords: Task complexity · Spatial ability · Teleoperation performance

1 Introduction

Teleoperation refers to an operation form that remotely control a robot, or system to accomplish a given task [9]. Teleoperation technology has been widely applied in various fields, especially in risky or unknown environment, ranging from search and rescue activities (e.g., search for survivors in the 911 event), underwater adventures, toxic or nuclear material processing to daily industrial and commercial systems such as microsurgery, mineral exploitation [1].

Teleoperation performance is influenced by many factors, such as operators' cognitive characteristics [8], visual interface providing real-time scene [11], and task complexity. Task complexity is an important task characteristic that affecting human

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 42–50, 2017.
DOI: 10.1007/978-3-319-58472-0_4

performance and behavior [6]. No general agreement on the relationship between task complexity and human performance is found. Based on previous studies, Liu et al. [7] summarized four common types to this relationship, namely, negative correlation, positive correlation, inverted U-shaped correlation, and dependent on other factors. In present study, we will tentative explore the relationship between task complexity and teleoperation performance.

Spatial ability is an individual's cognitive ability in the aspect of space or visual imagery [5]. To be specific, spatial ability is an individual's ability in the aspect of identifying, coding, storing, representing, decomposing, and integrating the environment spatial information. It is an important component of an operators' cognitive characteristics. Pan et al. [8] has found that spatial ability played key role in teleoperation. Therefore, the present work will also examine whether the relationship between task complexity and teleoperation performance is dependent on spatial ability, namely the interaction effects between task complexity and spatial ability.

2 Method

2.1 Subjects

Forty male subjects were recruited to participate in this study. They were all undergraduate engineering students, aged from 18–22 years old (mean age = 21.4, SD = 1.3), right handed, and without color blindness. They had no experience on teleoperation even without or only with a little knowledge about teleoperation. Before participating, they were informed about the details of the experiment and voluntarily signed the informed consent form. The experimental procedure was approved in advance by the ethics committee of China Astronaut Research and Training center.

2.2 Teleoperation Tasks

Based on the Virtual Robot Experimentation Platform (VREP), three kinds of simulated robotic arm teleoperations (namely point aiming, line alignment, and cross alignment) were designed. Different shapes were set for the end effector of a simulated robotic arm in different teleoperation tasks. As shown in Fig. 1(I), the end effector and target location were set as a sphere in the point aiming task. This teleoperation was successfully completed when three position deviations (X, Y, Z) satisfying accuracy requirements. As shown in Fig. 1(II), the end effector and target location were set as a cylinder in the line alignment task. This teleoperation was successfully completed when three position deviations (X, Y, Z) and two angle deviations (a, b) reaching accuracy requirements. As shown in Fig. 1(III), the end effector and target location were set as a three-dimensional cross. This teleoperation was successfully completed when three position deviations (X, Y, Z) and three angle deviations (a, b, c) satisfying accuracy requirements.

For each teleoperation task, a computer screen provided real-time global view, target view, end effector view, and position and angle deviations of the robotic arm teleoperation situation for the subjects. The subjects were asked to operate two 3DOF

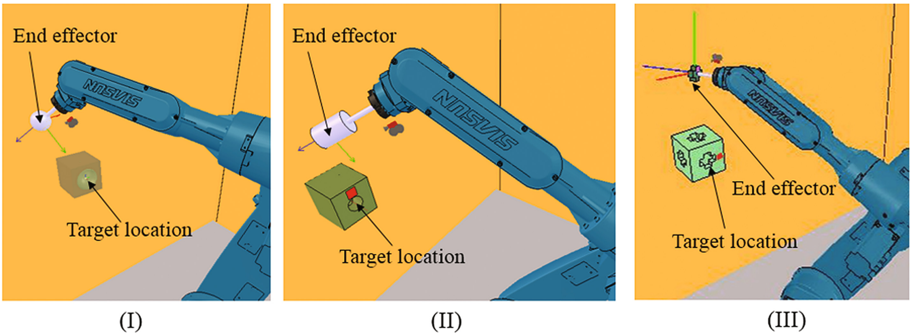


Fig. 1. Teleoperation tasks: (I) point aiming, (II) line alignment, (III) cross alignment

joysticks (Litestar PNX-2013) to control the movements and rotations of the simulated robotic arm's end effector. The position deviations (X, Y, Z) were adjusted by the movements of the end effector in three directions which the angle deviations (a, b, c) were adjusted by the rotations of the end effector along the three axes. After the subjects were introduced the experiment content and familiar with the control mode, they completed each teleoperation task twice. Teleoperation performance of each task was defined as the average performance of these two repeating operations.

During a teleoperation task, the VREP platform recorded the real-time position deviations, angle deviations, and distance moved of the end effector, movements and rotations of two joysticks, number of collision, and task completion time every 0.4 s. When a collision occurred, a prompting window will pop up to alert subjects and told them how many collisions they had made. When a teleoperation task completed successfully, a prompting window will also pop up to inform subjects, and the program automatically stop and return to the original state.

2.3 Dependent Variables

Teleoperation performance were measured from task completion time, rate of extra distance moved, number of operation slip and collision. Their definitions were descripted in Table 1, which were similar to definitions in [8]. Completion time reflects subjects' operation efficiency. Rate of extra distance moved and operation slip reflects subjects' operation effectiveness. Number of collision reflects subjects' operation reliability.

2.4 Independent Variables

Task complexity. This variable has three levels. Based on the different accuracy requirements of successfully completion, three kinds of teleoperation tasks (namely the point aiming, line alignment, and cross alignment) were respectively treated as a teleoperation with low, middle and high task complexity (see details in Sect. 2.2).

Table 1. Definitions of dependent variables [8]

Performance measurements	Definitions
Completion time	How long it took to complete a successful teleoperation. Shorter time indicates higher operation efficiency
Rate of extra distance moved	$\frac{total\ moved\ distance - initial\ position\ deviation}{initial\ position\ deviation}$ Lower ratio indicates better path planning
Number of operation slip	An operation slip was counted when two consecutive operations of joysticks within 2 s were completely opposite, except the residence time between these two operations is longer than the time spends on the first operation. Fewer operation slips indicate higher operation reliability
Number of collision	A collision was counted when any part of the simulated space manipulator collided with the environment or the target cube. Fewer collisions indicate higher operation reliability

In present study, task complexity was a within-subject variable. All subjects were asked to finish three kinds of teleoperation tasks with different task complexity.

Spatial ability. This variable has two levels. It was a between-subject variable. Based on subjects' scores of spatial ability tests, forty subjects were divided into two groups (high ability vs. low ability). Subjects' spatial ability were evaluated from two dimensions, namely mental rotation and perspective taking, separately by the revised Vandenberg test and the adapted Guay test.

The revised Vandenberg test, which has a high internal consistency and test-retest reliability [10], is widely used to evaluate an individual's mental rotation. This test included two sets of 12 items. As shown in Fig. 2(a), each item consists of a criterion figure and four stimulus figures (A, B, C and D). The subjects were asked to find out two figures from A, B, C, D, which matched the criterion figure. They had 4 min for each set.

The adapted Guay test is known as a standardized test for perspective taking ability [3]. This test included 24 items. As shown in Fig. 2(b), each item consists of an

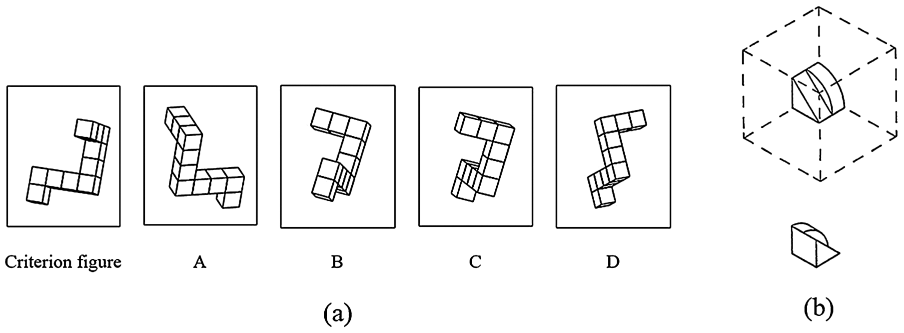


Fig. 2. Examples of spatial ability tests: (a) An example of the revised Vandenberg test; (b) An example of the adapted Guay test

isometric view of a 3-dimensional object depicted in the center of a see-through cube and another view for the same object depicted below the cube which is from a different viewpoint. The subjects were asked to find out the corner of the cube from which the second view of the object is taken. They had 8 min to complete this test.

2.5 Data Analysis

Repeated measures' analyses of variance (repeated ANOVA) was conducted to explore the relationship between task complexity and teleoperation performance, and that whether this relationship dependents on spatial ability. A Mauchly's test of sphericity was performed to guarantee that those data satisfy the assumptions for the repeated measure analysis. A Huynh-Feldt ε correction factor was used when the sphericity was violated ($p < 0.05$). Besides, post-hoc analyses were used to evaluate differences between teleoperation tasks with different complexity.

3 Results

Table 2 shows the descriptive statistic results of teleoperation performance. Table 3 presents the results of repeated ANOVA. Post-hoc analyses' results for task complexity are presented in Table 4.

Table 2. Mean (standard deviation) of performance measurements

		Completion time (s)	Rate of extra distance moved	Operation slip (#)	Collision (#)
Spatial ability	Low	211.04 (18.04)	2.93 (0.38)	16.2 (2.3)	3.5 (0.4)
	High	156.39 (17.58)	2.06 (0.37)	10.3 (2.2)	1.9 (0.4)
Task complexity	Low	104.38 (8.48)	1.21 (0.13)	8.1 (1.1)	0.5 (0.1)
	Middle	201.44 (14.88)	1.56 (0.19)	15.8 (2.3)	4.5 (0.6)
	High	245.33 (29.09)	4.72 (0.65)	15.8 (3.2)	3.1 (0.4)

Table 3. Results of repeated measures ANOVA

	Huynh-Feldt ε	Spatial ability		Task complexity		Spatial ability × Task complexity	
		F	*p*	F	*p*	F	*p*
Completion time (s)	.699	4.707	.037	15.684	<0.001	1.005	.347
Rate of extra distance moved	.599	2.689	.110	28.841	<0.001	.400	.568
Operation slip (#)	.743	3.477	.070	4.310	.028	.640	.486
Collision (#)	.828	10.180	.003	26.656	<0.001	2.958	.069

Table 4. Post-hoc analyses' results for task complexity (*p* value)

Task complexity		Completion time (s)	Rate of extra distance moved	Operation slip (#)	Collision (#)
Low	Middle	<0.001	.081	<0.001	<0.001
Low	High	<0.001	<0.001	.024	<0.001
Middle	High	.167	<0.001	.999	.049

3.1 Completion Time

From the results of repeated measures ANOVA in Table 3, it was found that both spatial ability and task complexity significantly influenced completion time ($p = 0.037$, $p < 0.001$). From the post-hoc analyses' results for task complexity in Table 4, it was found that the difference of completion time between the low complexity teleoperation and middle complexity teleoperation, and between the low complexity teleoperation and high complexity teleoperation were significant ($ps < 0.001$).

It can be seen from Table 2 that the subjects with high spatial ability significantly spent 25.9% less time than those with low spatial ability ($Mean_{\text{high spatial ability}} = 156.39$ s, $Mean_{\text{low spatial ability}} = 211.04$ s). Meanwhile, subjects spent 48.2% less time on the low complexity teleoperation (i.e., point aiming) than on the middle complexity teleoperation (i.e., line alignment) ($Mean_{\text{low complexity}} = 104.38$ s, $Mean_{\text{middle complexity}} = 201.44$ s), and spend 57.5% less time on the low complexity teleoperation than on the high complexity teleoperation (i.e., cross alignment) ($Mean_{\text{low complexity}} = 104.38$ s, $Mean_{\text{high complexity}} = 245.33$ s).

3.2 Rate of Extra Distance Moved

From the results presented in Table 3, it can be found that task complexity significantly influenced the rate of extra distance moved ($p < 0.001$). From the post-hoc analyses' results for task complexity in Table 4, it was found that the difference of the rate of extra distance moved between the low complexity teleoperation and high complexity teleoperation, and between the middle complexity teleoperation and high complexity teleoperation were significant ($ps < 0.001$). Meanwhile, the difference of the rate of extra distance moved between low complexity teleoperation and middle complexity teleoperation were marginally significant ($p = 0.081$).

3.3 Operation Slip

As shown in Table 3, the number of operation slip was marginally significantly affected by spatial ability ($p = 0.070$), and was significantly affected by task complexity ($p = 0.028$). From the post-hoc analyses' results for task complexity in Table 4, it was found that the difference of operation slip between the low complexity teleoperation and middle complexity teleoperation, and between the low complexity teleoperation and high complexity teleoperation were significant ($p < 0.001$, $p = 0.024$).

It can be seen from Table 2 that the subjects with high spatial ability significantly made 36.4% fewer operation slips than those with low spatial ability ($Mean_{\text{high spatial ability}} = 10.3$, $Mean_{\text{low spatial ability}} = 16.2$). Meanwhile, subjects made 48.7% fewer operation slips in the low complexity teleoperation (i.e., point aiming) than in the middle complexity teleoperation and in the high complexity teleoperation (i.e., line alignment, cross alignment) ($Mean_{\text{low complexity}} = 8.1$, $Mean_{\text{middle complexity}} = 15.8$, $Mean_{\text{high complexity}} = 15.8$).

3.4 Collision

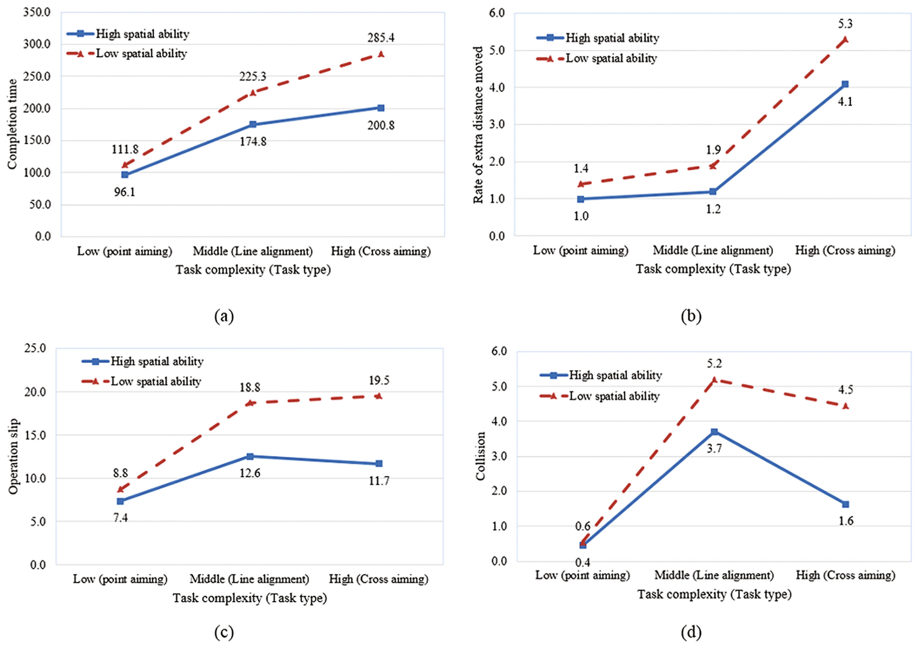
As the results presented in Table 3, the number of collision was significantly influenced by both spatial ability and task complexity ($p = 0.003$, $p < 0.001$). Meanwhile, the interaction effect between spatial ability and task complexity on the number of collision was marginally significant ($p = 0.069$). From the post-hoc analyses' results for task complexity in Table 4, it was found that the difference of collision between the low complexity teleoperation and the middle complexity teleoperation, between the low complexity teleoperation and the high complexity teleoperation, and between the middle complexity teleoperation and the high complexity teleoperation were all significant ($p < 0.001$, $p < 0.001$, $p = 0.049$).

It can be seen from Table 2 that the subjects with high spatial ability significantly made 45.7% fewer collisions than those with low spatial ability ($Mean_{\text{high spatial ability}} = 1.9$, $Mean_{\text{low spatial ability}} = 3.5$). Meanwhile, subjects made 88.9% fewer collision in the low complexity teleoperation (i.e., point aiming) than in the middle complexity teleoperation (i.e., line alignment), and made 83.9% fewer collisionin the low complexity teleoperation than in the high complexity teleoperation (i.e., cross alignment) ($Mean_{\text{low complexity}} = 0.5$, $Mean_{\text{middle complexity}} = 4.5$, $Mean_{\text{high complexity}} = 3.1$).

4 Discussion and Future Work

From Fig. 3, we can see that with the increase of task complexity, the difference of teleoperation performance (namely, completion time, rate of extra distance moved, operation slip, collision) between subjects with high and low spatial ability were also increasing. Those results implies that spatial ability plays a key role in teleoperation, especially for high complexity tasks. The actual teleoperation tasks, such as space robotic arm teleoperation, are much more complex than the simulated robotic arm teleoperation in present study. Therefore spatial ability should be considered as an important criterion for tele-operator selection.

Furthermore, as shown in Fig. 3(d), not only for the subjects with high spatial ability but also for the subjects with low spatial ability, they had more collisions in line alignment teleoperation than in cross alignment teleoperation. In the simulated robotic arm teleoperation of present study, collisions usually occurred in the fine-tuning stage. In the fine-tuning stage of line alignment, the subjects need to adjust the Z position deviation in a wide range keeping two angle deviations (a, b) unchanged. This is an action with strict constraints. During this kind of action, it is very easy to collide. But in point aiming and



(a)

(b)

(c)

(d)

Fig. 3. The interaction effects of spatial ability and task complexity on teleoperation performance

cross alignment, this kind of action does not exist. This result indicated that apart from accuracy requirement but the teleoperation process (i.e., the number of actions with constraints) should be also taken into account in the evaluation of teleoperation complexity.

According to reference [6], the conceptualization, identification, and measurement of task complexity is really complicated. It can be speculated that it is the same to teleoperation complexity. The present study is a tentative exploration on teleoperation complexity, and its relationship with teleoperation performance, especially for operators with different spatial ability. In future work, more research should be conducted to the measurement of teleoperation complexity. Moreover, the inverted U-shaped correlation between task complexity and performance is an important theoretical assumption [2, 4, 12]. If this assumption is proved in teleoperation. It will be a tremendous contribution for teleoperation design since an individual performs best in teleoperation with appropriate complexity based on the inverted-U shaped correlation. Therefore, empirical research is really necessary for this theoretical assumption in future work.

5 Conclusion

The present work found spatial ability significantly or marginally significantly influenced task completion time, collision, and operation slip in simulated robotic arm teleoperation. The subjects with high spatial ability performed significantly better than those with low spatial ability. That implies that spatial ability plays a key role in teleoperation. Furthermore, task complexity significantly affected completion time, extra distance moved, operation slip, and collision in the simulated teleoperation of present study. It was also found that the interaction effect of spatial ability and task complexity on collision was marginally significant. Compared to low complexity teleoperation, the performance difference between the subjects with low spatial ability and those with high spatial ability were larger in high complexity teleoperation. Therefore, spatial ability should be considered as an important selection criterion for tele-operators, especially those operators for high complexity teleoperation.

Acknowledgements. This study was supported by the National Natural Science Foundation of China (Project no. 71371174, 71671167).

References

1. Boboc, R.G., Moga, H., Talabă, D.: A review of current applications in teleoperation of mobile robots. Bull. Transilvania. Univ. Brasov. Ser. I: Eng. Sci. **5**(2), 9–16 (2012)
2. Day, R.H.: Relative task difficulty and transfer of training in skilled performance. Psychol. Bull. **53**, 160–168 (1956)
3. Guay, R.B.: Purdue Spatial Visualization Test. Purdue Research Foundation, West Lafayette (1976)
4. Hancock, P.A., Warm, J.S.: A dynamic model of stress and sustained attention. Hum. Factors. **31**, 519–537 (1989)
5. Linn, M.C., Petersen, A.C.: Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Dev. **56**(6), 1479–1498 (1985)
6. Liu, P.: Task complexity: conceptualization, identification, measurement, and effects. Tsinghua University (2014)
7. Liu, P., Li, Z.: Toward understanding the relationship between task complexity and task Performance. In: Rau, P.L.P. (ed.) IDGD 2011. LNCS, vol. 6775, pp. 192–200. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21660-2_22
8. Pan, D., Zhang, Y., Li, Z., Tian, Z.: Association of individual characteristics with teleoperation performance. Aerosp. Med. Hum. Perform. **87**(9), 772–780 (2016)
9. Park, S.H., Woldstad, J.C.: Design of visual displays for teleoperation. In: Karwowski, W. (ed.) International Encyclopedia of Ergonomics and Human Factors, pp. 897–900. CRC Press, Boca Raton (2000)
10. Vandenberg, S.G., Kuse, A.R.: Mental rotations, a group test of three-dimensional spatial visualization. Percept. Mot. Skills **47**(2), 599–604 (1978)
11. Xue, S., Jiang, G., Tian, Z., Jiang, T.: Progress of the key technologies in human-computer interface in space teleoperation. Manned Spaceflight **20**(5), 497–502 (2014)
12. Young, M.S., Stanton, N.A.: Attention and automation: new perspectives on mental underload and performance. Theoret. Issues Ergon. Sci. **3**, 178–194 (2001)

Model-Driven Payload Sensor Operation Assistance for a Transport Helicopter Crew in Manned–Unmanned Teaming Missions: Assistance Realization, Modelling Experimental Evaluation of Mental Workload

Christian Ruf(✉) and Peter Stütz(✉)

Institute of Flight Systems, University of the Bundeswehr Munich (UniBwM),
Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
`{christian.ruf,peter.stuetz}@unibw.de`

Abstract. One of the research fields at the Institute of Flight Systems (IFS) of the University of the Armed Forces (UniBwM) concerns the integration of reconnaissance sensor operator support in manned-unmanned teaming (MUM–T) transport helicopter (HC) missions. The purposive deployment of mission sensors carried by several unmanned aerial vehicles (multi–UAV) in such missions brings in new and impactful aspects, specifically in workload-intensive situations. An associate system offering variable automation levels supports the HC's crew by deploying machine-executable functionalities and high-level capabilities. The crews' work-processes to handle the reconnaissance payload as well as to derive and include relevant information in the mission progress are expected to induce additional mental workload (MWL) during operation. First, this paper gives an overview of the assistance concept for sensor operation to minimize the crews' MWL. Furthermore, an instance of a combined task- and resource model that describes MWL for several levels of automation in sensor guidance and payload sensor data evaluation is presented. Model parameters of human interaction for a holistic task- and activity set will be described. Finally, a method for demand parameter value determination from a dataset gained by an experimental campaign and results are presented.

Keywords: Human factors · Mental workload · Workload modelling · MUM-T · multi–UAV · Mission sensors · Levels of automation · Assistant system

1 Introduction

In this MUM-T approach, the UAVs shall enable the crew to directly reconnoiter the intended routes of flight, to survey certain areas such as operation sectors or potential landings zones and provide information on other mission-specific conditions. However, since there is no dedicated sensor operator as compared to legacy UAV systems, the HC' commander has to handle all related tasks including UAV guidance, mission sensor deployment and data assessment. Figure 1 shows the schematic team configuration which is focus of interest in this evaluation study.

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 51–63, 2017.
DOI: 10.1007/978-3-319-58472-0_5

Fig. 1. MUM-T approach illustrated in helicopter mission flight simulator

To address potential workload increase resulting from the broader task spectrum and higher overall mission complexity, the crew shall be supported by an adaptive, cognitive associate system [1–3]. It will provide situation-adapted support by continuous crew supervision and aims to balance the crews' workload. In [4], three basic requirements for associate system behavior were proposed, to be applied by associate systems for crew support. To reduce the overtaxing MWL, the associate system can involve suitable automation systems that provide context-depended, variable designed support.

Applying these design requirements to a sensor assistant system fosters the idea of situation-dependent crew support by executing automated machine-processes.

The goal is to achieve a solution that enables the crew to guide the UAVs directly from the helicopters cockpit during flight, in which the UAVs automatically supply mission-relevant reconnaissance results from sensor-perceived and pre-evaluated data.

2 Sensor Operation Assistance Concept

2.1 Motivation

When sensor payload operation now is to be automated in the cockpit, two main factors regarding human operators need to be considered:

- the effects on the crews' MWL situation [3] during system operation
- the operators' "Trust in Automation" [5]

These factors are essential in human machine cooperation investigated in this MUM–T configuration.

The crew commanders' working situation is mainly affected by the induction of MWL, caused by additional tasks in the field of reconnaissance sensor operation.

Especially perceiving and interpreting reconnaissance data causes additional MWL, depending on data bandwidth and degree of data abstraction. Therefore data bandwidth and data abstraction are the addressed independent parameters to influence the induction of MWL.

Regarding an operators' "Trust in Automation", the automation of airborne reconnaissance brings in domain specific conditions. Automated sensor evaluation systems often do not perform in a highly deterministic way, e.g. because of imperfection in sensor data evaluation [6] or varying operation environments. Out of a technical perspective, this circumstance can be addressed by the measure of "trustworthiness" of automated reconnaissance systems. In [7], a performance prediction method for image assessment algorithms is proposed. In this MUM–T application field, the trustworthiness of automated image assessment by algorithms directly affects reconnaissance performance. Figure 2 shows the effect "Trust in Automation" in the reconnaissance systems, besides existing effects that are well known in highly automated and complex cockpit environments ("Out-of-the-loop" [8], "Opacity-effect" [9, 10], "Over-Reliance" [11], "Brittleness" [9]).

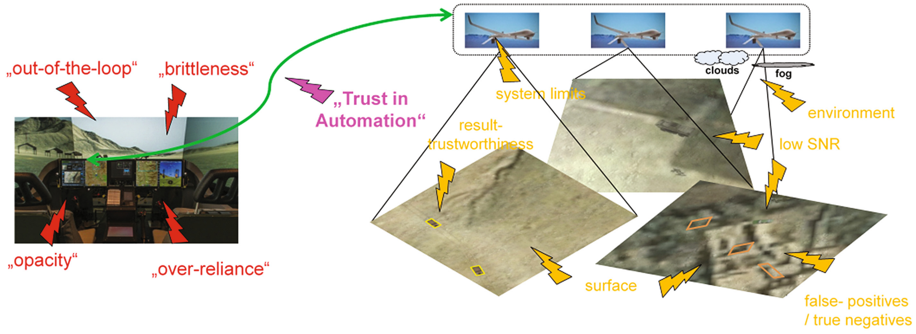


Fig. 2. Causal relationship of "Trust in Automation" and inconsistent automation character

As a result, decreasing automated reconnaissance performance or low trust in automation could lead to deeper operator involvement and more preference for executing a task manually, which both induces workload and contrasts the aim to reduce workload. In the event of decreasing automated reconnaissance performance, human MWL is required. Concluding, automated reconnaissance performance and MWL are contrary.

2.2 Solution Approach

The two dimensions, automated system performance (represented by trustworthiness) and human mental workload are triggers and dependencies for a variable interaction concept. So, the proposed general approach is to maximize the necessary reconnaissance performance by decomposing and balancing the antagonism of a tolerable workload and automated reconnaissance performance.

The proposed operation mode of assistance aims to purposefully modify the cooperative relationship between the human operator and machine processes. By adapting the reconnaissance system's automation level, changes in the crews' MWL-state are expected. According to [4], task transformation to an easier level is one major bullet in crew support, so changing to a higher automation degree is assumed to reduce workload and transform the non-manageable task demand situation to a manageable one. In contrast a forecasted decrease in machine reconnaissance performance would reduce the automation degree to preserve recon performance by involving more crew resources.

2.3 Realization

As described in [6], three major functional subsystems to implement such an assistance concept were introduced, realized by software implementations. They comprise

- functionalities for data presentation and assessment on different levels of automation,
- crew observation for workload based trigger generation as well as
- decision making for automation level selection.

With respect to the first, variable support means by application of the "levels of automation" (LOA) paradigm [12, 13] in the domain of sensor deployment was presented in previous work [6]. Here a repository of tools for reconnaissance sensor operation was realized on several levels of automation. This toolset includes functionalities for automated data preprocessing and evaluation as well the automated control of the payload gimbal.

With respect to the last bullet point, a management component was implemented to select a suitable level of automation. The tradeoff between the crews MWL and a maximized automated reconnaissance performance is solved by a machine decision process [6], which is able to automatically adapt the degree of automation.

As a prerequisite for decision making, knowledge about the crew's activity is needed. For this a task-model, containing knowledge of the crew' task load and associated MWL when performing tasks with different automated support, is utilized. This model is embedded in software and linked to an online crew activity determination. For crew activity determination, an external crew supervision system is used to generate the workload measure of the human crew [14] by referring model knowledge of all crew activities. Such activity determination enables an associate system to trigger automation level changes detecting anomaly in task and workload appearance.

This paper focuses on establishing such a task-model.

3 Modelling of Task-Related MWL

3.1 Modelling Principle

In [3], a unified theory for a task representation and context-rich representation of MWL was presented. This concept is applied to a holistic human-machine system. The term task is used to denote a means of communication between a human operator and an associate system as well as an interface between components of the associate system themselves. The task construct was also used as expression to describe mental workload. Also, the operationalization of crew tasks by a task model was presented. As well, the derivation of associated MWL from such a task model was demonstrated and a model instance covering the execution of a MUM-T- transport helicopter cargo mission was presented [3].

According to the method introduced in [3], the structure consists of elements representing mission tasks, tasks, complex tasks, actions, properties and relations (such as alternative, inheritance) between them and evidences for occurrence (Fig. 3). With these elements, structures can be built up, representing human activity and demand resource allocation.

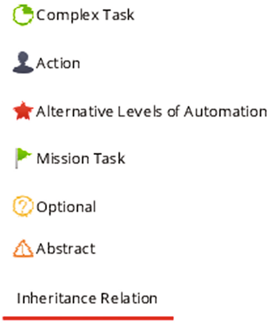


Fig. 3. Legend of model properties used for complex task representation

3.2 Domain-Applied Model Generation

Focus of this work is now to create a task model instance representing aspects of sensor guidance from the transport helicopter's cockpit. By applying the modelling principle, a depiction of different crew demands for collaboration with the automation system is aspired. For each necessary crew task performed on several automation levels, a corresponding demand representation was added to the task model. Beginning from the lowest elementary task type which is the first type to combine actions, a task structure was built up. For each elementary task, corresponding demands are assigned, represented by eight dimensions according to Wickens' multiple resource theory [15] (Fig. 4). The representation covers demand components of information perception (visual spatial, visual verbal, auditory spatial, auditory verbal), information processing (cognitive spatial, cognitive verbal) and response (manual spatial, vocal verbal).

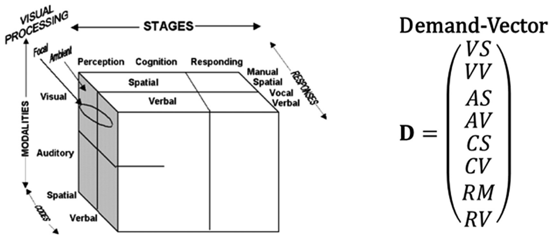


Fig. 4. Multiple resource model and demand vector according to [15]

A tree representation follows which re-uses subtasks and gives a model instance for crew interaction with the automated sensor system (Fig. 5).

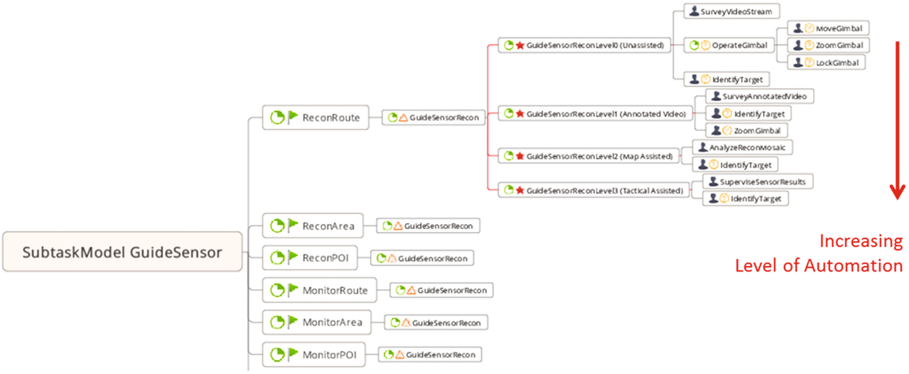


Fig. 5. Snapshot of task model describing crew interaction for several levels of automation

The model instance explicitly describes all crew activities, system interactions and demands for different automation levels. As can be seen in the model snapshot in, the detailed view for one complex mission task, in this example the reconnaissance of a helicopter flight route, shows that this task can be performed by support of three automation levels or performed manually (Fig. 5). For activity determination, observables are used. As described in [3], so called "evidences" (observable facts) are assigned in the model representation. Different observation channels with sensors like buttons, touch-sensitive displays and eye gaze tracking are used in this application and associated with modelled "Action" elements. By applying online activity determination, crew activity can be distinguished for different automation levels, and an alternative, advantageous automation level can be proposed by the associate system's workload projection if workload issues seem to occur.

4 Experimental Evaluation of Mental Workload

To investigate crew behavior in MUM-T missions, a dual-seat generic helicopter (HC) flight simulator, equipped with multi-touch displays and free configurable multi-function displays (MFD), was set up (Fig. 1). Graphical user interfaces for crew interaction with the automated reconnaissance system and visual reconnaissance data representation during flight missions were embedded in the helicopters MFDs (visible in Fig. 6).

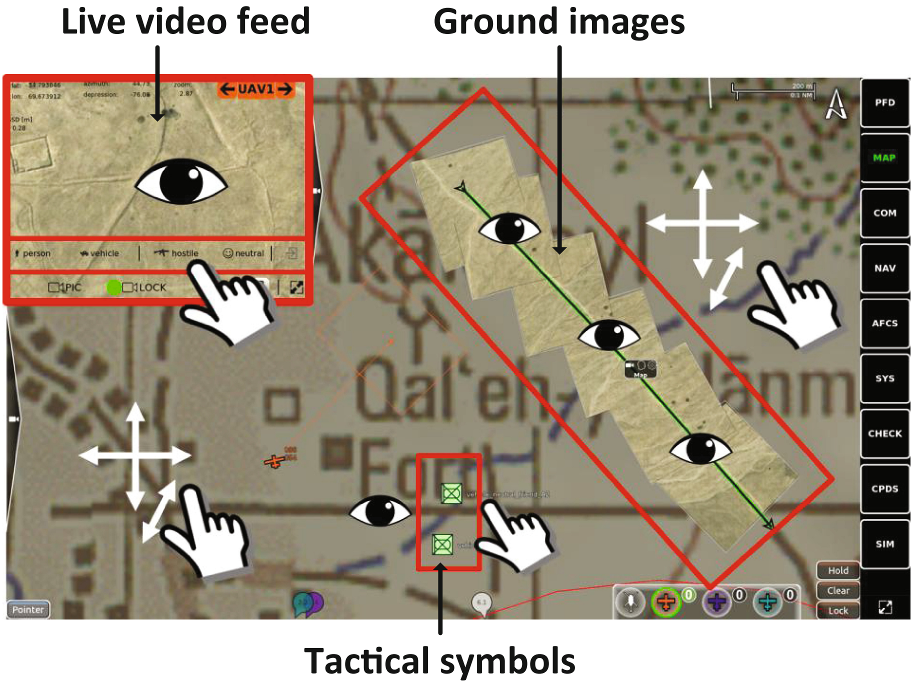


Fig. 6. User interface with annotations concerning different types of data representation and human interaction (Color figure online)

For crew observation, non-invasive, contact-less eye gaze measurement was applied. The used eye tracking system *smarteye pro* consists of four cameras per seat and provides functionality to capture, monitor and analyze the human operator's eye gaze movement. Measured values were current gaze positions on displays. Using a geometric model of the current MFD configuration, semantic references to graphical content currently shown on the surface is created. Furthermore, gaze tracking is complemented by synchronously capturing haptic interaction data to get more explicit evidences.

The goal is to determine human demand parameters [15] to be deposited in the task model described above. Therefore, a raw interaction dataset of visual and haptic interaction was gathered in human-in-the-loop experiments by crew observation sensors.

4.1 Interaction Dataset Recording

We defined and prepared separate use-cases of typical task constellations with several UAVs to be evaluated in the MUM–T mockup mission-simulator. For each use-case representing the task execution on several automation levels, we isolated phases containing the performance of reconnaissance task from the cockpit. Within the experiment, the reconnaissance automation levels were applied within a task-based-guidance concept [16]. A data recorder collected all user interactions with the running automation systems and reconnaissance result representation on the MFDs surfaces.

Figure 6 shows the user interface with observed and recorded interaction values and types, consisting of visual and haptic user interaction. Red borders in Fig. 6 show the interaction fields and data types of different visual data representation on different automation levels. Data representation comprises three different types of data; a live video feed, rectified and georeferenced ground images (image mosaicing) as well as tactical symbols.

Observable visual evidences are the operators gaze position on the live video feed, on the rectified ground images and on tactical map elements. Haptic evidences are user inputs by button presses and gestures on the map. These evidences are directly mapped to actions. For each automation level, different user demands exist. For specifying these demands from the raw data record, the data was analyzed and evaluated by an algorithm based method.

4.2 Derivation of MWL

To derive demand parameters of the raw data set, the two indicators time and interaction event amount were chosen. The factor of event quantity per observation time is the bandwidth of evidence observations. The derived demand values are normalized on the maximal human demand (*perception visual spatial* and *response manual*) during manual performance. For making different source data bandwidths comparable, the same amount of reconnaissance result to be presented was configured.

The following section illustrates the automated processing cycle for parameter extraction by pseudocode:

```
program demandExtraction
  for each observationInterval in observationIntervals
    for each interactionDataType in interactionDataTypes
      integrateInteractions()
      normalizeOnMaxDemand()
      fillCorrespondingDemandVectorElement()
    next
  next
end demandExtraction
```

4.3 Results

By evaluating the recorded dataset, demand components of information processing and response of the multiple resource model [15] were extracted successfully. For each automation level, the two most important components VS (*visual spatial*) and RM (*response manual*) could be determined.

Figures 7, 8, 9, 10 and 11 show the isolated interaction datasets of “complex” task [3] execution for all automation levels and one exemplary given single task of the task model instance. The corresponding demand parameters were reconstructed from the interaction dataset.

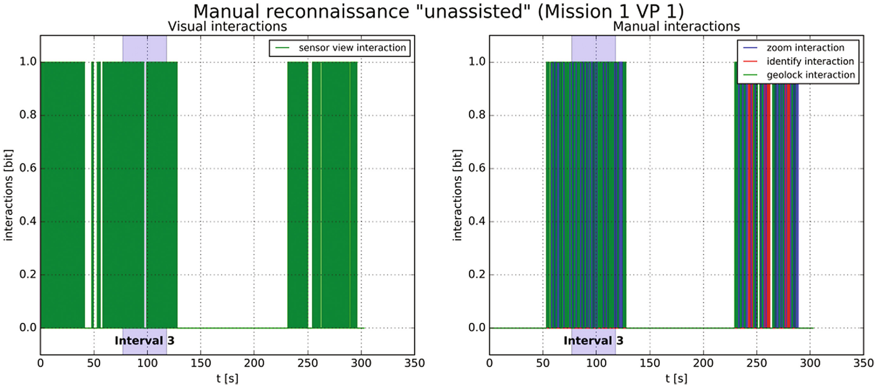


Fig. 7. Interaction dataset and observation interval for “unassisted” mode

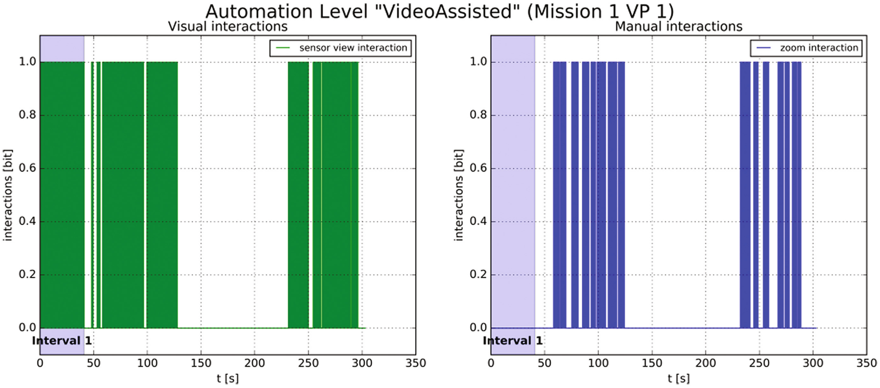


Fig. 8. Interaction dataset and observation interval for “Video Assisted” mode

The observation interval in Fig. 7 shows the interactions of “unassisted” manual reconnaissance. From this observation interval, the reference value RM for normalization was reconstructed. The operator’s haptic interaction activity is the highest

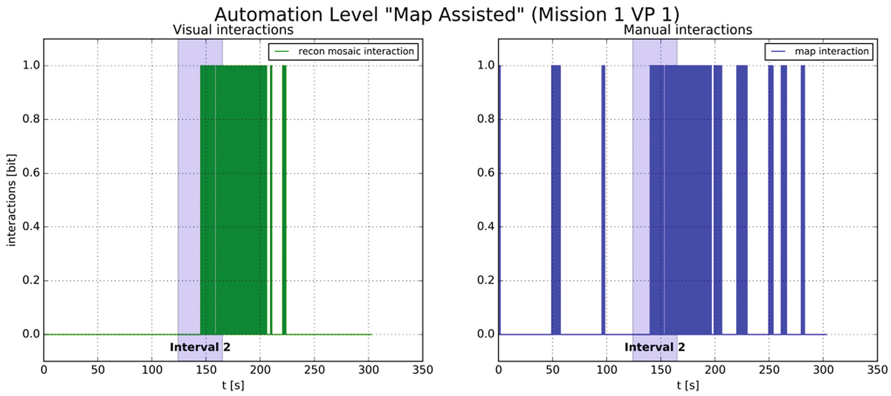


Fig. 9. Interaction dataset and observation interval for “Map Assisted” mode

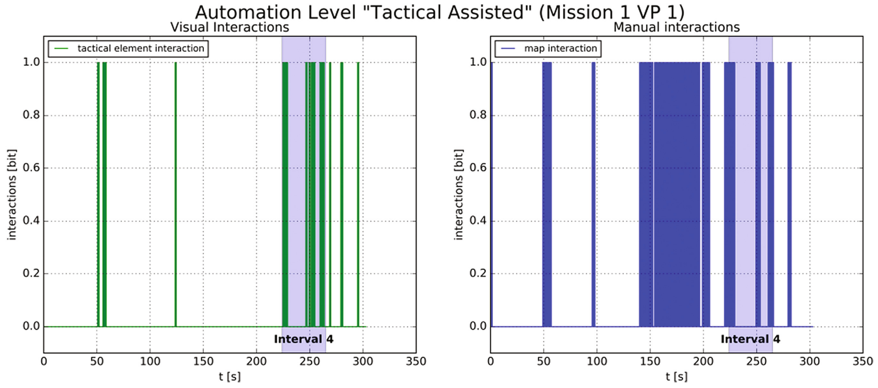


Fig. 10. Interaction dataset and observation interval for “Tactical Assisted” mode

possible, the sensor was guided fully manual and sensor data had to be evaluated visually.

```
+++++ Automation Level: "unassisted" ++++++
VS value: [0.9916710046850599]
RM value: [1.0]
```

Figure 8 shows the interactions of the “Video Assisted” automation level. From this observation interval, the reference value VS for normalization was reconstructed. The operator’s visual interaction activity is the highest possible, monitoring moving image is assumed as the most demanding visual activity. Sensor guidance was executed by automation which means that no manual interaction was required.

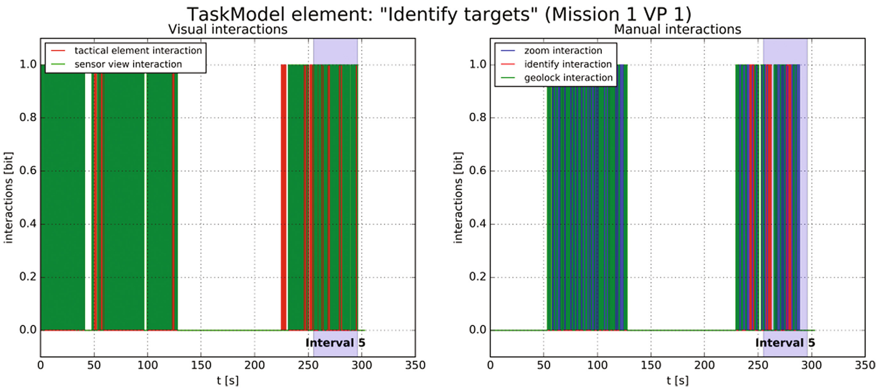


Fig. 11. Interaction dataset and observation interval for task "Identify targets"

```
+++++ Automation Level: "Video Assisted" ++++++
VS value: [1.0]
RM value: [0.0]
```

In the "Map Assisted" mode, the ground images had to be analyzed. The same amount of reconnaissance data could be evaluated in a shorter normative interaction time interval. Manual interaction occurred when shifting and zooming the ground image map. The extracted values were scaled to the VS and RM maximal values.

```
+++++ Automation Level: "Map Assisted" ++++++
VS value: [0.4809994794377928]
RM value: [0.37993527508090613]
```

The interactions in the "Tactical Assisted" mode (Fig. 10) occurred when the highest automation degree produced tactical elements on the map.

```
+++++ Automation Level: "Tactical Assisted" ++++++
VS value: [0.34096824570536177]
RM value: [0.21779935275080906]
```

Finally, the single task "Identify targets" was performed (Fig. 11). Three simulated objects had to be analyzed and identified in the live video feed.

```
+++++ Task: "Identify targets" ++++++
VS value: [0.9593961478396669]
RM value: [0.6210355987055016]
```

4.4 Result Interpretation and Reflection

By critically examining the demand parameter extraction and source dataset the fact became obvious, that there is no informational content about the demands of “information processing” components according to [15] (especially cognitive spatial) contained in the record. Another activity measurement type would be needed to retrieve such information content. The CS demand had to be determined by specialist knowledge applied in this domain. All the other information perception demand values did not occur in this task set (VV, AS, AV) and were set to 0, as well as the other response demand RV and the other processing demand CV.

In general, the applied method is able to fill the described parameters of larger models.

Analyzing systematic error sources, the experiment covered only several use-cases, not the complete task model parameter extraction. Therefore, small observation horizons were used that inherit the risk for high variance and normalizations become more prone to inaccuracy.

Operator observation processes, especially the gaze tracking method, is afflicted with measurement noise that may yield to wrong semantic associations of surface elements; such values were discarded in this experiment.

The goal of the campaign was to methodically determine specified model parts; the campaign and method is not representative in general. The results however show that the suggested automation levels reflect different user demands and the assistance approach is applicable for reduction of MWL.

5 Future Work

By connecting and linking the introduced subsystems, a software based chain to evaluate the sensor assistance concept was implemented in our MUM-T helicopter mission simulator. Future work comprises the application of all software modules in a full-mission scenario. A closed-loop-operation for functional demonstration of the holistic HC- associate system applying sensor automation with usage of the proposed sensor assistant system will be realized in the near future.

Experimental evaluation of the proposed concept as well as the effects on crew and mission performance by human-in-the-loop experiments with military transport helicopter crews is aspired.

References

1. Onken, R., Schulte, A.: System-Ergonomic Design of Cognitive Automation: Dual-Mode Cognitive Design of Vehicle Guidance and Control Work Systems. Springer Publishing Company Incorporated, Heidelberg (2012)
2. Strenzke, R., Uhrmann, J., Benzler, A., Maiwald, F., Rauschert, A., Schulte, A.: Managing cockpit crew excess task load in military manned-unmanned teaming missions by dual-mode cognitive automation approaches. In: AIAA Guidance, Navigation, and Control Conference, pp. 1–24 (2011)

3. Honecker, F., Brand, Y., Schulte, A.: A task-centered approach for workload-adaptive pilot associate systems. In: Proceedings of the 32nd Conference of the European Association for Aviation Psychology – Thinking High AND Low: Cognition and Decision Making in Aviation (2016)
4. Onken, R.: Funktionsverteilung Pilot-Maschine: Umsetzung von Grundlagenforderungen im Cockpitassistenzsystem CASSY. In: Gärtner, K.-P. (ed.) DGLR-Bericht 94-01, vol. 94–1, pp. 73–89. Deutsche Gesellschaft für Luft- und Raumfahrt, Berlin (1994)
5. Baker, A.L., Keebler, J.R.: Factors affecting performance of human-automation teams. Adv. Hum. Factors Robot. Unmanned Syst. **499**, 331–340 (2016)
6. Ruf, C., Stütz, P.: Model-driven sensor operation assistance for a transport helicopter crew in manned-unmanned teaming missions: selecting the automation level by machine decision-making. Adv. Hum. Factors Robot. Unmanned Syst. **499**, 253–265 (2016)
7. Hellert, C., Stütz, P.: Performance prediction and selection of aerial perception functions during UAV missions. In: AeroConf 2017 (2017, in press)
8. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. Hum. Factors J. Hum. Factors Ergon. Soc. **37**(2), 381–394 (1995)
9. Billings, C.E.: Aviation Automation: The Search for a Human-Centered Approach. Lawrence Erlbaum Associates, Mahwah (1997)
10. Wiener, E.L.: Human factors of advanced technology (glass cockpit) transport aircraft (1989)
11. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Factors J. Hum. Factors Ergon. Soc. **39**(2), 230–253 (1997)
12. Sheridan, T.B.: Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: distinctions and modes of adaptation. IEEE Trans. Syst. Man Cybern. - Part A Syst. Hum. **41**(4), 662–667 (2011)
13. Endsley, M.: The application of human factors to the development of expert systems for advanced cockpits. In: Human Factors Society 31st Annual Meeting, pp. 1388–1392 (1987)
14. Honecker, F., Schulte, A.: Konzept für eine automatische evidenzbasierte Online-Pilotenbeobachtung in bemannt-unbemannten Hubschraubermissionen. In 4. Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten – Verstehen, Beschreiben und Gestalten Kognitiver (Technischer) Systeme (2015)
15. Wickens, C.D.: Multiple resources and performance prediction. Theor. Issues Ergon. Sci. **3** (2), 159–177 (2002)
16. Uhrmann, J., Schulte, A.: Concept, design and evaluation of cognitive task-based UAV guidance. J. Adv. Intell. Syst. **5**(1) (2012)

Modeling of Performance Biases Induced by the Variance of Information Presentation to the Operator

Sen Tian(✉), Dan Huang, Lin Wang, and Shan Fu

School of Electronic Information and Electrical Engineering,
Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China
sjtutiansen@outlook.com

Abstract. The human factors have a significant impact on the flight performance. A computable human performance model (HPM) is proposed in this paper to simulate the operations and performances of a real human pilot. Incorporated with cognitive psychology theory, the proposed HPM attempts to explain the principles and mechanisms of the human factors such as the problem how the human operational time delay and errors were introduced. The proposed model contains a series of methods to expound the change of human time delay and human errors corresponding to information variations. Finally, a simulation experiment is carried out to study the impact of human time delay and decision-making biases on the flight performance.

Keywords: Human factors · Human performance model · Cognitive psychology · Human time delay · Decision-making biases

1 Introduction

Flight safety is one of the most significant issues in aviation. Various factors and the interaction among these factors may affect flight performance and safety [1]. Statistical data in recent years indicate among these factors the impact of human factors on flight safety issues is kept in high proportion. During the design of the aircraft, designer must consider human factors, but people usually find it difficult to predict and describe human behavior affected by human factors accurately. Thus, it is very important to construct practical methods to express and explain the impact of human factors for aircraft design, flight safety and so on.

In the last few decades, a wide range of techniques and approaches has been made by researchers to explain or predict human behavior. A crossover model was created by McRuer and Jex [2] to analyze the control quality of human. An optimal control pilot model was applied for different operator's performances by Kleinman etc. [3]. Later Hodgkinson [4] improved the optimal control model and add a compensatory pilot model to express adoption, composite application of gain and lag compensations. Recently, five cognitive modeling tools were applied to safety issues associated with pilot's operation and performance in the National Aeronautics and Space Administration Human Performance Modeling (NASA HPM) Project [5]. These five models

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 64–75, 2017.
DOI: 10.1007/978-3-319-58472-0_6

contain cognitive model, task network model and so on. All of these models can provide help to the human factors research at different levels. However, most of them can only reflect the impact of human factors on flight performance, while they can not give an answer to the formation mechanism and principle of human factors.

The main objective of this paper is to explore a method that can correctly explain the principles and mechanisms of the human factors such as the problem how the human operational time delay and errors were introduced, and reflect the impact of these human factors on flight performance by computer simulation. It has been proved that modeling is a good solution to this kind of problems [6]. Then, this paper attempts to build an accurate, applicable and computable HPM to simulate the operations and performances of a real human pilot. A human cognitive model [7] which contains three main components: information perception, decision making, and action execution is used as a basic structure of the HPM. Based on the cognitive psychology theory of information processing, a specific decision-making model in the HPM is proposed to express how the human pilot process the information obtained from outside world and make a decision of control rules. In order to improve computing efficiency, the HPM ignores other secondary human factors such as operation habits and mainly consider the human factors that have a greater impact on flight performance. In addition, this paper studies the problem that the variance of information will lead to human errors and the change of human time delay. A series of human time delay models and decision-making model are built to explain the problem. Finally, a simulation experiment is designed to validate the HPM. The simulation experiment is realized by the man–aircraft–environment (MAE) complex systems model [8], which makes it possible to generate a large number of flight data in a short time.

2 Specification of Human Performance Model

2.1 Overview of Human Factors in HPM

Among many human factors, human errors have the greatest impact on flight performance. Human errors has become an important, well-defined discipline and people has realized that human errors induced system failure is much more important and costly than the typical 100–500 ms delay observed in the reaction time study [9]. Such human factors often lead to the loss of the stability of the system. Human time delay and biases have second impact. Such human factors can affect the performance of the system such as the lag of the reaction and the decrease of the operation precision. Human habits such as operation habits, scan patterns have minimal impact on the system. So in this paper, the HPM is mainly concerned with human errors and human time delay while the other less affected human factors are ignored.

Human errors exist in both three components of cognitive process mentioned in the previous. However, in different component, the cause and the impact of human errors are different. About human errors, there are more systematic theory [10], but this paper does not consider too complex human errors, only consider the errors that have great impact and can be quantified accurately such as information missing (or overflow), decision-making mistakes, decision-making biases, the operation errors etc.

Like human errors, human time delay also exists throughout the cognitive process. In addition, human time delay has a coupling relationship with other human factors such as decision-making biases, operation error etc. The paper will explain the causes and effects of human errors and human time delay in different component in detail.

2.2 Structure of HPM

The behavior of pilots in the course of aircraft flight is similar to the process of human cognitive models. That is repeating the cycle process: get aircraft instrument information - determine the state of aircraft and environment – make decision – operation - get information. In the case of an aircraft system, the function of a human being is similar to an automatic control system. Every time a cognitive cycle is updated, the input of the controller is updated and the corresponding control variables are given by human, and then the human neuromuscular system is applied to the specific aircraft operation. However, different from the general controller, in many flight scenarios and flight tasks, two forms of control methods continuous control and discrete control coexist, and for different control objects, the proportion of continuous control and discrete control is different. For example, the control of the rudder and the elevator in the cruise task is continuous during the control period, which is similar to servo control. However, in the process of the task, if the other operations need to be occupied or the change of flight state leads to the change of control rules, the continuous control will be interrupted. The control of flaps and so on is discrete. Only certain conditions such as reaching a certain height are triggered will lead to changes of controlled variable, and once a change occurs, the control variable will remain constant over a long period of time until the next condition triggers. For the above two different situations, the controlled variable and the corresponding models are divided into two categories. One is the continuous control, of which the corresponding controlled variables include the elevator, rudder, throttle, etc. The other is the discrete control, of which the corresponding controlled variables include flaps, spoilers, landing gear etc. Since the rules have been determined, the time delay of the continuous control model in a single cognitive cycle is short. For the discrete control model, because people need to determine the state of aircraft and make complex decisions, the time delay is long. A special case is in continuous control model, the time delay will be longer if the state of the aircraft is changed and the current control rules need to be changed (Fig. 1).

2.3 Three Components of Human Performance Models

Information Perception

The component of information perception refers to the process of selective access to potentially useful information from the outside world. In the flight, the pilot can obtain visual, auditory, somatosensory and other information from the outside world, but only some of the information is potentially useful for pilots. The information is potentially useful is based on a subjective judgment of the pilot. The pilot has a conscious selective access to information to reduce unnecessary time consumption. For pilots, potentially

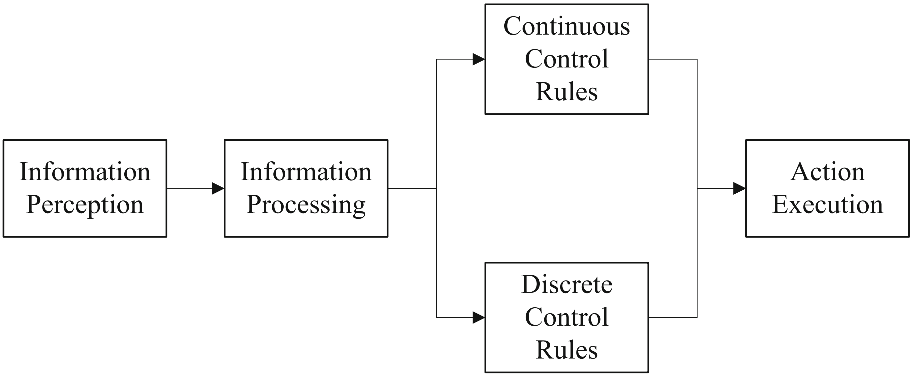


Fig. 1. The structure of HPM

useful information include aircraft instrument information (aircraft status), window visual information, voice indication information, somatosensory information (including spatial location, acceleration, etc.). In the subsequent experimental verification process, because it is difficult to quantify the later information in the program, only aircraft instrument information is taken into account.

The time required to obtain the instrument information is calculated by the empirical formula obtained by the serial self-terminating search (SSTS) model [11]. The SSTS model considers that the search time ST required to search for N instruments is proportional to N:

$$ST = a_p + bN \tag{1}$$

Where a_p is a constant used to describe the inherent time of the search, including head movement time, attention shift time and so on, b is search and dwell time of a single instrument and the different instruments have different b. The values of these two constants are empirical values obtained by experiments. ST is equal to the time delay of the information perception in single cycle. For all discrete controlled variables, this time delay is present throughout the flight, because pilot need to obtain enough useful information to determine the status of aircraft and environment. For the continuous controlled variables, because of the small number of instruments to be observed, the information acquisition and the operation can be carried out simultaneously. In this case time delay is a relatively small value. No matter what kind of control, the number of instruments required to search N is determined by the current flight mission and flight phase. The main human errors in the information acquisition is the omission of information and the false recognition of information content. This part of the error will have an impact on the subsequent decision making process, so it will be described later in detail.

Decision-Making

Information Process. Decision-making contains the front information processing section. Information processing is to get some useful information from the part of information perception then select and fuse the information in order to get a complete description of the aircraft states. The selection of information is to select the information that is required for the operation to be performed from all the useful information. For example, the pilot may repeatedly check the airspeed indicator, altimeter, flight attitude indicator etc. However, in the implementation of the left turn 90 degrees flight mission, the operation requires only a small amount of information such as yaw angle, speed, etc. For information fusion, it is difficult to describe exactly how people integrate information into a whole concept, but the process of information fusion does exist and work. This paper does not consider the specific mechanism of information fusion, but in the process of decision-making, the multi information is combined into the judgment condition as the form of information fusion. Through the experiment, it is difficult to distinguish the time delay of information processing and the time delay of decision-making, and only their common time delay can be obtained. Because the information processing does not need to carry on the complex logical inference, and has the subconscious participation, the actual process is quite rapid. A very small constant is used to express the time delay.

Decision-making. The decision-making process includes two meanings: the generation of decision rules and the choice of decision rules. The generation of decision rules is a complex process. There is still no universal theory in the fields of biomedicine, psychology and so on to explain how the human produce ideas. Therefore, the HPM model does not show how these control rules are generated. The HPM use a method similar to expert control, which write the standard control rules for different flight scenes according to the experience of experts. The flight check list, flight crew operating manual (FCOM), and so on, are use this method to give an answer to the pilot's handling of the aircraft in the standard or some emergency situations. These rules that have been written in advance belong to the normative decision making. In fact, the decision made by pilots in most of the time during flight belongs to this type. In the remaining situation that has not been considered in advance, pilots need to determine the situation and generate the decision rules. This type of decision belongs to heuristic decision making [12]. Although this decision, especially in the face of unexpected situations will be more important (this is the main reason that the current computer can not completely replace the pilots), as mentioned earlier, it is impossible to accurately describe the process. Therefore, heuristics is not considered in this paper. The core problem of normative decision making is how to make decision or choice according to the best reference frame. In this paper, a task network model [13] is used to simulate human decision making. The whole flight can be divided into takeoff, cruise and other stages, and each stage can be divided into many specific flight tasks. The process of the choice of decision rules is to determine which task needs to be executed by the aircraft status and the overall flight plan, then in each task, it is necessary to determine which control rules need to be selected to control the aircraft in accordance with the more specific aircraft status information. The entire process is similar to borrowing from the library, the control rules written in advance are classified by the actions, tasks, stages, etc.

The decision making process of pilots is to determine which rule applies to the current state of the aircraft through a series of If… Else… conditions.

For the machine, it is possible to check all the necessary information in each cycle (the machine refresh clock) and then give the value of all the controlled variables. Because of the limited resources and capabilities, it is impossible for human to do a large amount of information processing and decision-making behavior in a short period of time (the order of the machine refresh clock). Human make decision in a similar way to Time Division Multiplexing. For two pilots, tasks of continuous control and discrete control can be assigned to the main pilot and the copilot respectively. For one pilot, if the continuous control has confirmed the rules and maintained this rule for a longer period of time (for example, manual control of aircraft Cruise), it will only take up fewer resources. The pilot can execute other controls at the same time. On the contrary, because of the complexity of the state determining and decision making, only one discrete control can be carried out at the same time. If every time to obtain all useful information and make decision, the time delay will be too large. In the real situation, the pilot will not do so. This paper presents a form of regular and trigger to simulate the real process. The pilots mainly execute continuous control during the flight. The main concern of pilots is the information that is closely related to the continuous control such as speed, height, etc. In this case, the cycle time of each HPM is very short. Every once in a while (This time can be modified according to the actual operation of the pilot, in this paper it is set to 10 times the period of the continuous control in HPM) the pilot checks all the useful information to determine if any other minor continuous or discrete control is needed. This cycle corresponds to a longer time delay. In addition, at any time, once other control rules is triggered by a certain information, all operations are interrupted. If a new operation (other controls) is added, which leads to the amount of control reaches the upper limit of the pilot's control, one of the previous operations will be suspended until the new operation is returned (Fig. 2).

The Human Time Delay of Decision-Making. The human time delay of decision making is approximately calculated by hick-hyman-law [14, 15]. Hick-hyman-law is a formula used to calculate the choice reaction time (RT). The choice reaction time can be regarded as the time required for human decision making (here is the choice of decision rules). The Law points out that the choice reaction time is proportional to the amount of information contained in the choice:

$$RT = a + bH \tag{2}$$

Where a and b are both constant obtained by experiment. *a* describes the sum of processing latencies that are independent of uncertainty reduction, such as the amount of time spent on stimulus encoding and response execution, *H* is the average amount of information for a certain rule calculated by the formula:

$$H_{avg} = \sum_{i=1}^{n} P_i [log_2 \left(\frac{1}{P_i} \right)] \tag{3}$$

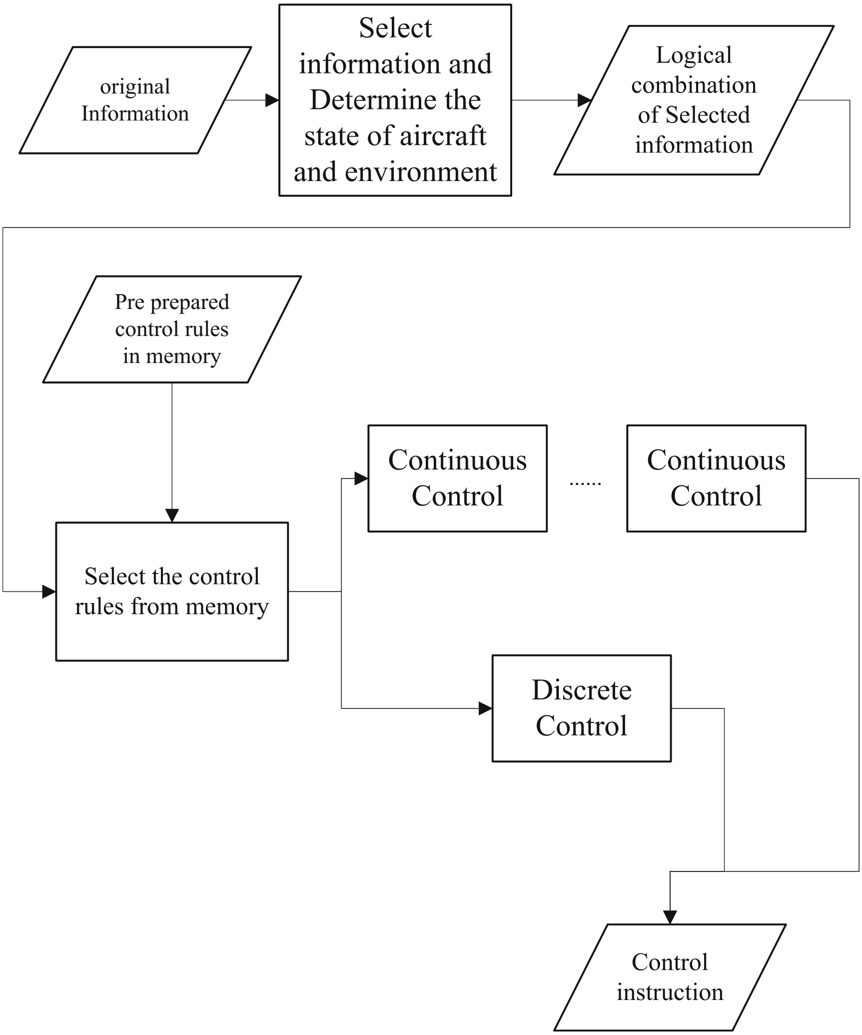


Fig. 2. The process of decision-making

Where P_i is the probability of the occurrence of a certain information in the rule. The sum of the probabilities of all information used to determine this rule is 1. The time delay of continuous control and discrete control in decision-making can be calculated by Hick-hyman-law. Specially in the continuous control, due to the existence of a certain period of time the control rules remain unchanged, then the corresponding decision-making time delay is 0.

The Human Errors of Decision Making. Although most of the time people are expected to be able to make accurate decisions in accordance with the best reference frame, people can not correctly perform the operation due to some subjective or objective reasons, which means the human errors and biases. This paper mainly studies the human errors induced by the change of information. Where the change of information is refer to the change of the amount of information (missing or overflow). When the information is missing, it will lead to the failure to select the right rules which will cause decision-making biases or errors. This is simulated in HPM as follows: If at a certain moment, the combination of the condition of the rules that should be used is A&B&C (A, B, C are conditions containing different information, & refers to a certain logical operation) but the information that has been obtained is only A&B. Then a rule will be randomly selected from all the rules of the A&B condition. If the information overflow, although it will not lead to the choice of wrong rules, the extra time cost of information perception will also have an impact on the performance.

Action Execution

Action execution mainly involves continuous control. Because the discrete controlled variables such as flaps, landing gear are generally switch control, only the time delay of which has to be considered. In the paper, it is an empirical value. For continuous control, different control methods and different control precision etc. will affect the results. In this paper, the control rules are not written for the whole flight but for the stage final approach and landing. During the final approach and landing, the continuous controlled variables mainly includes the three control objects, the elevator, rudder and throttle. The PI control methods is used for a simple simulation of human's control methods. For example, the control rules of the elevator δe in the approaching process is:

$$\delta e = (K_{pv} + K_{iv} * T_{delay}) * \delta V + (K_{pp} + K_{ip} * T_{delay}) * \delta P \tag{4}$$

Where K_{pv} is the proportional coefficient of velocity deviation, K_{iv} is the integral coefficient of velocity deviation, T_{delay} is the time delay of single cycle. δV is the deviation between the expected speed and the instrument speed. Similarly, the latter part of the formula is the coefficient of the position and position deviation.

The time delay of action execution is related to the operation precision. According to speed-accuracy operating characteristic (SAOC) [16], there is a linear relationship between the logarithm of operation time and operation accuracy:

$$\log\left[\frac{P(true)}{P(false)}\right] \propto T \tag{5}$$

Where $P(true)$ is the probability of true operation and $P(false)$ is the probability of false. They meet $P(false) + P(true) = 1$. The operation accuracy is expressed by a random number. For example, if the elevator operation accuracy is 95%, the relationship between output δe_{actual} and expected output δe_{expect} is:

$$\delta e_{actual} = \delta e_{expect} * random(0.95, 1.05) \tag{6}$$

Where $random(0.95, 1.05)$ is a random number of 0.95 to 1.05.

3 Simulation Results and Discussion

3.1 Simulation Experiment Configuration

Simulation experiment is based on the MAE complex systems model. The model includes a human model, an environment model and an aircraft model. Human model is realized by previously mentioned HPM. The specification of HPM time delay parameters is listed in Table 1:

Table 1. HPM time delay parameters

Time delay of Information perception	
ap	N(0.2,0.04)
b	N(0.1,0.02)
Time delay of decision making	
a	N(0.2,0.04)
b	N(0.1,0.02)
Time delay of action execution	
a	N(0.5,0.1)
Precision	0.95

Where N(0.2,0.04) refers to a random number according to the standard normal distribution, whose expectation is 0.2, variance is 0.04.

The environment model includes an atmosphere model and a wind speed model. They are based on the international standard atmosphere (ISA). The aircraft model is implemented by a six-degree-of freedom (6-DOF) aircraft flight dynamics model [17]. The parameters of aircraft is based on the dynamic model of Boeing 747-400 [18], the specification is listed in Table 2:

Table 2. The parameters of Boeing 747-400

Span	59.74 m
Chord	8.32 m
Reference area	510.97 m^2
Mass	288775 kg
Max propulsion	28803 kg*4
Moment of Ineria	
Ixx	24675887 kg*m^2
Iyy	44877574 kg*m^2
Izz	67384152 kg*m^2
Ixz	1315143 kg*m^2

Table 3. Initial status of aircraft

Initial location	(0,0,1000) m
Initial velocity	(100,0,0) m/s
Initial attitude	(0,0,0) rad
Initial angular velocity	(0,0,0) rad/s
Initial angle of attack	0 rad

The scenario of simulation experiment is set to a final approach and landing mission. In this mission, aircraft is expected to track the flight path with down slope equals to −1:29. The initial status of aircraft are listed in Table 3 and the expected final status of aircraft are listed in Table 4:

Table 4. Expected final status of aircraft

Final location	(29000,0,0) m
Final velocity	(72.27,0,3) m/s
Final attitude	(0,0.01,0) rad
Final angular velocity	(0,0,0) rad/s
Final angle of attack	0.05 rad

Final velocity and touchdown offset (final location) are selected to reflect the flight performance. If the final vertical velocity is greater than 3 m/s, it is considered to be hard landing. If the final airspeed is less than 120 Knot, it is considered to be stalled. The smaller touchdown offset means that the better flight performance.

Four types of experiments are designed as Table 5:

Table 5. Types of experiments

	Type of information obtained	Types of control rules
Type 1	Position, velocity	Position, velocity
Type 2	Position, velocity, acceleration	Position, velocity
Type 3	Position, velocity, acceleration	Position, velocity, acceleration
Type 4	Position, velocity (randomly omit)	Position, velocity

In type 1, pilot is concerned with position and velocity information for continuous control. And only the rules of position and velocity have been written in the continuous control rules. In type 2, pilot is concerned with acceleration information in addition to position and velocity information. The continuous control rules of type 2 is the same as type 1, which is equivalent to completely ignoring acceleration information. In type 3, pilot is concerned with position, velocity and acceleration information. The continuous control rules of type 3 include the rules of position, velocity and acceleration (The acceleration rule is added on the original rule, and the logic structure of the original rule will not be changed). In type 4, pilot is concerned with position and velocity information. But sometimes the pilot omits the velocity information (a random probability function). The continuous control rules of type 4 is the same as type 1.

3.2 Results and Discussion

Because the time delay, decision rules choice and operation precision all contain uncertainty, each simulation result is different from the other. Each type of experiment was carried out 500 times. The Table 6 show the results of four types of experiments:

Table 6. The results of experiment

	Time delay per cycle	Stalled	Hard landing
Type 1	3.2 s (average value)	0	0
Type 2	3.7 s	0	0
Type 3	3.7 s	0	0
Type 4	3.2 s	0	32%

	Touchdown offset						
	>10 m	>20 m	>30 m	>40 m	>50 m	>60 m	>70 m
Type 1	69%	52%	38%	28%	10%	0%	0%
Type 2	74%	56%	43%	33%	15%	6%	0%
Type 3	74%	57%	41%	29%	19%	3%	0%
Type 4	78%	72%	72%	72%	70%	16%	0%

The results show that the final velocity of all the experiments except type 4 is satisfied. The average human time delay of type 1, type 4 are the same and that of type 2, type 3 are the same. The average time delay of type 1 is less than that of type 2. It is obvious that the increase in the amount of information will lead to an increase in time delay. Type 2 contrasts with type 1, the touchdown offset has increased due to the increase of the amount of information. The experimental results of type 2 and type 3 are similar. It can be seen that the control rules corresponding to the acceleration information have little impact on the flight performance. Type 4 contrasts with type 1, the touchdown offset has increased obviously, which shows that velocity information has larger impact on flight performance than the acceleration information.

4 Conclusions

Based on cognitive psychology, this paper constructs a human performance model containing human time delay, human decision biases, operation deviation etc. The impact of human factors and the interactions between them on flight performance is explained by the HPM. The HPM shows that there is a close relationship between human time delay and the complexity of decision-making and there is a coupling between human time delay and operation precision. The pilot controls the aircraft by continuous and discrete control. The decision-making bias occurs mainly in continuous control and it is mainly caused by the change of information. A serious information omission will lead to decision-making errors, which has a great impact on flight safety.

References

1. Salas, E., Dan, M., Curtis, M.: Chapter 1 - Human factors in aviation: an overview. In: Human Factors in Aviation, pp. 3–19. Elsevier Inc. (2010)
2. Mcruer, D.T., Jex, H.R.: A review of quasi-linear pilot models. IEEE Trans. Hum. Factors Electron. **HFE-8**(3), 231–249 (1967)
3. Kleinman, D.L., Baron, S., Levison, W.H.: An optimal control model of human response part I: theory and validation. Automatica **6**(3), 357–369 (1970)
4. Hodgkinson, J.: Aircraft handling qualities, pp. 100–130. AIAA, Reston (1999)
5. Foyle, D.C., Hooey, B.L.: Human Performance Modeling in Aviation. CRC Press, Boca Raton (2007)
6. Byrne, M.D., Pew, R.W.: Chapter 8 A history and primer of human performance modeling. Rev. Hum. Factors Ergon. **5**(1), 225–263 (2009)
7. Yang, C., et al.: Human factors quantification via boundary identification of flight performance margin. Chin. J. Aeronaut. **27**(4), 977–985 (2014)
8. Yin, T., Huang, D., Fu, S.: Human strategy modeling via critical coupling of multiple characteristic patterns. J. Aircraft **52**(2), 617–627 (2015)
9. Wickens, C.D., Hollands, J.G., Parasuraman, R., Banbury, S.: Engineering Psychology and Human Performance, 4th edn., pp. 218–230. Pearson, London (2012)
10. Salvendy, G.: Handbook of Human Factors and Ergonomics, 3rd edn. (2006)
11. Sternberg, S.: High-speed scanning in human memory. Science **153**(3736), 652–654 (1966)
12. Kahneman, D., Klein, G.: Conditions for intuitive expertise. Am. Psychol. **64**(6), 515–526 (2009)
13. Schweickert, R., Fisher, D.L., Proctor, R.W.: Steps toward building mathematical and computer models from cognitive task analyses. Hum. Factors J. Hum. Factors Ergon. Soc. **45**(1), 77–103 (2003)
14. Hick, W.E.: On the rate of gain of information. Q. J. Exp. Psychol. **4**(1), 11–26 (1952)
15. Hyman, R.: Stimulus information as a determinant of reaction time. J. Exp. Psychol. **45**(3), 188–196 (1953)
16. Drury, G.G.: The speed—accuracy trade-off in industry. Ergonomics **37**(4), 747–763 (1994)
17. Zipfel, P.H.: Modeling and Simulation of Aerospace Vehicle Dynamics (AIAA Education), pp. 367–480. American Institute of Aeronautics and Astronautics, Reston (2003)
18. Roskam, J.: Airplane Flight Dynamics and Automatic Flight Controls, Part I & II. Part I Design Analysis & Research Corporation (2001)

Can Fixation Frequency Be Used to Assess Pilots' Mental Workload During Taxiing?

Xiaoyan Zhang[1,2], Hongjun Xue[2(✉)], Xingda Qu[1], and Tao Li[2]

[1] Institute of Human Factors and Ergonomics, Shenzhen University,
Shenzhen 518060, China
zxyliuyan@163.com, quxd@szu.edu.cn
[2] School of Aeronautics, Northwestern Polytechnical University,
Xian 710072, China
xuehj@nwpu.edu.cn, 603649479@qq.com

Abstract. Increased mental workload is often associated with increased risk of pilots' operational errors. Mental workload is a multifactor problem. The paper aims at defining an appropriate indicator for pilot mental workload, under the assumption that vision is the single most important source of information for the pilot. We design a series of experiments about the task difficulty, which have different visibility, taxiing length and turns. The results showed that the fixation frequency did not increase with the task difficulty. Fixation frequency is not sensitive with the mental workload. This study concludes that the Fixation frequency cannot be used as the evaluation criteria of mental workload. But the results may be caused by small sample sizes (three pilots).

Keywords: Mental workload · Pilot · Fixation frequency · Fixation · Eye movement

1 Introduction

Pilots need to perform various visually demanding tasks in flight, such as monitoring various displays in the cockpit, navigating, and cockpit reconfiguration [1]. The culmination of the required tasks, the difficult operating environment, and the lack of aiding technologies result in a very high-workload of their visual system, which is very likely to induce operational errors [2]. According to the statistical summary of commercial jet airplane accidents reported by Boeing, human errors due to the overloaded visual system accounted for nearly 90% aviation accidents [3]. Thus, there is an urgent need to control pilots' operational errors induced by the overloaded visual system.

Subjective-ratings and Physiological measurements are two traditional methods used to assess inflight pilot workload. Notable Subjective-ratings currently employed are NASA-TLX (subject demand assessment) [4, 5], a Bedford workload scale, SART (a situational awareness assessment analysis questionnaire), and the Modified Cooper-Harper Rating Scale is four generally used subjective method [5]. But Subjective measures have some limitations. The probes of questionnaires confused pilots easily, and they always require a conscious, subjective response from the pilots to ensure the accuracy [6]. Ahlstrom et al. concluded that eye movement activity can

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 76–84, 2017.
DOI: 10.1007/978-3-319-58472-0_7

provide a sensitive measure of controller workload, but subjective ratings might not capture more transient fluctuations in workload levels during system or display interactions [7]. Therefore, objective, physiological measures would be preferable and become a key issue. Results from previous studies report that eye movement activity, heart parameters, and respiration, EEG signals are all sensitive to mental workload.

The blink of the eye is believed to be a good indicator of both fatigue and workload [7–9]. Blink Rate (BR) increases as a function of time on task (TOT), and blink duration (BD) decrease as visual workload increases. Besides eye blink activity, some researchers found that pupil size can also be used to assess workload. Pupil dilates most because of mental effort. Pupil dilation has been successfully used for distinguishing different levels of difficulty of various cognitive tasks.

The paper aims at defining an appropriate indicator for driver visual workload, under the assumption that vision is the single most important source of information for the pilot. Fixation frequency had been proposed to assess mental workload of pilot during taxiing. We design a series of experiments about the task difficulty, which have different visibility, taxiing length and turns. During the experiment, the eye movement data had been collected, and analyzed.

2 Methods

2.1 Subject

Three aircrews, each consisting of one captain and one FO (the First Officer) from Shanghai Eastern Flight Training CO., LTD. who flew the Airbus320, participated in the study. The participants are all male. The information of the three captains executing the experiment are given in Table 1.

Table 1. The information of three captains

No.	1	2	3
Flight hours (h)	6500	15000	20000
Gender	Male	Male	Male

2.2 Tasks

The trial is conducted in Airbus 320 simulator (shown in Fig. 1) in Shanghai Eastern Flight Training CO., LTD. In all scenarios, the pilot is asked to land from the last approach point and taxiing to the gate successfully according to the flight plan established before. Two variables of interest were investigated in the study: visibility and taxiing way.

Visibility. Trials were conducted in three different visibilities shown in Fig. 2. First is clear day. Pilot has unlimited visibility and the trial conducted in Visual Meteorological Condition (VMC). Second is night. The airport is under night mode and pilot should flight in Instrument meteorological Condition (IMC). To investigate the visual performance under low visibility, the runway visibility is 200 m, which is the lowest visibility according to the flight rules, and pilot should flight in IMC.

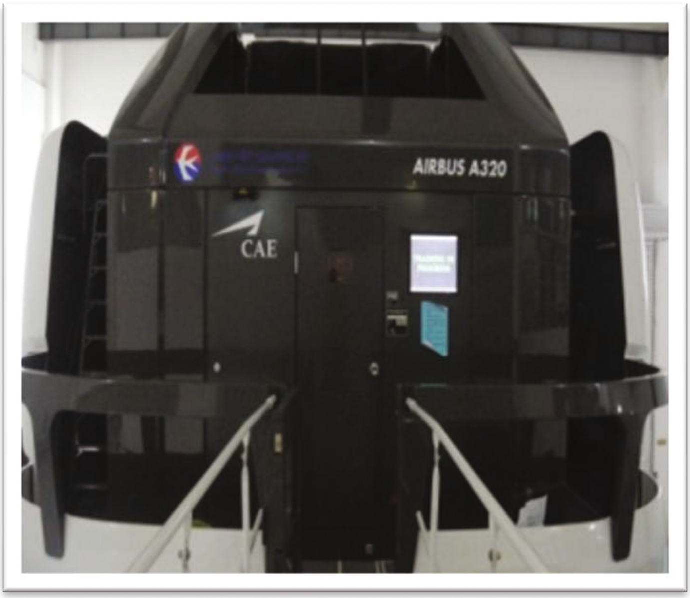
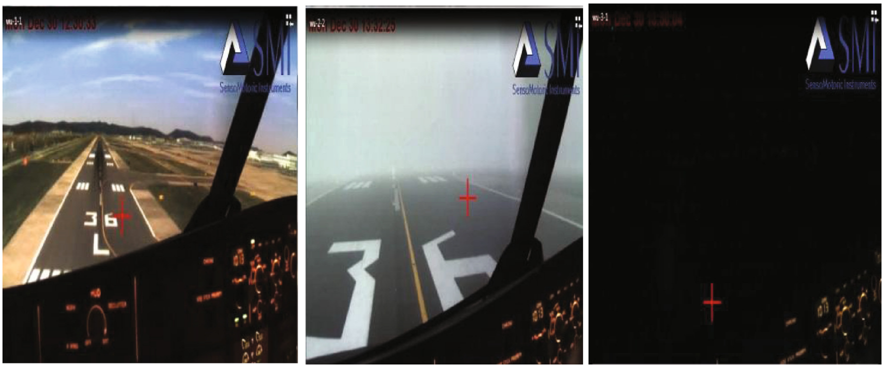


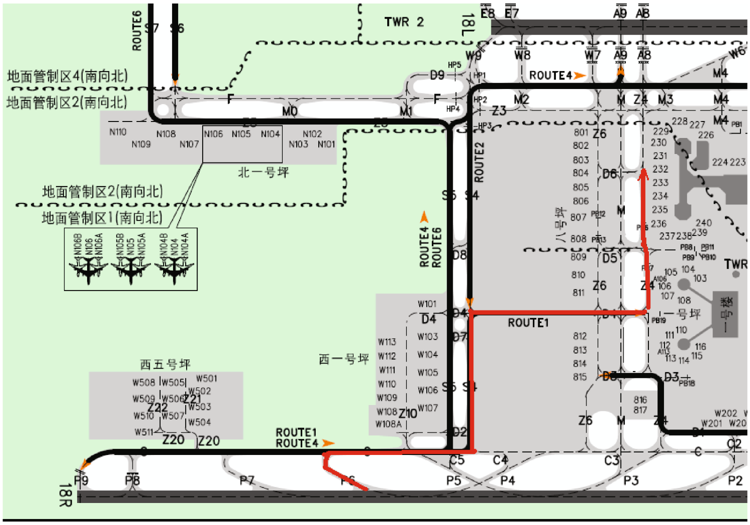
Fig. 1. Simulator used in the experiment



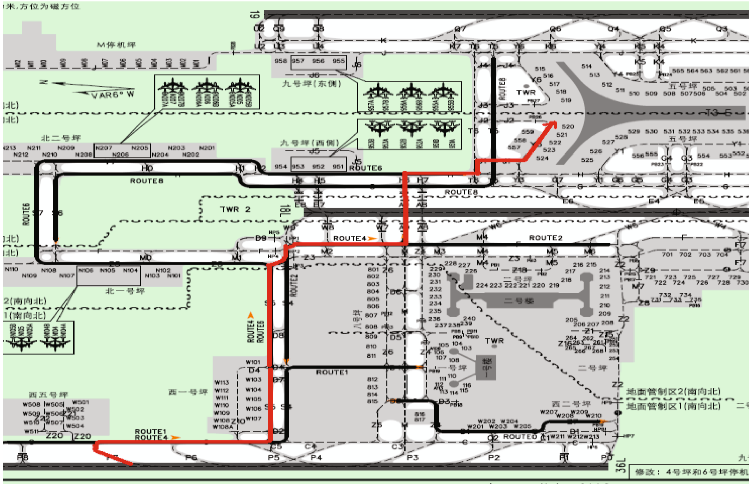
(a) Clear day (b) Low visibility III (c) Clear day, night

Fig. 2. Three different visibilities

Taxiing Way. There are two different difficulty taxiing ways from Beijing international airport shown in Fig. 3. The two has different number of turns and length. Besides, the difficult taxiing way will across runway, which needs the pilot to wait for a runway clearance from ATC (Air Traffic Control). The two different taxiing ways are shown in Table 2.



(a) Runway 36L to 232 gate (simple one)



(b) Runway 36L to 520 gate (difficult one)

Fig. 3. Two different taxiing ways in Beijing international airport

Table 2. Two different taxiing ways

	The number of turns	Length (km)
Simple	10	10.5
Difficulty	18	19

2.3 Variable Design

The experiment is in-subject design, and each group should flight six times. The order of the blocks were clear day & simple way(CS)- clear day & difficulty way(CD) -night & simple way (NS)-night & difficulty way (ND)- low visibility & simple way(LS) - low visibility & difficulty way (LD).

2.4 Experimental Procedure

In all scenarios, the participants were asked to land from the last approach point and taxiing to the gate successfully according to the flight plan established before.

At the beginning, the flying instructor communicated the flight plan including landing airport, runway, taxiing plan and airport gate with the aircrews as given. And then the instructor set the airplane to the last approach point. The participant's task, then, was to arm the autopilots, set the altitude, and engage landing gear, and altitude settings. After all the configurations done, eye tracker (SMI iView X HED) would be took on by the captain. A five-point calibration was made for each participant. Then the flight instructor activates the simulator, and the trial begins. Pilots should take full manual control during the trial.

2.5 Eye Tracker

The eye movement data had been collected by SMI head- mounted eye tracker shown in Fig. 4. The sample frequency is 240 Hz, and accuracy is 0.01.

Fig. 4. Pilot with the eyetracker in the experiment

2.6 AoI Designs

The definition of Area of Interests (AoIs) was directly connected to the pilot fixation pattern. AoIs were defined throughout the cockpit, namely electronic centralized aircraft monitoring display (EVAM), the navigation display (ND), primary flight display (PFD), and the centerline of taxiing way out of the window (OTWM), the left signs besides the taxiing way (OTWL) and the right signs besides the taxiing way (OTWR). The division of AoI is shown in Fig. 5.

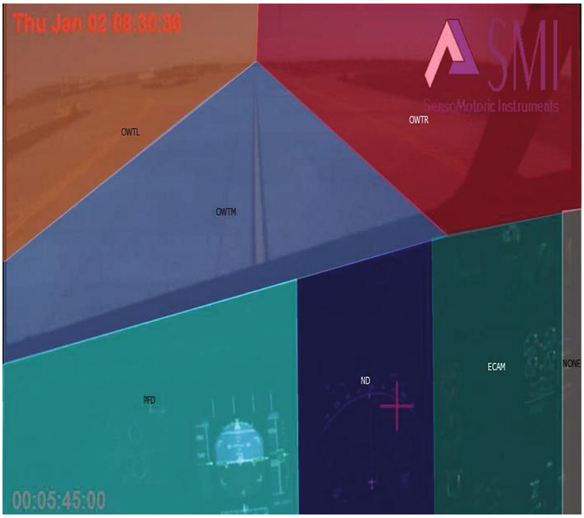


Fig. 5. Division of AoIs

3 Results

The F_f (Fixation frequency) of different pilots in the experiment is shown in Tables 3, 4 and 5.

Table 3. F_f of the pilot No. 1

Item	Scenario					
	CS	CD	NS	ND	LS	LD
Number of fixations	473	707	583	832	500	668
Fixation time(s)	174.364	236.915	244.581	316.104	179.294	259.848
F_f	2.7	3.0	2.4	2.6	2.8	2.6

Table 4. F_f of the pilot No. 2

Item	Scenario					
	CS	CD	NS	ND	LS	LD
Number of fixations	509	674	576	659	560	569
Fixation time(s)	176.662	237.538	230.793	255.806	193.267	214.546
F_f	2.9	2.8	2.5	2.6	2.9	2.6

Table 5. F_f of the pilot No. 3

Item	Scenario					
	CS	CD	NS	ND	LS	LD
Number of fixations	465	617	540	684	427	615
Fixation time(s)	175.401	205.520	190.930	256.580	147.775	235.186
F_f	2.7	3.0	2.8	2.7	2.9	2.6

A significant effect was found between scenes by different flight experience pilots, $F(5,17) = 5.365$, $P < 0.05$. However, there is not significant effect between different flight experience pilots. The F_f didn't changed regularly with the difficulty of taxiing way or the visibility.

4 Discussion

The aim of this paper was to contribute to understanding the pilot's fixation frequency is an effective indicator of mental workload. Due to the difficulty in recruiting professional pilots, the sample size of the present study was relatively small. This could explain why we did not find many significant differences between the experience conditions and visibility conditions.

Besides the statistical significant difference discussed above, we also found some interesting trends for the factors of interests shown in Fig. 6. The result showed that the F_f is significant different among different scenes. But the result can not be used as the evidence that Ff can be used to assess the workload since F_f is not changed regularly. F_f is higher in simple taxiing way than in difficult one, while it is smaller in worse visibility than in better one. Environment factors were also found to have effects on mental workload. McCann et al. (1997) reported that the low-visibility and night conditions contributed to increased workload [9]. Foyle (1996) also reported that a variety of environmental conditions (airport complexity, and visibility) could affect the workload [10]. But F_f did not changed regularly with the environment changed. Besides, experience is one of the factors that may affect mental workload. Some researchers reported that expert and non-expert pilots have different physiological responses to tasks at different difficulty levels [9, 12–18]. The value of F_f has no significant effect among different pilots.

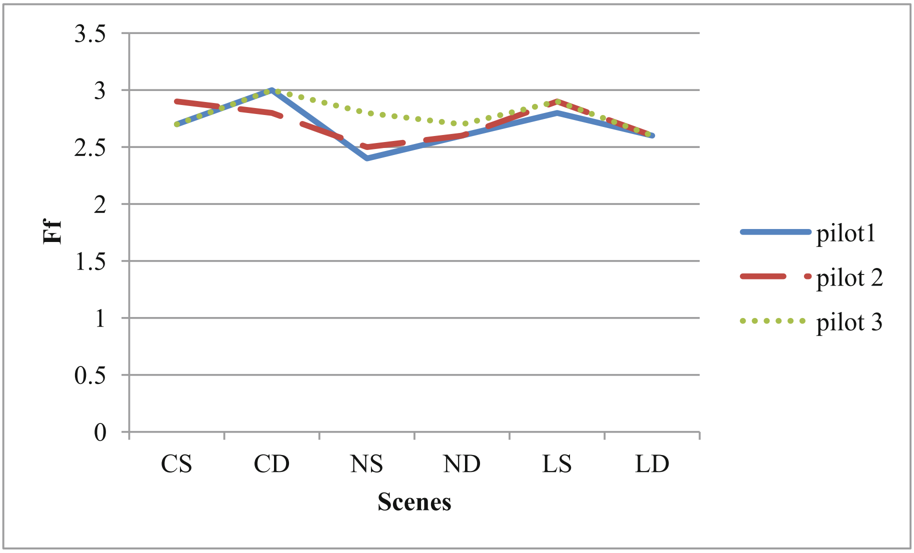


Fig. 6. F_f by scene and pilot

5 Conclusions

Interestingly, there was no significate effect found in expertise level and visibility condition in fixation frequency. We did not confirmed that the fixation frequency of pilots was an effective method to assess pilot mental workload. Although a significant effect had been found among different scenes, but it changed irregularly. The change of fixation frequency can neither indicate the change of mental workload influencing by the task difficult nor by the pilot experience. The lack of significance in this main effect could be due to small sample sizes (three pilots). For the future study, we should recruit more pilots to validate that whether fixation frequency could be an effective indicator to evaluate mental workload among different pilots or different visibilities.

References

1. Haslbeck, A., Bengler, K.: Pilots' gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task. Cogn. Technol. Work **18**, 529–540 (2016)
2. Borghini, G., Isabella, R., Vecchiato, G., Toppi, J., Astolfi, L., Caltagirone, C., Babiloni, F.: Brainshield HREEG study of perceived pilot mental workload. Ital. J. Aerosp. Med. **5**, 34–47 (2011)
3. Statistical Summary of Commercial Jet Aircraft Accidents. Boeing, Seattle (2003)
4. Hart, S.G., Staveland, L.E.: Development of NASA task load index (TLX): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 239–250. North Holland Press, Amsterdam (1988)

5. O'Brien, T.G., Charlton, S.G.: Handbook of Human Factors Testing and Evaluation. Hillsdale, England (1996)
6. Cain, B.: A review of the mental workload literature. English **35** (2007)
7. Ahlstroma, U., Friedman-Bergb, F.J.: Using eye movement activity as a correlate of cognitive workload. Int. J. Ind. Ergon. **36**, 623–636 (2006)
8. Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J.-L., Besson, P.: Cognitive workload and psychophysiological parameters during multitask activity in helicopter pilots. Aerosp. Med. Hum. Perform. **86**(12), 1–6 (2015)
9. Mccann, R.S., Andre, A.D., Begault, D., et al.: Enhancing taxi performance under low visibility: are moving maps enough? Proc. Hum. Factors Ergon. Soc. Ann. Meet. **41**, 37–41 (1997)
10. Foyle, D.C., Andre, A.D., Battiste, V.: Advanced navigation aids in the flight deck: effects on ground taxi performance under low visibility conditions. In: Proceedings of the 41st Annual Meeting of the Human Factors and Ergonomics Society (HFES) (1996)
11. Parasuraman, R., Jiang, Y.: Individual differences in cognition, affect, and performance: behavioral, neuroimaging, and molecular genetic approaches. NeuroImage **59**(1), 70–82 (2012)
12. Ottati, W.L., Hickox, J.C., Richter, J.: Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation. In: Proceedings of the Human Factors and Ergonomics Society, 43rd Annual Meeting, pp. 66–70. Human Factors and Ergonomics Society, Santa Monica (1999)
13. Doppelmayr, M., Finkenzeller, T., Sauseng, P.: Frontal midline theta in the pre-shot phase of rifle shooting: differences between experts and novices. Neuro Psychologia **46**(5), 1463–1467 (2008)
14. Benbassat, D., Abramson, C.I., Williams, K.W.: Comparative approach to pilot error and effective landing flare instructions. Int. J. Comp. Psychol. **15**(2–3), 249–255 (2016)
15. Benlic, U., Brownlee, A.E.I., Burke, E.K.: Heuristic search for the coupled runway sequencing and taxiway routing problem. Transp. Res. Part C Emerg. Technol. **71**, 333–355 (2016)
16. Haslbeck, A., Bengler, K.: Pilots' gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task. Cogn. Technol. Work **18**(3), 529–540 (2016)
17. Itan, L., Yitzhaky, Y.: Performance of visual search tasks from various types of contour information. J. Opt. Soc. Am. Opt. Image Sci. Vis. **30**(3), 392–402 (2013)
18. Johnson, M.E., Zhao, X., Faulkner, B.: Statistical models of runway incursions based on runway intersections and taxiways. J. Aviat. Technol. Eng. **5**(2), 15–26 (2016)

Psychological and Emotional Issues in Interaction

MINIMA Project: Detecting and Mitigating the Negative Impact of Automation

Bruno Berberian[1(✉)], Oliver Ohneiser[2], Francesca De Crescenzio[3], Fabio Babiloni[4], Gianluca Di Flumeri[4], and Andreas Hasselberg[2]

[1] French Aerospace Lab, ONERA, Salon Air, France
bruno.berberian@onera.fr
[2] German Aerospace Center (DLR), Braunschweig, Germany
[3] University of Bologna, Forlì, Italy
[4] BrainSigns srl, Rome, Italy

Abstract. In this paper, we present the preliminary steps conducted in the framework of the research project Mitigating Negative Impacts of Monitoring high levels of Automation (MINIMA). The main objectives of MINIMA are (i) to develop vigilance and attention neuro-physiological indexes, and (ii) to implement them in a system that can adapt its behavior and guide the operator's attention. The goal is to mitigate negative impacts of the foreseen increasing automation in future Air Traffic Control (ATC) scenarios on Air Traffic Controller (ATCo) performance. The first step of research activities consists of better comprehension of Out-Of-The-Loop (OOTL) phenomena and of current methods to measure and compensate such effects. Based on this State of the Art, we propose the MINIMA concept, i.e. a dynamic adaptation of the task environment which is foreseen as a major requirement to keep the human 'in the loop', perfectly aware of the traffic situation. In the second part of this paper, we give details about the highly automated Terminal Manoeuvring Area selected as case study. Additionally, we describe the adaptation mechanisms that are planned to be implemented into this task environment and analyzed in the MINIMA project. Finally, the document provides information about the technical implementation of the vigilance and attention measurement that will be used to trigger adaptation of the task environment.

Keywords: Air traffic controller · Terminal Manoeuvring Area · Automation · Vigilance · Attention · Adaptive task and support activation · Attention guidance · Electroencephalography

1 Introduction

Over the past few years the global air traffic growth has exhibited a fairly stable positive trend, even through economic immobility, financial crisis and increased security concerns. According to a prevailing opinion, this trend is unlikely to change in the future, although a number of contextual factors, such as political climate, economy, environmental issues, safety issues and security issues may affect its actual rate. Further, according to the 'Free Flight' and the '4D Trajectory Management' concepts, different types of aircraft, such as manned, unmanned, and autonomous aircraft, as well

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 87–105, 2017.
DOI: 10.1007/978-3-319-58472-0_8

as all kinds of rotorcrafts, will operate simultaneously in a ‘structure-less’ and ‘time based’ environment [1, 2] allowing for much more direct and continuous trajectories to be used. Also, brand new airspace designs, possibly dynamic, may be required.

Within this picture, traffic flow patterns will become more complex, making conflicts and situations harder to identify for a human operator, putting immense pressure on the air traffic control system. Several solutions have been proposed for modernizing air traffic control and meet the demands for enhanced capacity, efficiency, and safety. As envisaged by both SESAR JU and HALA! Research Network, higher levels of automation will help ATCos to deal with increasingly complex airspace scenarios, enabling them to manage complex situations in a safe and efficient way.

In the present case, the changes in the ATC environment will also cause a shift of *Air Traffic Controllers*’ (ATCo) tasks from active managing of aircraft to monitoring [3]. In the future, ATCos’ actions will only be necessary if an aircraft deviates from its scheduled plan. However, ATCos being less actively involved in the ATC task may be affected by the *Out-Of-The-Loop* phenomena including performance degradation during their work. Such a *new ATCo* may show a “diminished ability both to detect system errors, and subsequently to perform manual tasks in facing automation failures, compared with operators who normally perform the same tasks manually” [4].

MINIMA will address these performance issues. Its aim is to identify thresholds in future ATM scenarios identifying out-of-the-loop behavior and to find solutions to minimize the negative impact of monitoring high levels of automation on the human operator’s performance. In the following parts, we will first describe the performance consequences of the OOTL phenomenon and discussed how to explain such degradation. Then, we will present solutions described in the literature to mitigate OOTL phenomenon. Finally, we will propose a technical solution to detect and compensate this OOTL phenomenon.

2 OOTL Characterization

Increasing the automation of ATM will result in new roles for ATCo. They will mainly monitor highly automated system and intervene seldom. Such change (from manual to supervisory control) is far from trivial. The key difference between passive information processing and direct action on the process is that the former involves functions similar to those maintained during process monitoring (e.g., scanning information sources); whereas, the latter involves manual control functions including process planning, decision making, selecting responses and implementing strategies. The problems due to automation are related to these new roles that are created for operators when their tasks are changed from manual to supervisory control.

Indeed, empirical data suggest that human operator is not always very efficient in supervisory task [4, 5]. Particularly, several incidents highlight that when the automatic equipment fails, supervisors seem dramatically helpless for diagnosing the situation and determining the appropriate solution because they are not aware of the system state prior to the failure. Numerous experimental results confirm such difficulties. For example, Endsley and Kiris [4] provided evidence that performance during failure mode following a fully automated period were significantly degraded, as compared to a failure mode following a fully manual control. Merat and Jamson [6] reported similar conclusions. In a driving simulation task, they demonstrated that drivers' responses to critical events were slower in the automatic driving condition than in the manual condition.

This so called Out-Of-The-Loop (OOTL) performance problem represents a key challenge for both systems designers and human factor society. However, after decades of research, this phenomenon remains difficult to grasp and treat. In the following section, we aim to bring a better understanding of this crucial phenomenon.

2.1 Becoming Out of the Loop

In the current research, we consider this phenomenon as it occurs in the context of human machine interaction. The human involvement and role in control of automatic systems depends on the level of automation. In manual control, the operator acts on the object, hence performing control on a lower level of aggregation. When a high level of automation is used, all of the control levels mentioned above are active and the operator acts as a supervisor on a high level where system and sub-system functionality is monitored rather than individual objects. The automation took care of the lower level actions and the human operators simply watched over the system, presumably ever-alert for deviations and problems. In other words, operators are relegated to passive information processor: they are "out of the loop" [7–9].

In other words, the OOTL phenomenon corresponds to a **lack of control loop involvement** of the human operator. Automation technology has created an increasing distance between human operator and loop of control, making him disconnected from the automation system. Such a removal leads to a decreased ability of the human operator to intervene in system control loops and assume manual control when needed in overseeing automated systems (see following sections). Because automation is not powerful enough to handle all abnormalities, this difficulty in takeover is a central problem in automation design.

2.2 OOTL Performance Problem and Decrease in Situation Awareness

The origins of these takeover difficulties have been largely debated. In the current accepted picture, the degradation of the Situation Awareness (SA) appears as a key component to understand the OOTL performance problem. Situation awareness is defined as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [10]. Situation awareness encompasses three processes: the perception of what is happening (Level 1), the understanding of what has been perceived (Level 2) and the use of what is understood to think ahead (Level 3).

Nowadays, it is clear that a loss of situation awareness underlies a great deal of the out-of-the-loop performance problem and that OOTL phenomenon is characterized by both a failure to detect and to understand the problem and by difficulties to find appropriate solutions.

Failure to detect (Level 1)

Numerous incidents have been attributed to a lack of operator awareness of automation failures and a decrease in detection of critical system state changes when involving in automation supervision (for a review see [4]). The near crash of Air China's Boeing 747 in 1989 illustrates this difficulty. In this case, the aircraft experienced a gradual engine failure that the human pilot was not aware of because of autopilot compensation (through rudder control) up until the point of failure of the autopilot, itself. Subsequently, the jet stalled and plummeted thousands of feet, being recovered within a few seconds of the ocean surface [11]. Numerous empirical evidences confirm that ATCos may be poor in detecting aircraft-to-aircraft conflicts when they are not actively controlling the airspace but nevertheless have to monitor for occasional anomalies. As illustration, Endsley and Rodgers [12] found that ATCos showed poor performance in detecting conflicts in recorded traffic when they were passively monitoring the traffic. Galster et al. [13] found that passive monitoring with airborne control of aircraft separation, which would be the case under mature Free Flight, led to a marked decrease in conflict detection performance by ATCos under high traffic load.

Failure to understand (Level 2 and Level 3)

In addition to delays in detecting that a problem has occurred necessitating intervention, operators may meet difficulties to develop sufficient understanding of the situation and to overcome the problem. For example, Wickens and Kessel [14] demonstrate longer system recovery times and poor response accuracies for operators who had been removed from control loops in advance of critical events requiring intervention. This delay may prohibit operators from carrying out the very tasks they are required to perform or diminish the effectiveness of actions taken.

Further, during failure modes, operators who have been removed from system control may not know what corrective actions need to be taken to stabilize the system and bring it into control. Several examples of incidents and accidents resulting from these system misunderstandings have been reported [15–17]. "Automation surprises" are a direct instantiation of these difficulties in automation understanding and take-over situations [18] and correspond to situation where the operator is surprised by the behavior of the automation. These "automation surprises" are particularly well documented [19–21] and have been listed as one of the major cause of incidents.

2.3 Origins of the Out-Of-The-Loop Performance Problem

Changes in attention/vigilance mechanisms appear as a first concern to explain the OOTL performance problem and the SA deterioration. And for good reasons since research on vigilance has shown that humans are poorly suited for monitoring role [22]. Pilots' reports of incidents have notably highlighted difficulties in monitoring automated systems. For example, Mosier and collaborators [23] examined NASA's Aviation Safety Reporting System database and found that 77% of the incidents in which over-reliance on automation was suspected involved a probable vigilance failure. Similarly, Gerbert and Kemmler [24] studied German aviators' anonymous responses to questionnaires about automation-related incidents and reported failures of vigilance as the largest contributor to human error. Two main sources could explain such decrease in vigilance: the inability to maintain a high level of vigilance in time and the complacency effect.

Difficulties to maintain high level of vigilance
One unintended consequence of automation for human operators is boredom. Indeed, highly automated environments require maintaining high levels of vigilance during a long period of time. In many phases of operation, operators are reduced to monitoring activities, waiting for the unlikely system anomaly. Resulting boredom increases the likelihood of operator distraction, which ultimately can affect system performance if operators miss or respond late to critical events.

Interestingly, several studies show that sustained attention over hours cannot be achieved [22, 25]. Research on vigilance suggests that time on task decreases significantly the discrimination of infrequent and unpredictable signals from a noisy background of nonsignals [26–28]. Moreover, there is some consensus for the existence of a decrease of human operator vigilance in case of interaction with highly automated system [29–31]. Both change in vigilance level and deterioration of the attentional mechanisms could cause degradation of the monitoring process involved in supervisory task and decrease performance in failure detection and system understanding.

Over-trust/Complacency
Together with this difficulty to maintain high level of vigilance in time, decrease in vigilance could also result from an overreliance on automation, the so-called complacency phenomenon [32]. Complacency defines the cognitive orientation toward high reliability automation, particularly prior to the first time it has failed in the user's experience [33]. Overreliance or complacency is created as operators form beliefs of the technical system as being more competent than it actually is. This overreliance on automation represents an important aspect of misuse that can result from several forms of human error, including decision biases and failure of monitoring [34–36]. A typical illustration is the case of the crash of Northwest Airlines at Detroit Airport in 1987. The McDonnell Douglas MD-80 crashed due to improper configuration of the flaps and slats of the aircraft. All persons were killed because an automated take-off configuration warning system, which the crew relied on, failed to function. They did not realize the aircraft was improperly configured for take-off and failed to check manually [37]. In this case like in others, when computer control facilities failed, operators, out of the direct control loop, were unaware of the state of the system and encounter difficulties to compensate for the failure mode before an accident occurred.

The previous section illustrates how the lack of operator involvement in automated systems control and the vigilance decrements induced will contribute to the loss of operator situation awareness. The following will present the different solutions currently proposed to mitigate this OOTL performance problem.

3 Current Solutions for OOTL Mitigation

Solutions for solving the OOTL-Problem are diverse. Following the perspective of an "Interaction Problem", a solution can either target at the system or target at human operators to make them less prone to OOTL-Problems.

3.1 Human Operator “Adaptation”

A first solution for OOTL problems targeting human operators is acting on their training. Human operator can be explicitly trained for situations in which OOTL problems can occur. For example, in a laboratory experiment using a process control simulation, Bahner and colleagues [38] showed that a preventive training in which participants were exposed to rare automation failures could significantly reduce complacency. Careful selection of the operators is presented as another solution targeting the human relates. Today, ATCos are carefully selected based on the key ability required by the working environments. Regarding the increase in automation, the ability to monitor automated systems and to switching immediately from monitoring to decision making will become an important competence in the selection of future ATCos [39]. If promising, such solution needs time to become effective and empirical evidence needs to be collected regarding its effectiveness.

3.2 System Adaptation

Since OOTL problems are caused by changing the system and introducing higher levels of automation, it seems likely that it can also be solved by changing the systems. For example, MABA-MABA-like methods (Men Are Better At-Machines Are Better At) rest on the idea that you should exploit the strengths of both humans and machines differently. The basic premise is: give the machines the tasks that they are good at, and the humans the things that they are good at (see for example [40]). However, Dekker and Woods [41] argued that such methods are misleading as automation often has unexpected effects [42]. These include the OOTL problems discussed above. It is now clear that introducing automation does not simply transfer the execution of functions to the machine, but instead create completely new functions and transform human practice. They conclude that automation needs to support cooperation with human operators – in standard and unexpected situations. Also, Rieth and collaborators [43] argued for better design of Human-Machine-Systems. They showed that the visual salience of standard indicators “generally do not draw attention to the information needed to identify emerging problems” and suggested other formats by which better mapping the task-relevance of information to the visual salience of how it is displayed. A holist approach is to develop automation in such a way that it can be seen as a partner. Human operator and automation should form a team that works cooperatively together, in a highly adaptive way to achieve its objectives. They have to adapt to each other and to the context in order to guarantee fluent and cooperative task achievement. For example, Klein and colleagues [44] defined ten challenges to improve human machine cooperation (model the others’ intentions, be delectable, make their status and intentions obvious and be able to interpret the status and intention of others, be able to engage in goal negotiations and enable a collaborative approach, be able to participate in managing attention, and help controlling the costs of coordinated activity).

3.3 Adaptive Automation as a Solution

A system that can be considered as cooperative must be able to adapt to the needs and the state of the user in real time. It is able to meet the changing needs of operators often without requiring the human operator to explicitly state his needs or trigger the adaptations. Making a system adaptive enables it to behave like a good human assistant. A technical solution for some of these challenges is the concept of adaptive systems.

The concept of Adaptive Automation (AA) concentrates on the dynamic allocation of function between operators and systems. This means, in that the Level of Automation of such system is not fixed but is adapted during the runtime according to the current needs of the operator. Particularly, the level of automation of such system is not fixed, but it is adapted during the activity according with the current needs of the operator [45]. Consequently, adaptive automation enables the level or modes of automation to be tied more closely to operator needs at any given moment [46] without requiring the human operator to explicitly state his/her needs or trigger the adaptations. Several evidences have proved the AA can improve operator's performance and moderate workload in complex environment [47, 48]. Besides the dynamic allocation of functions, other aspects of a system can be adapted during operations like, for example, the modality which is used to provide information, the amount of information that is presented to the operator or the lay-out of the information. The MINIMA project aims to design such cooperative/adaptive system. In the last section, we will describe the solution proposed in MINIMA.

4 MINIMA Concept

MINIMA will develop a dynamic adaptation of the task environment which is foreseen as a major requirement to keep the human 'in the loop', perfectly aware of the traffic situation. As a consequence of the developed concept, not all tasks potentially automated will be automated every time. To trigger adaptations of the automation, MINIMA will develop a real-time monitoring system that constantly measures the operators' state regarding the OOTL phenomenon. Because vigilance decrement is considered as one the major index of OOTL phenomenon, we will focus on vigilance and attention levels. This is called "Vigilance and Attention Observer" in MINMA. A component called "Adaptive Task and Support Activation" will decide based on the measured vigilance and attention level which adaptions of the task environment should be activated. On Overview of the MINIMA Concept is shown in Fig. 1.

4.1 Use Case Selected in MINIMA

A highly automated Terminal Manoeuvring Area (TMA) has been selected as use case in MINIMA. This task environment represents an air traffic control task as it is expected for the future: Most of the interaction with the aircraft is automated. A principle assumption of MINIMA is that Air Traffic Controllers (ATCos) are required to intervene in a few situations as error-free automation cannot be guaranteed.

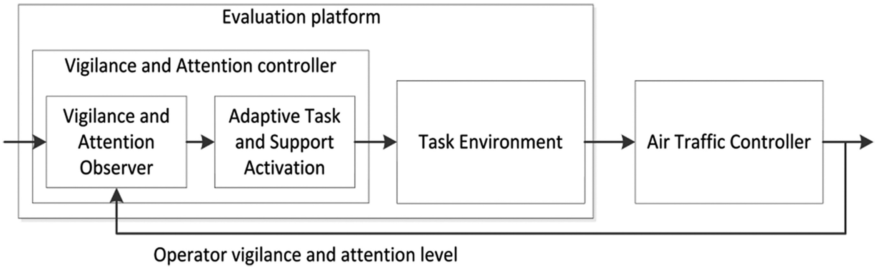


Fig. 1. Evaluation environment for MINIMA and influences of components on others

In the MINIMA use case, the arrival management will be highly automated. On-board Flight Management Systems (FMS) will negotiate with an Arrival Manager (AMAN) on 4D-Trajectories automatically. However, these trajectories are only guaranteed to be conflict-free at a merging point. Conflicts between arrival aircraft at other locations, conflicts between arrival and departures, and deviations form 4D-Trajectoreis are still possible - but seldom - and need to be managed by the ATCos.

One controller will be responsible for monitoring arriving and departing traffic. He/She only needs to intervene in cases of conflicts or emergencies.

4.2 Vigilance and Attention Observer

A critical challenge when designing such adaptive automation is to determine how changes among modes or levels of automation will be triggered. In other words, what should determine and “trigger” allocation of functions between the operator and the automation system. In MINIMA, we propose to use biophysical measure of vigilance and attention to trigger changes among the modes of automation, a module called *Vigilance and Attention observer.*

The importance of vigilance decrement for understanding human performance in a variety of industrial and military systems is now largely accepted. Several studies have shown that accidents ranging in scale from major to minor are often the result of vigilance failures [49]. Hawley [50], for example, described the role of vigilance and situation awareness in fratricide incidents in the Iraq war involving the highly automated Patriot missile system. Moreover, vigilance decrement is also considered as one the major index of OOTL phenomenon as described previously. In MINIMA project, we assume that both change in vigilance level and deterioration of the attentional mechanisms could cause degradation of the monitoring process involved in supervisory task. In this context, the aim of the operator vigilance and attention level observer is to measure both the current vigilance and attention levels of the human operator in view to quantify the OOTL phenomenon.

Biomarkers of vigilance/Attention

Defining a relevant index of alertness and sustained attention level in an operational context could be considered as the first step required for the development of such an inference system. In the method propose in MINIMA, physiological signals that reflect central nervous system activity would serve as a trigger for shifting among modes or levels of automation. Several reasons explain our choice. First, the measures can be obtained continuously with little intrusion [51, 52]. Second, these measures have been found to be diagnostic of multiple levels of arousal, attention, and workload. Even if there are still many critical conceptual and technical issues (e.g., making the recording equipment less obtrusive and obtaining reliable signals in noisy environments), numerous works have proved that it is indeed possible to obtain indices of one's brain activity and use that information to drive an adaptive automation system to improve performance and moderate workload in complex environment [47, 48].

Our literature review identified several biopsychometrics sensitive to changes in vigilance/sustained attention suggesting them as potential candidates for triggering adaptive automation such as electroencephalographic (EEG), near-infrared spectroscopy (NIRS), transcranial Doppler sonography (TCD), oculometrics, electrocardiogram (ECG) or skin electric potential (GSR). Regarding both advantages and disadvantages of the different techniques available, we decide to use EEG as biomarkers of change in vigilance. Compared to others neuroimaging devices, EEG offers the best compromise between spatial and temporal resolution, practical use and cost. Together with the use of EEG for vigilance monitoring, we propose to use eye tracking techniques to take into account how the available attentional resources is used by ATCo.

EEG and Vigilance

Electroencephalography (or EEG) is the recording of electrical activity produced by the firing of neurons within the brain. Several studies have demonstrated the suitability of EEG for real-world monitoring of mental states [53] and for brain-computer-interface (BCI) applications [54, 55].

Interestingly, EEG is assumed as one of the most reliable indicators of vigilance [56] and a number of EEG markers have been specifically correlated with vigilance. Particularly, change in power spectral densities (PSD) within the classically defined frequency bands (alpha, beta, theta, delta, and gamma) or ratios between these frequencies bands seem directly linked to change in vigilance state. Nowadays, there is a very large literature concerning the relationship of oscillatory activity and attention/vigilance [57–59] and brain dynamics associated to vigilance are well known. The general evidence is that lower levels of vigilance are related to increases in lower frequencies (theta and alpha) in EEG spectrum [60]. Several BCI systems have been designed based on this idea [54, 55, 61, 62]. Interestingly, these studies show that mental (de)activation may be monitored by changing balance between brain activity regions. Beta activity (12–30 Hz) is predominant when the participant in the study is generally awake and alert, while the activity dropping to Alpha activity (8–12 Hz) indicates developing drowsiness, and going further down into the theta region (5–8 Hz) may even lead to falling asleep.

In conclusion, summarizing the evidences in order to set a proper experimental design, vigilance could be evaluated investigating theta, alpha, and beta activity on the frontal, frontotemporal and parietal sites of the brain, thus the EEG system will be set up accordingly.

Technical implementation of vigilance measurement
The vigilance monitor device proposed will be based on the power spectral density (PSD) distribution. It will encompass four functions (see Fig. 2):

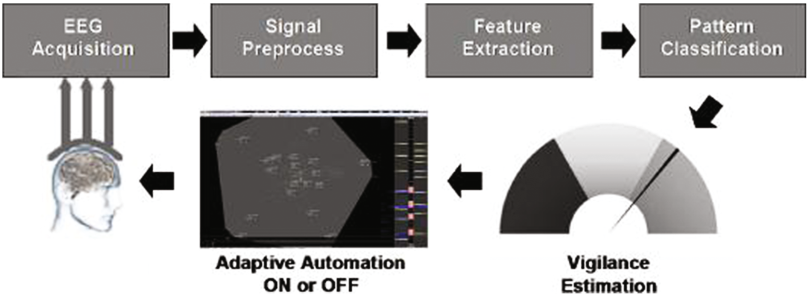


Fig. 2. General Framework of the proposed BCI system for vigilance analysis and estimate. Multi-channels raw EEG signals are first collected as inputs, and then EEG signal must be denoised and removed artefacts. After the preprocess of EEG data, vigilance level would be analyzed by feature extraction and pattern classification algorithms, the results would trigger the Adaptive Automation when the vigilance decrement condition occurs.

- Function 1: EEG acquisition and amplifying,
  The EEG device used will be the Galileo BEPlus (EB Neuro Spa, Italy), a system able to record up to 64 EEG channels with a sampling frequency up to 1024 Hz. Galileo BEPlus is a wired EEG system with wet electrodes. The sample frequency will be 256 Hz. EEG caps of different size will allow recording correctly the signal from different subjects independently of their head size. We will first use a high number of electrodes (between 30 and 40) in order to investigate deeply the brain sites involved in the vigilance process. Then, once relevant signal identified and algorithms trained, we will propose to decrease the number of electrodes to 16 electrodes in order to reduce the intrusiveness and improve the wearability of the system proposed. The electrodes will be placed based on the Standard International 10–20 system, mainly on the prefrontal, frontal, frontotemporal, and parietal sites. Electrodes in mastoids will be used as reference.
- Function 2: Noise and Artefact elimination
  Removing the polluted signal segments or contaminated channels and reducing the noise influences from EOG, EMG and other channels are all necessary in the EEG signal preprocessing work. In the process of data preprocessing, we will first remove the EEG signals from the damaged channels. Then, the EEG signals will be band-pass filtered (5th order Butterworth filter) to eliminate the noise. Because the brain potential is generally between 1 Hz and 40 Hz, so the band-pass filter cut-off frequencies are set to 1 Hz and 40 Hz. At this point, artefacts elimination will be performed. The EEG signal will be segmented into epochs of 2(s), shifted of 0.125(s). The Fpz channel will be used to remove eyes-blink artefacts from the EEG data by using the regression-based algorithm REBLINCA [63]. For other sources of artefacts (i.e., ATCos normally communicate verbally and perform several

movements during their operational activity), specific procedures (Threshold criterion, Trend estimation, Sample-to sample difference) available in the EEGLAB toolbox [63] will be applied.

- Function 3: Feature extraction
  To make sense of the recorded EEG signal, feature extraction and data dimension reduction are needed. Due to the close relationship between the EEG spectrum and the subject's vigilance state, the rhythm activities, that is, EEG power in the three specified bands, Theta (4–8 Hz), Alpha (8–13 Hz), and Beta (13–30 Hz) and their ratios, are calculated as features. In MINIMA, the EEG Power Spectral Density (PSD) will be estimated by using the Fast Fourier Transform (FFT) in the EEG frequency bands. Such frequency bands will not be defined equally for all the subjects (e.g., alpha equal to the 8÷12 [Hz] band), but the Klimesch approach will be adopted by using the Individual Alpha Frequency, in order to take into account the physiological subjective aspects of brain activity [64].
- Function 4: Pattern classification algorithms
  Various classifiers performing classification task have been proposed. Amongst other, support vector machine (SVM), artificial neural networks (ANN) and autoregressive (AR) are the most commonly used classifiers in the EEG research domain. In the case of MINIMA, a machine learning approach will be thus adopted in order to compute a Vigilance index based on the selected brain features. Particularly, during the calibration scenarios (LowVigilance and HighVigilance), we calibrate the algorithm before the Testing scenarios presentation. In particular, the Power Spectral Density of EEG epochs related to each calibration scenario (LowVigilance and HighVigilance) will be calculated by using only the frequency bands directly correlated to the vigilance state. The EEG frequency bands [frequency resolution of 0.5(Hz)] of interest will be defined for each ATCo by the estimation of the Individual Alpha Frequency value, as stated previously. At this point, the classification algorithm automatic stop Stepwise Linear Discriminant Analysis (asSWLDA, [65]) will be used to identify the most relevant discriminant features among the two different experimental conditions related to the lowest and the highest level of vigilance. Once identified, the asSWLDA classifier will assigns to each significant feature specific weights plus a bias. These parameters will be used later on to compute online the vigilance level index of the user during the testing scenarios.

Oculometric measure and attention.

Often at work, it is asked to the operator to attend more than one thing at the same time. This is possible thanks to the human ability to move the attentional focus, and is essential for every multitasking activity [66]. Of course, measuring only the attentive intensity aspect, i.e. the vigilance in our case, does not pledge that this amount of vigilance is properly oriented toward the tasks composing the work activity: On the contrary, other distracting activities could cause great levels of vigilance but consequently low performance on the work activity itself. For such a reason, it is crucial to monitor also which is the attentional focus of the operator, i.e. where his/her visual attention is addressed. In this context, we propose to use eye tracking device.

Tracking eye movements has the potential to provide a more direct measure of where attention is deployed since the direction of gaze is generally considered to be tightly coupled to the orienting of attention — at least, under normal circumstances. It can provide crucial information about the "attentional path" of the subject, revealing for example if the gaze is correctly directed and if there are particular fixation points [67]. Uncoupling of gaze direction and attention can, of course, occur as Posner's task clearly demonstrates. The value of eye tracking is that in natural scene viewing — where the visual environment is complex compared to many simple experimental situations — it should provide a good guide to the locus of attention. In recent years, researchers have capitalized on this possibility, seeking eventually to understand how attention and gaze are deployed to make sense of the visual world.

Technical Implementation of Visual Attention Measurement

In MINIMA, we choose to use a remoted system (Tobii EyeX) with 60 Hz sampling rate. This system can deliver information in real time (i.e., gaze direction). In contrast to EEG signal, the interpretation of eye tracking data is something relatively easy. Even if many different methods of exploring eye data exist, the eye tracking data do not require any particular skill in terms of analysis, since the software itself provides results in terms of gaze movements, fixations and so on. For MINIMA, we propose to use eye tracking to identify in real time where, when and what people look at and what they fail to see.

Basically, we will use fixation and saccades as measures of visual attention and interest. Based on fixation position (where?) and timing information (when?), we could compute different index, like the time and the number of fixation spent on a specific area of interest (AOI) or the Time to First Fixation (or TTFF), that is the amount of time it takes a respondent to look at a specific AOI from stimulus onset. These different metrics will inform us where and when the ATCos look at and how different events in the simulation will catch attention of the ATCo.

Using these measures of the controller attention area, we aim to help the controller in keeping his attention at the relevant display areas. In case of critical event and whatever the current level of vigilance, three different situations could be envisaged: (1) ATCo is focused on the relevant part of the radar display regarding the current situation, (2) ATCo is focused on a non-relevant area of the radar display, (3) ATCo is exploring the environment without specific area of interest. Oculometric measure could help us to detect problematic situation as cases 2 or 3 and help ATCo to focus on the relevant information. To make it possible, we need (1) to evaluate which ATC event on the radar display is the most relevant for the controllers, (2) to detect which radar areas the controller actually focuses his attention on, e.g. via eye-tracking, and (3) to guide the controllers' attention to the relevant radar display area if his attention is somewhere else. How to guide vigilance and attention is precisely the purpose of the following section.

4.3 Attention Vigilance Guidance Module

In MINIMA, we propose to design a module, called the *adaptive task and support activation*, able to modify both the level of automation and the feedback sent by the automation technology in order to maintain the ATCos in the loop of control and improve their performance in monitoring task.

In order to identify relevant vigilance and attention guidance tools, a workshop concerning the future ATC was first conducted with four ATCos (2 female, 2 male, average age of 39.3 years, average work experience as an ATCo of 20.5 years). The following sections briefly depict some of the ideas for vigilance and attention guidance within MINIMA, derived from the workshop results.

- *Attention guidance to separation conflicts (Short term conflict prediction):*
  Whenever a situation arises that is classified as potential loss of separation or might become a future loss of separation if the ATCo does not interact, the system highlights the affected aircraft. In order to do so, the system calculates the distances between all aircraft based on the current position and the predicted trajectories and detects all separation conflicts between these trajectories. A separation conflict is detected and highlighted if the distance between two aircraft is predicted to be below 3 NM horizontally and below 1000ft vertically.
- *Attention guidance to aircraft that cannot meet agreed target times:*
  This adaptation is also based on prediction. It is checked if the position of aircraft on their route is according to the last agreed trajectory. If there is a difference, there is the risk that this aircraft might be too late or too early and that it will cause a conflict with another aircraft. If deviations above a threshold of 0.5 NM are detected, the aircraft is highlighted. For example, the difference is shown in the aircraft label.
- *Attention guidance with eye tracker:*
  The system monitors the eye movement of the operator and determines, based on a normative model, which areas are monitored insufficiently or not at all (monitoring loss). Additional information is provided to the ATCo to update his/her situation awareness in this specific area.
- *Centerline Separation Range*:
  The Centerline Separation Range (CSR) [68] is a visual hypothetical aircraft final visualization (see Fig. 3) similar to HungaroControl's tool MergeStrip [69]. The calculated remaining flight distances of all approaching aircraft relative to other aircraft are displayed as distance-to-go (DTG) in nautical miles on one single arrival flow line for each runway (white numbers). When summing up all DTGs of previous aircraft in the sequence, the DTG of the current aircraft results (e.g. sum up 1.4 + 7.54 + 8.01 to get the DTG of DLH500). The angle brackets indicate the trend of increasing or decreasing relative separation between two sequenced aircraft since last radar update. Aircraft are colored depending on their weight category (e.g., light, medium, heavy). The geometry is to distinguish between real (triangle) and projected aircraft positions. If there are multiple parallel runway centerlines, these will be reflected in the number of CSRs. In MINIMA, all arriving aircraft will be projected on the CSR. It will not show aircraft only on the Centerline. All objects move from the right to the left of the display depending on their speed. The leftmost number shows the distance to threshold. The CSR is displayed below the radar situation on the same display.

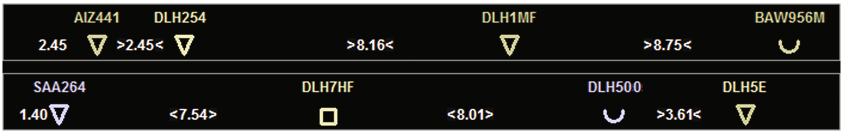


Fig. 3. Centerline separation range

- *Sequence optimization:*
  The AMAN calculates an optimized sequence for the arriving aircraft. The trajectories are generated in such a way that the aircraft arrive at the merging point according to the planned sequence. The AMAN adds aircraft entering the planning horizon of the AMAN to the list. However, except for adding new aircraft, the list will not be updated automatically. If ATCos would like to update the list, they have to modify the sequence manually (e.g., using drag and drop on a sequence display). Having the possibility to modify the sequence would allow offering a new kind of service to airlines. Airlines have additional restrictions that are not met today. Considering that increasing automation reduces the workload of controllers and that safety is kept on the same level, the services provided by ATCos could be increased. It could be possible for airlines to change sequences of their arriving aircraft if it only affects their own aircraft. This can only happen if the aircraft of one airline are in sequence and are controlled by the same ATCo. This additional task is triggered by airlines that already know the sequence of their arriving aircraft and would like to change. This could be simulated with an additional tool that provides the special demands issued by airlines and pilots.
- *Advisories*:
  The trajectories generated by the Arrival Manager will have phases of climb/descent in altitude, increase/decrease in speed. In the MINIMA application example, the aircraft will enter new phases included in the agreed trajectory automatically. If the aircraft were guided conventionally, the ATCo would have to give a clearance if an aircraft needs to enter a new phase of climb/descent or change its speed. It is possible to extract these necessary controller commands from the trajectory. This functionality was developed to support the controller by showing these commands as “Advisories”. The advisory for a controller can include aircraft call sign, command type, command value, and a countdown, when this command should be executed by the pilot. An example could be “AFR376 DESCEND Alt 4000 15 s” (see also Fig. 4). These advisories can either be displayed in a stack or directly at the radar label. As the aircraft follow their trajectories automatically and ATCos do not have to give clearance in the MINIMA use case, they could be unaware of the changes of aircraft behaviour. For example, an aircraft can start to descend without the ATCo noticing. In MINIMA, the “Advisories” could help to increase the ATCos awareness regarding the behaviour of the aircraft and changes in the flight profile, e.g., by showing these Advisories or by requesting the ATCo to confirm these Advisories. In case Advisories are not confirmed, the related changes of the flight profile are not executed by the aircraft. These aircraft are no longer guided automatically, but have to be guided manually by the ATCo.

Fig. 4. Advisory countdown (yellow) shown at the controller radar display (Color figure online)

- *Provision of additional information to increase service.*
  This idea is also considering that the safety is kept on the same level trough automation and that the ATCo has more time to provide better service. The ATCo should provide the aircraft with additional information that are not provided today and that could be of relevance for the pilots. The task is triggered by the pilots that request the information. The ATCo then can decide if his workload allows him to provide the requested information. Examples for requests are: predicted minimum separation, additional weather information, or the parking position.
- *Adaptation of Sector Size*:
  A common method to balance the workload of ATCos today is splitting or merging their airspace sectors. If high amounts of traffic are expected, the sector is split so that the traffic can be handled by two controller positions. The same could be applied in MINIMA. In principle, two types of adaptation of the sector size are possible. At first, the ATCo in the MINIMA use case can either be responsible for one or for both runways of the simulated airport. Further, it is possible to hand over aircraft from adjacent sectors earlier. The ATCo would be responsible longer for each individual aircraft and consequently would have to handle more aircraft at a time. In the next steps of this project, we will implement and test the solution proposed.

5 Conclusion

An increase of automation in air traffic control can have negative effects on the air traffic controller's performance. The effects are known as out-of-the-loop phenomenon. The MINIMA Project will develop a vigilance and attention controller to mitigate these effects. A highly automated arrival management task will be used as a case study. Psychophysiological measurements like EEG will be used to identify the state of the Air Traffic Controller and combined with adaptive task activation. This will allow for activating tasks based on the Air Traffic Controllers state to keep their performance on a high level and to ensure safe operations.

Acknowledgement. This project has received funding from the SESAR Joint Undertaking under grant agreement No. 699282 under European Union's Horizon 2020 research and innovation program. This paper reflects only the authors' view and the JU is not responsible for any use that may be made of the information it contains.

References

1. Birkmeier, B., Korn, B., Kügler, D.: Sectorless ATM and advanced SESAR concepts: complement not contradiction. In: 2010 IEEE/AIAA 29th Digital Avionics Systems (DASC). IEEE (2010)
2. Neal, A., Flach, J., Mooij, M., Lehmann, S., Stankovic, S., Hasenbosch, S.: Envisaging the future air traffic management system. Int. J. Aviat. Psychol. **21**(1), 16–34 (2011)
3. SESAR: The roadmap for delivering high performing aviation for Europe – European ATM Master Plan, Brussels (2015)
4. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. Hum. Fact.: J. Hum. Fact. Ergon. Soc. **37**(2), 381–394 (1995)
5. Kaber, D.B., Endsley, M.R.: Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. Process. Saf. Prog. **16**(3), 126–131 (1997)
6. Merat, N., Jamson, A.H.: How do drivers behave in a highly automated car. In Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 514–521 (2009)
7. Norman, S.D., Orlady, H.W.: Flight deck automation: promises and realities (1989)
8. Rasmussen, J., Rouse, W.B.: Human detection and diagnosis of system failures (1981)
9. Wiener, E.L., Curry, R.E.: Flight-deck automation: promises and problems. Ergonomics **23** (10), 995–1011 (1980)
10. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 32, no. 2, pp. 97–101 (1988)
11. Wickens, C.D.: Engineering Psychology and Human Performance, 2nd edn. Harper Collins, New York (1992)
12. Endsley, M.R., Rodgers, M.D.: Distribution of attention, situation awareness and workload in a passive air traffic control task: implications for operational errors and automation. Air Traffic Control Q. **6**(1), 21–44 (1998)
13. Galster, S.M., Duley, J.A., Masalonis, A.J., Parasuraman, R.: Air traffic controller performance and workload under mature free flight: conflict detection and resolution of aircraft self-separation. Int. J. Aviat. Psychol. **11**(1), 71–93 (2001)
14. Wickens, C.D., Kessel, C.: The effects of participatory mode and task workload on the detection of dynamic system failures. IEEE Trans. Syst. Man Cybern. **9**(1), 24–34 (1979)
15. Billings, C.E.: Aviation automation: the search for a human-centered approach (1997)
16. Funk, K., Lyall, B., Wilson, J., Vint, R., Miemcyzyk, M., Suroteguh, C.: Flight deck automation issues. Int. J. Aviat. Psychol. **9**, 125–138 (1999)
17. Sarter, N.B., Woods, D.D.: How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum. Fact.: J. Hum. Fact. Ergon. Soc. **37**(1), 5–19 (1995)
18. Sarter, N.B., Woods, D.D., Billings, C.E.: Automation surprises. Handb. Hum. Fact. Ergon. **2**, 1926–1943 (1997)

19. Degani, A., Heymann, M.: Some formal aspects of human automation interaction. NASA Technical Memorandum Number 209600. NASA Ames Research Center, Moffett Field, CA (2000)
20. Palmer, E.: Murphi busts an altitude: a murphi analysis of an automation surprise. In: Proceedings of 18th Digital Avionics Systems Conference, vol. 1, p. 4-B. IEEE (1999)
21. Van Charante, E.M., Cook, R.I., Woods, D.D., Yue, L., Howie, M.B.: Human-computer interaction in context: physician interaction with automated intravenous controllers in the heart room. In: Analysis, Design, and Evaluation of Man-Machine Systems, pp. 263–274 (1992)
22. Davies, D.R., Parasuraman, R.: The Psychology of Vigilance. Academic Press, New York (1982)
23. Mosier, K.L., Skitka, L.J., Korte, K.J.: Cognitive and social psychological issues in flight crew/automation interaction. In: Human Performance in Automated Systems: Current Research and Trends, pp. 191–197 (1994)
24. Gerbert, K., Kemmler, R.: The causes of causes: determinants and background variables of human factor incidents and accidents. Ergonomics **29**(11), 1439–1453 (1986)
25. Matthews, G., Davies, D.R.: Individual differences in energetic arousal and sustained attention: a dual-task study. Pers. Individ. Differ. **31**(4), 575–589 (2001)
26. Mackworth, N.H.: The breakdown of vigilance during prolonged visual search. Q. J. Exp. Psychol. **1**(1), 6–21 (1948)
27. Parasuraman, R.: Memory load and event rate control sensitivity decrements in sustained attention. Science **205**(4409), 924–927 (1979)
28. Warm, J.S.: An introduction to vigilance. In: Sustained Attention in Human Performance, pp. 1–14 (1984)
29. O'Hanlon, J.F.: Boredom: practical consequences and a theory. Acta Physiol. (Oxf) **49**(1), 53–82 (1981)
30. Strauch, B.: Investigating Human Error: Incidents, Accidents, and Complex Systems. Ashgate, Burlington (2002)
31. Wiener, E.L.: Cockpit automation. In: Wiener, E.L., Nagel, D.C. (eds.) Human Factors in Aviation, pp. 433–461. Academic, San Diego (1988)
32. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation induced 'complacency'. Int. J. Aviat. Psychol. **3**(1), 1–23 (1993)
33. Rovira, E., McGarry, K., Parasuraman, R.: Effects of imperfect automation on decision making in a simulated command and control task. Hum. Fact.: J. Hum. Fact. Ergon. Soc. **49** (1), 76–87 (2007)
34. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Fact.: J. Hum. Fact. Ergon. Soc. **39**(2), 230–253 (1997)
35. Moray, N., Inagaki, T.: Attention and complacency. Theor. Iss. Ergon. Sci. **1**(4), 354–365 (2000)
36. Sheridan, T.B., Parasuraman, R.: Human–automation interaction. In: Nickerson, R.S. (ed.) Reviews of Human Factors and Ergonomics. Human Factors and Ergonomics Society, vol. 1, pp. 89–129. Santa Monica, CA (2006)
37. Endsley, M.R.: Automation and Situation Awareness. Automation and Human Performence: Theory and Application, pp. 163–181 (1996)
38. Bahner, J.E., Hüper, A., Manzey, D.: Misuse of automated decision aids: complacency, automation bias and the impact of training experience. Int. J. Hum. Comput. Stud. **66**(9), 688–699 (2008)
39. Bruder, C., Eißfeldt, H., Maschke, P., Hasse, C.: A model for future aviation. Aviat. Psychol. Appl. Hum. Fact. **4**(1), 13–22 (2014)

40. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man, Cybern. A **30**(3), 286–297 (2000)
41. Dekker, S.W., Woods, D.D.: MABA-MABA or abracadabra? Progress on human–automation co-ordination. Cogn. Technol. Work **4**(4), 240–244 (2002)
42. Bainbridge, L.: Ironies of automation. Automatica **19**(6), 775–779 (1983)
43. Rieth, C.A., Cook, M.B., Ngo, M.K.: Displays for effective human-agent teaming: evaluating attention management with computational models. In: Shumaker, R., Lackey, S. (eds.) International Conference on Virtual, Augmented and Mixed Reality, vol. 9179, pp. 216–227. Springer International Publishing, Cham (2015)
44. Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J.: Ten challenges for making automation a "team player" in joint human-agent activity. IEEE Intell. Syst. **19**(6), 91–95 (2004)
45. Scerbo, M.W.: Adaptive automation. In: Karwowski, W. (ed.) International Encyclopedia of Ergonomics and Human Factors, 2nd edn, pp. 1893–1896. CRC/Taylor & Francis, Boca Raton (2006)
46. Parasuraman, R., Bahri, T., Deaton, J.E., Morrison, J.G., Barnes, M.: Theory and design of adaptive automation in aviation systems. Naval Air Warfare Center, Warminster, PA, USA (Report, NAWCADWAR-92 033-60) (1992)
47. John, M.S., Kobus, D.A., Morrison, J.G., Schmorrow, D.: Overview of the DARPA augmented cognition technical integration experiment. Int. J. Hum.-Comput. Interact. **17**(2), 131–149 (2004)
48. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Hum. Fact.: J. Hum. Fact. Ergon. Soc. **45** (4), 635–644 (2003)
49. Molloy, R., Parasuraman, R.: Monitoring an automated system for a single failure: vigilance and task complexity effects. Hum. Fact.: J. Hum. Fact. Ergon. Soc. **38**(2), 311–322 (1996)
50. Hawley, J.K.: Patriot fratricides: the human dimension lessons of operation Iraqi freedom. FA J. **11**(1), 18 (2006)
51. Eggemeier, F.T.: Properties of workload assessment techniques. Adv. Psychol. **52**, 41–62 (1988)
52. Kramer, A.F.: Physiological metrics of mental workload: a review of recent progress. In: Damos, D.L. (ed.) Muliple-Task-Performance, pp. 329–360. Taylor & Francis, London (1991)
53. Makeig, S., Gramann, K., Jung, T.P., Sejnowski, T.J., Poizner, H.: Linking brain, mind and behavior. Int. J. Psychophysiol. **73**(2), 95–100 (2009)
54. Lin, C.T., Chuang, C.H., Huang, C.S., Tsai, S.F., Lu, S.W., Chen, Y.H., Ko, L.W.: Wireless and wearable EEG system for evaluating driver vigilance. IEEE Trans. Biomed. Circuits Syst. **8**(2), 165–176 (2014)
55. Zhang, Z., Luo, D., Rasim, Y., Li, Y., Meng, G., Xu, J., Wang, C.: A vehicle active safety model: vehicle speed control based on driver vigilance detection using wearable EEG and sparse representation. Sensors **16**(2), 242 (2016)
56. Oken, B.S., Salinsky, M.C., Elsas, S.M.: Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin. Neurophysiol. **117**(9), 1885–1901 (2006)
57. Foxe, J.J., Snyder, A.C.: The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Front. Psychol. **2**, 154 (2011)
58. Frey, J.N., Ruhnau, P., Weisz, N.: Not so different after all: the same oscillatory processes support different types of attention. Brain Res. **1626**, 183–197 (2015)
59. Jensen, O., Kaiser, J., Lachaux, J.P.: Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci. **30**, 317–324 (2007)

60. Steriade, M., Gloor, P., Llinas, R.R., Da Silva, F.L., Mesulam, M.M.: Basic mechanisms of cerebral rhythmic activities. Electroencephalogr. Clin. Neurophysiol. **76**(6), 481–508 (1990)
61. Makeig, S., Elliott, F., Postal, M.: First demonstration of an alertness monitoring, management system. Technical report, pp. 93–36 (1993)
62. Ji, H., Li, J., Cao, L., Wang, D.: A EEG-based brain computer interface system towards applicable vigilance monitoring. In: Wang, Y., Li, T. (eds.) Foundations of Intelligent Systems, vol. 122, pp. 743–749. Springer, Heidelberg (2011)
63. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods **134**, 9–21 (2004)
64. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. **29**(2), 169–195 (1999)
65. Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., Babiloni, F.: A passive brain–computer interface application for the mental workload assessment on professional ATCOs during realistic air traffic control tasks. Progr. Brain Res. **228**, 295–328 (2016)
66. Simons, D.J., Levin, D.T.: Change blindness. Trends Cogn. Sci. **1**(7), 261–267 (1997)
67. Stern, J.A., Boyer, D.J., Schroeder, D.J., Touchstone, R.M., Stoliarov, N.: Blinks, saccades; and fixation pauses during vigilance task performance: 2: gender and time of day (1996)
68. Schaper, M., Temme, M.M., Gluchshenko, O., Christoffels, L., Pick, A.: Coupling of ATM planning systems with different time horizons. In: Air Transport and Operations Symposium, Delft (2012)
69. HungaroControl: HungaroControl's MergeStrip Process. HungaroControl, Budapest (2013)

Cognitive Considerations in Auditory User Interfaces: Neuroergonomic Evaluation of Synthetic Speech Comprehension

Adrian Curtin[1,2] and Hasan Ayaz[1,2,3,4](✉)

[1] School of Biomedical Engineering, Science and Health Systems,
Drexel University, Philadelphia, PA 19104, USA
ayaz@drexel.edu
[2] Cognitive Neuroengineering and Quantitative Experimental Research
(CONQUER) Collaborative, Drexel University, Philadelphia, PA 19104, USA
[3] Department of Family and Community Health,
University of Pennsylvania, Philadelphia, PA 19104, USA
[4] Division of General Pediatrics, Children's Hospital of Philadelphia,
Philadelphia, PA 19104, USA

Abstract. Automated spoken language interfaces have seen a remarkable proliferation in recent years, integrating with automotive, household, industrial, and mobile platforms to shape the way in which we interact with our devices. While the use of an auxiliary auditory information stream has the potential to decrease interference and prevent disengagement from operation of traditional visual/mechanical interfaces, evidence from behavioral and neuroimaging studies have suggested that the brain mechanisms underlying the perception and comprehension of synthetic speech may be different from naturally produced speech, resulting in an unnecessary additional cognitive burden.

In this neuroergonomics study, functional Near-Infrared Spectroscopy (fNIRS) over the anterior prefrontal cortex has been measured to determine the influence of synthetic speech quality during a sentence comprehension and quality assessment task. Eight participants were asked to listen to topical sentences from real-world audio interfaces employed in car driving scenarios and then answer questions regarding the content of the messages and rate the quality (*Intelligibility* and *Naturalness*) of the audio. Results indicate that the behavioral performance during assessment of speech content and rated *Intelligibility* were negatively impacted when using lower quality synthetic voices. Performance costs associated with low-quality synthetic voices were related to increased cognitive load as measured by increased medial prefrontal cortex activity. Approaches and concepts described here can be used to guide next-gen speech synthesizer design and future research for decreasing the cognitive load in driving scenarios.

Keywords: Prefrontal cortex · fNIRS · Synthetic speech perception · Auditory

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 106–116, 2017.
DOI: 10.1007/978-3-319-58472-0_9

1 Introduction

Users of modern day electronics are witnessing an increasing ubiquity of connected devices with wide variety of interface designs. While some of these devices require voice-based interfaces due to a lack of mechanical input controls or any visual display, others might offer these as add-on features for specific operational requirements, or simple convenience. A decreasing cost of both electronic components and processing power paired with increased usability has invited auditory interfaces into a wide variety of commercially available products, in turn leading more users to expect these interfaces. Voice-based interfaces allow users to interact in a hands-free manner, allowing and perhaps encouraging the user to engage in multiple activities. Presentation of auxiliary information to the user along an auditory channel may let the user continue to perform additional tasks which require visual and motor attention without a conscious awareness of the additional cognitive demands. Although in non-critical situations this compartmentalization of tasks can be easily made, in situations of distraction, overload, or fatigue the operator or other's safety may be jeopardized [1, 2]. Therefore, the burdens placed on cognition as a result of interacting with synthetic speech-based interfaces must be assessed in order to minimize these costs.

A vital characteristic of human speech perception is a tolerance to natural variability of voices as generated by different speakers [3]. This tolerance allows individuals to flexibly engage in speech perception under varying external conditions which may degrade either the auditory quality or introduce potential distractions, negatively impacting speech perception. In spite of the relatively resilient nature of speech, the presence of synthetic speech requires operators to adapt to an introduction of unnatural acoustic, phonetic, and prosodic properties [4, 5]. Behavioral and neuroimaging studies have suggested that despite maintaining successful comprehension of sentence content, this adaptation can introduce a measurable performance penalty and a corresponding increase in cognitive load [6]. One explanation of this phenomena suggests that the mechanisms underlying the perception of synthetic speech may be different from naturally produced speech [7, 8].

Understanding the neural mechanisms that contribute to the acquisition, development, and use of cognitive skills is an important goal for cognitive neuroscience research and for applications of neuroscience to work and everyday activities. Neuroergonomics, defined by late Prof. Raja Parasuraman as studying "the brain at work" [9–12] is an emerging interdisciplinary research field at the intersection of cognitive neuroscience, systems engineering, human factors, and psychology. By utilizing portable and wearable brain imaging sensors, unique information such as mental workload and state can be captured. This mental state is independent of performance measures, and can, in turn, be used to guide product or complex machine interface design or adaptation during field use.

Functional Near-Infrared Spectroscopy (fNIRS [13]) is a non-invasive, safe, silent, and portable neuroimaging technology well suited for the study of speech perception and language. fNIRS measures cortical correlates of neural activity via relative changes in cortical oxygenated (HbO) and deoxygenated hemoglobin (HbR), taking advantage of transmissive and diffusive properties of tissue when using near-infrared light [14].

The technique allows research to be conducted practically under real-world settings and has become increasingly popular in auditory research [15, 16] and applied research [17], linking operational characteristics with the underlying cognitive functions.

In this neuroergonomics study, the influence of synthetic speech quality during a sentence comprehension and quality assessment task was assessed using self-reported, behavioral, and fNIRS measures. Participants listened to topical sentences from real-world audio interfaces employed in car driving scenarios, then answered questions regarding the content of the messages and rated the quality of the audio to assess perceived *Intelligibility* and *Naturalness*. Three levels of synthetic speech quality (low, medium, and high) were assessed in addition to naturally recorded speech to identify cognitive considerations in synthetic speech systems.

2 Methods

2.1 Participants

Eight right-handed participants (7 Male, 1 Female) between the ages 18 to 35 volunteered for this study. Participants reported no hearing impairment, neurological or psychiatric history. All participants were medication-free, with normal or corrected-to-normal vision. Participants gave written informed consent for the study, which was approved by the Institutional Review Board at Drexel University, and were paid for their participation.

2.2 Experiment Protocol

Synthetic speech recordings used in the study were originally developed at Intel Labs, while the experimental protocol itself was implemented in a custom protocol presentation package. After a 5 s baseline period, subjects were asked to listen to short 5–10 s sentences with topics adapted from real-world audio interfaces employed in car driving scenarios. Following audio presentation, subjects were asked a question regarding the content of the message as a measure of comprehension and then asked to evaluate the quality of the audio.

Participants listened to 5 different sentence categories under 4 levels of audio quality (natural + 3 levels of synthetic voice). Synthetic speech synthesizer quality reflected the different system requirements required for operation. Synthesizer S1 required 250 MB system memory while S2 required 50 MB and S3 only required 1 MB. Comprehension questions were varied such that no two questions were repeated. *Voice*, *Category*, and *Comprehension* were pseudo randomized to account for order effects. Quality was evaluated on a 1 to 5 scale for the metrics of *Intelligibility* (1: Understood none – 5: Understood all) and *Naturalness* (1: Nothing like a human – 5: Exactly like a Human). Each evaluation period lasted approximately 30 s and as is presented in Fig. 1. The entire Listening task lasted about 20 min. Audio was presented to the subject utilizing a professional headphone amplifier (Head Acoustics HPS IV), transducer, and high-fidelity headphones (Sennheiser HD 600).

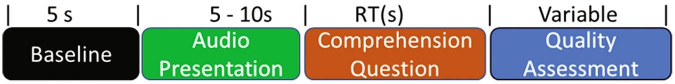


Fig. 1. Trial timeline block diagram

2.3 fNIRS Acquisition

Prior to the starting the task, subjects were fitted with a continuous wave fNIRS system (fNIR1100; fNIR Devices LLC; www.fnirdevices.com). The fNIRS system includes a flexible sensor pad with 4 dual-wavelength (730 nm, 850 nm) LED light sources and 10 light detectors arranged spatially with a 2.5 cm separation and time-multiplexed to allow for 16 measurement locations [14]. The sensor was placed directly above the eyebrows over the forehead to allow measurement of the cortical areas directly underlying as seen in Fig. 2. Due to experimental setup conditions, the sensor was placed such that the sensor was centered over the midline and then offset by 1.58 cm corresponding with an offset of one optode in the horizontal direction.

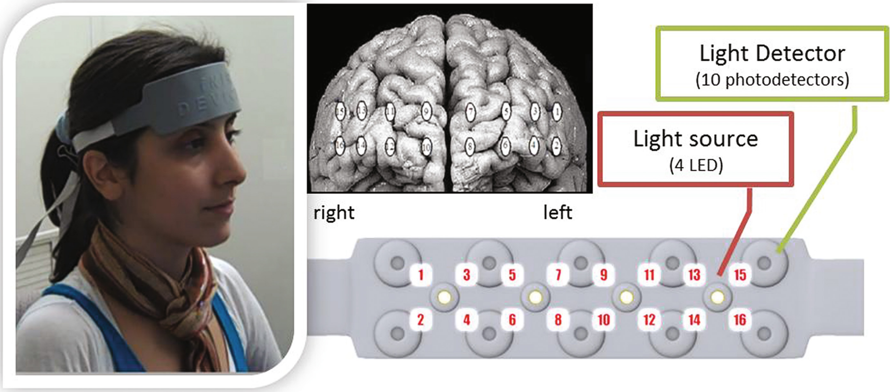


Fig. 2. Functional Near Infrared Spectroscopy sensor (headband) and optode locations visualized on anterior view brain surface image [18].

COBI Studio [19] software was used for data acquisition and visualization. fNIRS data was collected continuously at a 2 Hz sampling rate. A serial cable between the presentation computer and the fNIRS acquisition computer was used to relay event markers used to synchronize neuroimaging data with audio stimulation onsets and other events.

2.4 Data Analysis

fNIRS data was processed using Matlab (R2016b) while R (3.3.2) was used for statistical testing. Channels were assessed for basic data quality and channels contaminated by excessive light saturation, insufficient signal, or subject motion were rejected prior to fNIRS evaluation. Each participant's raw fNIRS data was low-pass filtered with a finite impulse response, linear phase filter with order of 20, and cut-off frequency of 0.25 Hz to attenuate noise from high frequency sources [18]. Motion artifacts were rejected automatically using a sliding motion artifact rejection technique [20]. Relative changes in blood oxygenation were calculated using the modified Beer Lambert Law (mBLL) [21] from changes in optical density measured during the pre-trial baseline period. Average change in oxygenation [Oxy], calculated as the difference in [HbO] and [HbR], from 4 to 7 s from the initiation of the audio presentation period were used as the dependent measure according to the estimated delay in the hemodynamic response function [22].

Main effects for dependent measures, including both self-reported, behavioral, and biomarker measurements were analyzed using repeated measures ANOVA. Subject and *Category* were used as fixed effects to control for the topic, audio length, and Subject variability. Tests of linear-hypotheses were corrected for multiplicity using the False Discovery Rate. A criterion of $\alpha = 0.05$ was designated as the threshold of statistical significance.

3 Results

3.1 Self-reported Measures

The subjective measures of *Intelligibility* and *Naturalness* were assessed using a repeated measures one-way ANOVA that adjusted for the content *Category* (Calendar, Email, Navigation, SMS, Weather) of the audio presented. There was a significant within subject main effect for Voice on *Naturalness* ($F_{3,117} = 156.5$, $p < 0.001$). Subjects appeared to easily order *Voices* in order of quality with Natural speech determining the highest level and Synthetic levels 1 to 3 (S1, S2, S3) showing decreasing levels of *naturalness*. Post-hoc tests determined that Natural speech was significantly different from all levels of Synthetic speech and that all levels of Synthetic speech showed significant differences from each other ($q_{(117,0.05/3)} = -2.84$ to -12.84). Average levels of *Naturalness* for each *Voice* are presented in Fig. 3a.

Despite most audio presentations in this study being highly Intelligible (> 3), *Voice* showed a significant main effect on *Intelligibility* ($F_{3,117} = 11.34$, $p < 0.001$). Post-hoc tests showed that Natural speech was the most highly-intelligible *Voice* with Synthetic speech decreasing in *Intelligibility* according to level. Natural Speech was significantly different from both S2 and S3 ($q_{(117,0.05/3)} = -4.69..-5.33$), and approached trend level differences with S1 ($q_{(117,0.05/3)} = -1.71$). S1 was also significantly different from both S2 and S3 levels ($q_{(117,0.05/3)} = -2.99..-3.63$). Levels S2 and S3 appeared to continue a decreasing trend in *Intelligibility*, but they were not statistically different after correction for multiple-hypotheses. Average rated *Intelligibility* for each *Voice* is presented in Fig. 3b.

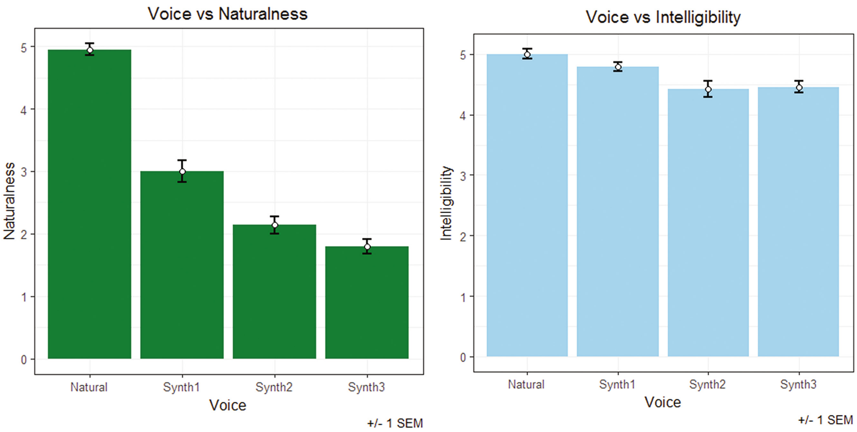


Fig. 3. Average self-reported ratings of *Naturalness* (left) and *Intelligibility* (right) for each *Voice*. Error bars are standard error of the mean (SEM).

3.2 Behavioral Results

Workload during sentence comprehension can often be assessed behaviorally by measuring the length of time required to process and respond to information of increased difficulty, typically associated with increased processing time. The response time (RT) during the *Comprehension Question* phase was used as a metric of cognitive demand required for each *Category* and *Voice* combination. A significant within-subjects effect was observed for *Voice* on Question RT ($F_{3,102} = 3.504$, $p = 0.018$) when adjusted for different categories. Post-hoc differences showed that Natural and S1 Voices were significantly different from S3 ($q_{(102,0.05/3)} = 3.10..3.16$) and approached trend differences with S2 ($q_{(102,0.05/3)} = 1.84$) after adjusting for Subject differences and *Category*. Average RT during the sentence comprehension period for each *Voice* is presented in Fig. 4.

3.3 fNIRS Measures

One-way repeated measures ANOVAs were performed separately for each Optode to assess the impact of *Voice*, *Intelligibility*, and *Naturalness* on fNIRS biomarkers. Oxygenation changes as measured by [Oxy] were analyzed using one-way repeated measures ANOVA. A main effect for *Voice* was observed in Optodes 3 [$F_{(3,61)} = 2.983$, $p = 0.038$] and 14 [$F_{(3,40)} = 2.841$, $p = 0.049$]. Post-hoc Tukey tests showed that in Optode 14, the response to Natural Voice was significantly different from S1 ($q_{(61,0.05/3)} = 2.72$) and S3 ($q_{(61,0.05/3)} = 2.73$), with a trend difference for S2 ($q_{(61,0.05/3)} = 2.235$). However, responses between Synthetic voices were undifferentiated.

When examining *Naturalness*, Optode 14 also showed a significant main effect [$F_{(4,39)} = 4.03$, $p = 0.008$]. Post-hoc tests showed that highly Natural ratings (5) were associated with decreased activity relative to low Natural ratings (2,3)

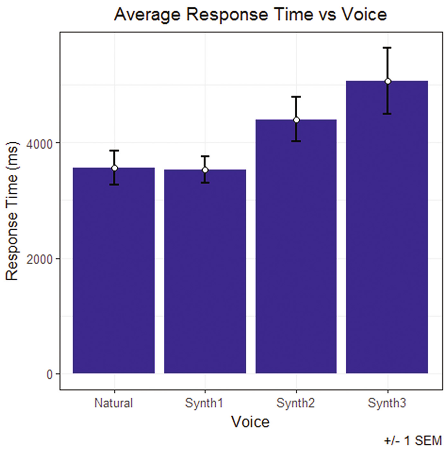


Fig. 4. Average response time for each speech quality level. Error bars are +/− standard error of the mean (SEM)

($q_{(39,0.05/4)} = (-4.11$ to $-2.64)$ and a trend difference for the lowest rating (1) ($q_{(39,0.05/4)} = -2.325$). Average changes in [Oxy] for Optode 14 for different *Voice* groups and across varying levels of self-reported *Naturalness* are presented in Fig. 5.

Optode 11 revealed a significant main effect for *Intelligibility* [$F_{(3,74)} = 3.85$, $p = 0.013$]. Tukey post-hoc tests indicated that highly Intelligible (4–5) presentations were significantly different than ratings of lower *Intelligibility* (2–3) ($q_{(76,0.05)} = -3.68$). Average changes in [Oxy] measured in Optode 11 across varied values of self-reported *Intelligibility* are presented in Fig. 6.

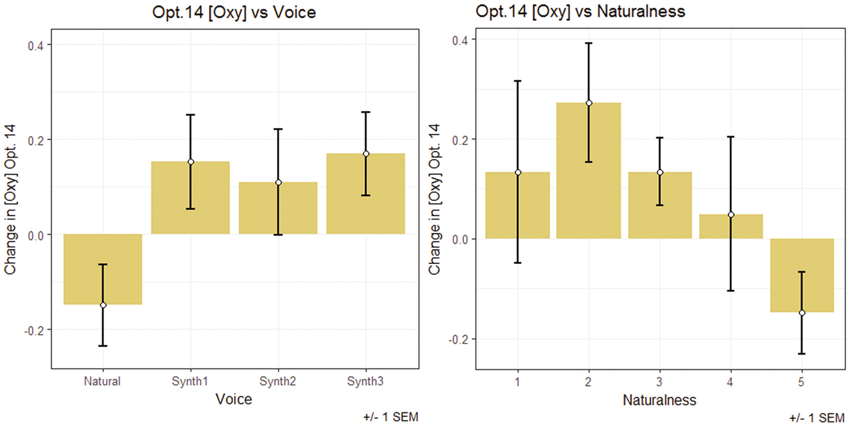


Fig. 5. Average oxygenation changes in Optode 14 across different *Voices* (left) and different rated levels of *Naturalness* (right). Error bars are standard error of the mean (SEM).

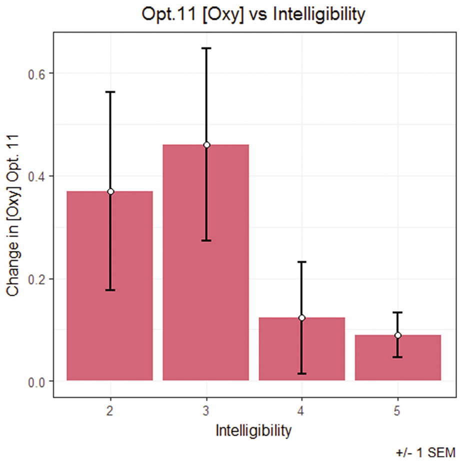


Fig. 6. Average oxygenation changes in Optode 11 across different rated levels of *Intelligibility*. Error bars are standard error of the mean (SEM).

4 Discussion

Extensive behavioral research has been conducted to assess the cognitive factors associated with the differential perception of synthetic and natural speech [4]. Researchers have largely demonstrated that natural and synthetic speech feature a variety of differences not only in the somewhat subjective quality of the voice itself, but in its accurate perception and comprehension. It is unsurprising that natural speech is typically viewed as superior to synthetic speech, as natural speech features a number of characteristics which even high quality synthetic speech cannot replicate. In human speech prosody is frequently varied, adapting the pace, emphasis, and even emotion in a manner that is listener, content, and context dependent, whereas synthetic speech features a more "mechanical" sounding quality, typical of rule-based generation, which is typically inflexible to changing listening conditions. The unnatural characteristics of synthetic speech systems may be further exacerbated by low-quality synthetic speech.

This study sought to identify the behavioral and cortical response to varying levels of synthetic audio quality during the comprehension of contextual audio frequently found in hands-free systems. Behavioral responses in terms of increased *Response Time* suggested that comprehension of lower-quality synthetic speech was more difficult than highly-natural and natural speech. Our fNIRS results appear to indicate that these trends can be anticipated by cortical evoked activity during active listening. Optodes 11 and 14, both located near the middle frontal gyrus, present significantly different evoked responses depending on presented audio quality. Increased cortical activity at Optode 14 associated with comprehension of synthetic speech appeared to suggest that natural speech was less cognitively demanding than its synthetic counterparts. A second comparison with the rated *Naturalness* showed a clearer trend towards reduced demand with increased *Naturalness*. Audio with lower ratings of *Intelligibility* appeared to recruit additional cognitive resources as evidenced by increased activity in

Optode 11. These results describe a relationship between both the qualitative rating of *Naturalness*, the different quality levels of the *Voices*, and their associated cognitive demand. Specifically, the use of natural, and highly-natural synthetic speech appeared to reduce the cognitive workload required to correctly comprehend the sentence as measured both in behavioral and biomarker representations.

The task used in this study was an active listening task which required listeners to both comprehend the audio content and verify comprehension by answering simple questions about the content. While cortical roles associated with language perception are heavily left-lateralized, results from PET studies have suggested that reduced synthetic speech quality recruits additional domains in the prefrontal cortex including the right-MFG [23] as reported in this study. Previous behavioral studies have suggested that comprehension tests required more cognitive load during synthetic speech than required by natural speech [4, 24, 25]. Working memory (WM), which is necessarily recruited during comprehension tasks, is increasingly tasked under synthetic speech where misleading acoustic-phonetic structures and absent natural cues increase phrase ambiguity [26]. As WM is a shared and limited resource, increasing WM demands has also been reported to further compound penalties associated with low-quality synthetic speech [24]. These findings describe a situation where the presence of low-quality synthetic speech imposes a baseline cognitive burden and necessitates the minimization of cognitive workload in the design of Auditory interfaces.

Since the original development of synthetic Text-to-speech (TTS) systems, progress on speech synthesis has evolved substantially, as has the contexts in which they are frequently used. In the past, TTS systems were often used primarily for accessibility devices, but the recent introduction of conversational assistants such as Apple's Siri, Microsoft's Cortana, Google Assistant, and Amazon's Echo, have again changed the nature of speech systems. Although auditory speech systems try to emulate natural language, at present these interfaces limit the ways in which the user can interact, meaning that users are already constrained to use these systems in ways completely unlike natural conversation. These new roles for synthetic speech systems require new cognitive design considerations as proper assessment of Human Computer Interaction requires an understanding of both the functional purpose and context in which the device is being used. However, as every scenario cannot be anticipated, design choices must be made regarding synthetic speech systems which optimize not just intelligibility [27] but overall user experience in the face of limited adaptability.

The results presented in this study suggest that design of synthetic speech affect not only the behavioral performance, but can also impact the listener's cognitive load directly during message comprehension. This study expands the research on the comprehension of synthetic speech and templates a role in the use of neuroimaging techniques to assess baseline cognitive requirements of such systems. The flexibility of fNIRS as an emerging portable and wearable neuroimaging technique enables comprehensive exploration of the cognitive demands of modern computer interfaces. Consistent with the Neuroergonomic approach, such fNIRS sensors could be used to guide design of speech synthesizers and audio interfaces not only in artificial laboratory contexts, but also monitor the way in which users actually interact with these systems in real world settings.

Acknowledgement. The authors would like to thank Dr. Paul Crawford for his help with this study. This study is made possible by a research award from the Intel Corporation. The content of the information herein does not necessarily reflect the position or the policy of the sponsor and no official endorsement should be inferred.

References

1. Treffner, P.J., Barrett, R.: Hands-free mobile phone speech while driving degrades coordination and control. Transp. Res. Part F Traffic Psychol. Behav. **7**(4–5), 229–246 (2004)
2. Dehais, F., Roy, Raphaëlle N., Gateau, T., Scannella, S.: Auditory alarm misperception in the cockpit: an EEG study of inattentional deafness. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) AC 2016. LNCS (LNAI), vol. 9743, pp. 177–187. Springer, Cham (2016). doi:10.1007/978-3-319-39955-3_17
3. Nusbaum, H., Magnuson, J.: Talker normalization: phonetic constancy as a cognitive process. In: Talker Variability Speech Process, pp. 109–132 (1997)
4. Delogu, C., Conte, S., Sementina, C.: Cognitive factors in the evaluation of synthetic speech. Speech Commun. **24**, 153–168 (1998)
5. Pisoni, D.: Perception of synthetic speech. In: Progress in Speech Synthesis, pp. 541–542 (1997)
6. Ayaz, H., Crawford, P., Curtin, A., Syed, M., Onaral, B., Beltman, W.M., Shewokis, P.A.: Differential prefrontal response during natural and synthetic speech perception: an fNIR based neuroergonomics study. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) AC 2013. LNCS (LNAI), vol. 8027, pp. 241–249. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39454-6_25
7. Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D.J., Pylkkänen, L.: Syntactic structure building in the anterior temporal lobe during natural story listening. Brain Lang. **120**(2), 163–173 (2012)
8. Benson, R.R., et al.: Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain Lang. **78**(3), 364–396 (2001)
9. Parasuraman, R., Christensen, J., Grafton, S.: Neuroergonomics: the brain in action and at work. Neuroimage **59**(1), 1–3 (2012)
10. Parasuraman, R.: Neuroergonomics: research and practice. Theor. Iss. Ergon. Sci. **4**(1–2), 5–20 (2003)
11. Parasuraman, R., Wilson, G.F.: Putting the brain to work: neuroergonomics past, present, and future. Hum. Factors **50**(3), 468–474 (2008)
12. Parasuraman, R., Rizzo, M.: Neuroergonomics. Oxford University Press, New York (2006)
13. Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P.A., McKendrick, R., Parasuraman, R.: Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. Front. Hum. Neurosci. **7**, 871 (2013)
14. Ayaz, H., Shewokis, P.A., Curtin, A., Izzetoglu, M., Izzetoglu, K., Onaral, B.: Using MazeSuite and functional near infrared spectroscopy to study learning in spatial navigation. J. Vis. Exp. **56**, 3443 (2011). doi:10.3791/3443
15. Quaresima, V., Bisconti, S., Ferrari, M.: A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. Brain Lang. **121**(2), 79–89 (2012)

16. Liu, Y., et al.: Measuring speaker–listener neural coupling with functional near infrared spectroscopy. Sci. Rep. **7**, 43293 (2017)
17. Derosière, G., Mandrick, K., Dray, G., Ward, T.E., Perrey, S.: NIRS-measured prefrontal cortex activity in neuroergonomics: strengths and weaknesses. Front. Hum. Neurosci. **7**, 583 (2013)
18. Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B.: Optical brain monitoring for operator training and mental workload assessment. Neuroimage **59**(1), 36–47 (2012)
19. Ayaz, H.: Analytical software and stimulus-presentation platform to utilize, visualize and analyze near-infrared spectroscopy measures. Drexel University (2005)
20. Ayaz, H., Izzetoglu, M., Shewokis, P.A., Onaral, B.: Sliding-window motion artifact rejection for functional near-infrared spectroscopy. Conf. Proc. IEEE Eng. Med. Biol. Soc. **2010**, 6567–6570 (2010)
21. Cope, M.: The Application of Near Infrared Spectroscopy to Non Invasive Monitoring of Cerebral Oxygenation in the Newborn Infant. University College, London (1991)
22. Lindquist, M.A., Meng Loh, J., Atlas, L.Y., Wager, T.D.: Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modelingb. Neuroimage **45**(1 Suppl), S187–S198 (2009)
23. Sharp, D.J., Scott, S.K., Wise, R.J.S.: Monitoring and the controlled processing of meaning: distinct prefrontal systems. Cereb. Cortex **14**(1), 1–10 (2004)
24. Francis, A.L., Nusbaum, H.C.: Effects of intelligibility on working memory demand for speech perception. Attention Percept. Psychophys. **71**(6), 1360–1374 (2009)
25. Ralston, J., Pisoni, D.: Comprehension of synthetic speech produced by rule: word monitoring and sentence-by-sentence listening times. Hum. Factors **33**(4), 471–491 (1991)
26. Nusbaum, H.C., Pisoni, D.B.: Constraints on the perception of synthetic speech generated by rule. Behav. Res. Methods Instrum. Comput. **17**(2), 235–242 (1985)
27. Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., Tang, Y.: Evaluating the intelligibility benefit of speech modifications in known noise conditions. Speech Commun. **55**(4), 572–585 (2013)

Dynamic Changes of ERPs in Gestaltzerfall Phenomena: Analysis Using Multi-data Selecting and Averaging Method

Mariko Funada[1(✉)], Tadashi Funada[2(✉)], and Yoshihide Igarashi[3(✉)]

[1] Department of Business Administration, Hakuoh University,
Oyama, Tochigi, Japan
mfunada@fc.hakuoh.ac.jp
[2] College of Science, Rikkyo University, Toshima-ku, Tokyo, Japan
04ca045w@rikkyo.ac.jp
[3] Gunam University, Kiryu, Gunma, Japan
fiftystorms@lilac.plala.or.jp

Abstract. In the design of BCI (Brain-Computer Interface) systems, brain signals distributed in a certain rage are often regarded as the identical signals. However, there exist some cases where such rough measurement of signals and data-collection are not good enough. Phenomena called "Gestaltzerfall" are examples of such cases. For example, Gestaltzerfall in the recognition of a letter is a phenomenon that a human does not recognize the meaning of the letter when he/she keeps watching the letter for a while. The purpose of this paper is to grasp the features of Gestaltzerfall and to understand the process getting to Gestaltzerfall by observing analysing the change of ERPs (Event-Related Potentials). From our experiments and data analysis we suggest that brain-signals are useful data to detect the variation in the pattern recognition process. We believe that such detection of the variation is applicable to the design of comfortable and universal BCI systems.

Keywords: Gestaltzerfall · ERPs (Event-Related Potentials) · Discriminant analysis · m-DSAM (multi-Data Selecting and Averaging Method)

1 Introduction

Present-day society relies very much upon AI (artificial intelligence) and HCI (human-computer interaction). We are asked to prepare suitable environments so that such present-day science and technology can be properly applied to increase the quality of human life. It is difficult but interesting how we can adjust the differences among individual characters, personalities and abilities when we develop some systems and/or devices relied on AI and HCI.

BCI (brain-computer interface) is one of the most important fields of HCI. In the design of BCI systems, brain signals distributed in a certain and proper range are often regarded as the identical output from the brains. However, there exist some cases where such rough measurement of signals and data-collection are not good enough. Phenomena called "Gestaltzerfall" are examples of such cases. For example, Gestaltzerfall in the

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 117–127, 2017.
DOI: 10.1007/978-3-319-58472-0_10

recognition of a letter is a phenomenon that a human does not recognize the meaning of the letter when he/she keeps watching the letter for a while. For such a case as a Gestaltzerfall phenomenon, the output signals from the brain may often go beyond their expected range. It is usually difficult for a BCI system to cope properly such a case.

In this paper we examine the following hypotheses:

Hypo1: It is possible to measure the occurrence of Gestaltzerfall in the letter recognition by ERPs.

Hypo2: It is possible to grasp the features of Gestaltzerfall and to understand the process getting to Gestalzerfall in the letter recognition by observing the changes of ERPs.

From our investigation we predict that Gestaltzerfall in the letter recognition occurs when high dimensional brain functions, which are related to the pattern recognition mechanism in the brains, do not work well [2, 3]. Furthermore, we confirm that ERPs [4] reflect somewhat significantly the recognition and judgement potentials as well as the visual evoked potentials for stimuli. Especially, ERPs are useful signals to measure the brain status caused by the high dimensional brain functions.

If the hypotheses listed above are affirmatively verified, then we would be able to design better BCI systems by considering Gestaltzerfall.

2 Methods

2.1 Experimental Methods

Since individual differences of ERPs are relatively large, the experiments were carried out by "single-case experimental design [5]".

The subject is a right-handed and 22 years old man. The experiments were carried out in the laboratory of the first author at Hakuoh University. We use 92 kinds of stimuli in the experiments. Each stimulus is one of 46 "Hiraganas" (Japanese phonetic symbols) in one of the two different fonts as shown in Fig. 1 (Ms-gothic and MS-Mincho). Each stimulus is displayed 40 times in a CRT display of 19 inches for one second. The time interval between two consecutive stimuli is 10 ms and the white screen is displayed during each interval.

A subject watches a stimulus, and inputs "0" or "1" by the keyboard. If here cognizes the meaning of the letter displayed in the screen, then he inputs "0", and otherwise inputs "1".

The single polar and eight channels of the "International 10 – 20 method" are used for the measurement of EEGs (electroencephalograms). The positions of the measurement are at C3, C4, Cz, and Pz. The base is A1 that is connected to A2. The sampling frequency for the A/D converter is 1 kHz.

(a)MS-Gothic (b)MS-Mincho

Fig. 1. Stimuli in the experiment

2.2 Preprocess of EEGs

The recorded EEGs are sent to an adaptive filter [11, 12]. For the proper function of the adaptive filter, we need to determine its cut-off frequencies as a function of time. This means that the numbers of data supplied to the adaptive filter are given as a function $g(t)$ of time ($0 \leqq t \leqq 000$ [ms]). From the previous research, $g(t)$ is in the form $k(1 - e - t/c)$. As shown in Fig. 2, we adopt that $k = 50$ and c is 100 [ms]. The data after the adaptive filter are normalized by taking the standard deviation of the averaged waveforms of EEGs. In this way, we obtain 100 repetitious data (EEGs) $D = \{x_1(t), x_2(t), \ldots, x_{40}(t)\}$ ($t = 1, 2, \ldots, 1000$). Then ERPs are derived from the normalized data of EEGs by the AM (Averaging Method) and the m-DSAM (multi-Data Selecting Averaging Method) [6].

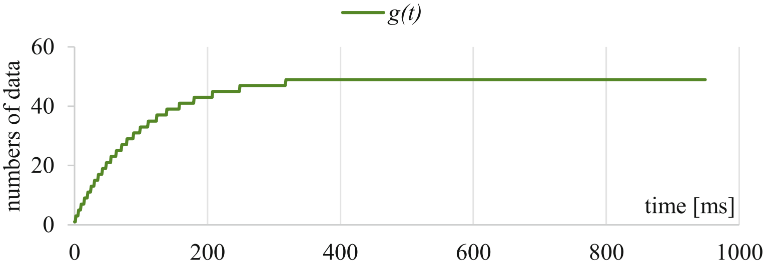


Fig. 2. The numbers of data sent to the adaptive filter, $g(t)$

2.3 The m-DSAM

Let $x_i(t)$ be a filtered and normalized data, where i is a data index and t is the latency (the time delay from the time when a letter is displayed). Let D be a data set of $x_i(t)$, and D_{before} and D_{after} be data subsets of D, which denotes "before Gestaltzerfall" and "after Gestaltzerfall" respectively. These subsets are determined by the inputs of the subject. In the following steps, D denotes D_{after} or D_{before}. The m-DSAM is a sequence of the calculation of the following steps:

- Step 1: For each i, $x_i(t)$ is transformed into two binary sequences $b_i^+(t)$ and $b_i^-(t)$ by the following formulae with threshold value L (in this paper, from the signal-noise ratios we determine $L = 0.5$):

$$b_i^+(t) = \begin{cases} 1 & \textit{if} \;\; x_i(t) > L \\ 0 & \textit{if} \;\; x_i(t) \le L \end{cases}, \qquad b_i^-(t) = \begin{cases} 0 & \textit{if} \;\; x_i(t) \ge -L \\ 1 & \textit{if} \;\; x_i(t) < -L \end{cases}.$$

- Step 2: The sum $B^+(t)$ of all $b_i^+(t)$ and the sum $B^-(t)$ of all $b_i^-(t)$ are calculated.
- Step 3: The maximum value MB^+ of $B^+(t)$ around the latency of a positive peak P is found. The minimum value MB^- of $B^-(t)$ around the latency of a negative peak N is found, too. Let T_P and T_N be the latencies such that $B^+(T_P) = MB^+$ and $B^-(T_N) = MB^-$, respectively.

- Step 4: We find the subsets D_{P} and D_{N} of D such that

$$D_{\mathrm{P}} = \{x_i(t)|x_i(T_{\mathrm{P}}) > L,\ x_i(t) \in D\},\quad D_{\mathrm{N}} = \{x_i(t)|x_i(T_{\mathrm{N}}) < -L,\ x_i(t) \in D\}$$

- Step 5: We calculate

$$ERP_{\mathrm{P}}(t) = \frac{1}{n_{\mathrm{P}}} \sum_{x_i(t) \in D_{\mathrm{P}}} x_i(t),\quad ERP_{\mathrm{N}}(t) = \frac{1}{n_{\mathrm{N}}} \sum_{x_i(t) \in D_{\mathrm{N}}} x_i(t),$$

where n_{p} and n_{N} are the numbers of elements in D_{p} and D_{N}, respectively.

- Step 6: For each peak P or N, we estimate $ERP_{\mathrm{P}}(t)$ or $ERP_{\mathrm{N}}(t)$, by the following formulae (a normal distribution determined by a certain algorithm [2]):

$$ERP_{\mathrm{Q}}(t) \approx f_{\mathrm{Q}}(t) = (-1)^{h(Q)} w_{\mathrm{Q}} \frac{1}{\sqrt{2\pi} s_{\mathrm{Q}}} \exp\left(-\frac{(t - T_{\mathrm{Q}})^2}{2 s_{\mathrm{Q}}^2}\right) \tag{1}$$

where Q is a P or an N, h(Q) is a function such that if Q = P then h(Q) = 0, and otherwise, h(Q) = 1, w_{Q} is the amplitude of peak Q, s_{Q} is the standard deviation, and T_{Q} is the latency of Q.

- Step 7: We calculate

$ERP_{\mathrm{m-DSAM}}(t) = \frac{1}{\bar{n}} \sum_{\mathrm{Q}} n_{\mathrm{Q}} f_{\mathrm{Q}}(t)$, where $\bar{n}$ is the average of all n_{Q}'s.

3 Results

3.1 Numbers of Repetition Until Gestaltzerfall Occurs

The Numbers of repetition of stimuli until the subject recognized Gestaltzerfall are shown in Table 1. The maximum and minimum numbers of repetitions are the same between both fonts, respectively. However, there was a significant difference of average between the number of repetitions in MS-Gothic font and MS-Mincho font ($t_{45} = -4.259$, $p < 0.001$).

Table 1. Numbers of repetitions until Gestaltzerfall occurred

Statistics	Fonts	
	MS-Gothic	MS-Mincho
Mean	19.48	23.48
Standard deviation	5.52	5.27
Maximum	35	35
Minimum	10	10

3.2 ERPs Calculated by AM and Distribution of Amplitude by m-DSAM

ERPs obtained by AM are shown in Fig. 3. The horizontal axis is time after a stimulus is displayed, and the vertical axis shows average of amplitude. The curves show ERPs before the subject recognized Gestalzerfall, and the dashed curves represent after the subject recognized Gestaltzerfall. Each number in parentheses shows the number of averaging operations. The dashed curves of "Gothic_after" includes six potentials called P1, N1, P2, N2, and P3. In both curves of "Gothic_before" and "Mincho_before", P2 and N2 do not appear clearly. Since watching a stimulus and judging the occurrence of Gestaltzerfall are easy tasks, visual evoked potentials P1 and N1 appear and P3 is also observed. P3 is concerned with recognition and judgement.

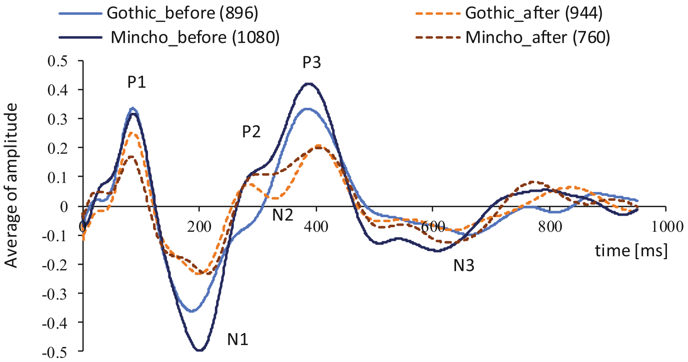


Fig. 3. ERPs calculated by AM (Position: Cz)

The averages of $B^+(t)$ and $B^-(t)$ for D_{before} and D_{after} are shown in Fig. 4. The solid curves represent the averages of $B^+(t)$, and the doted curves are the averages of $B^-(t)$. P1, N1, P3 and N3 appear clearly, and the frequencies of P2 and N2 are very small.

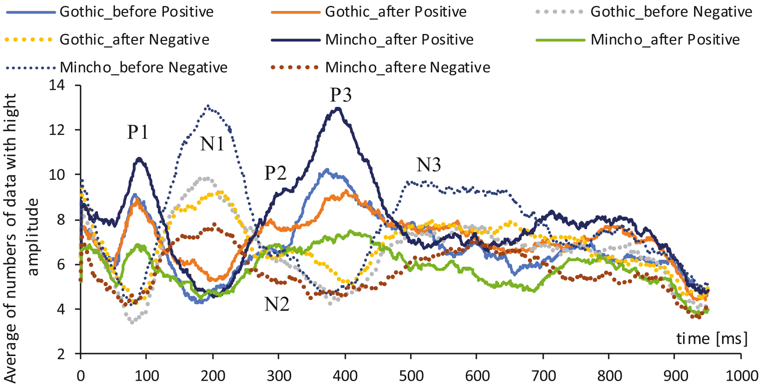


Fig. 4. Average of $B^+(t)$, $B^-(t)$ for "before" and "after" Gestaltzerfall (Position: Cz, Number of averaging operations: the same numbers in Fig. 3)

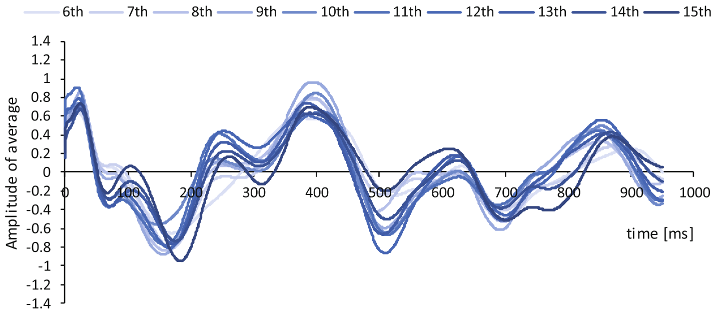


Fig. 5. The first 11-point moving averaged ERPs in the experiment (Position: Cz, letter: え)

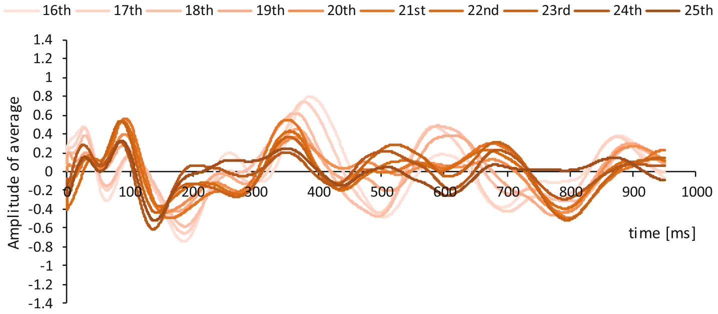


Fig. 6. The second 11-point moving averaged ERPs in the experiment (Position: Cz, letter: え)

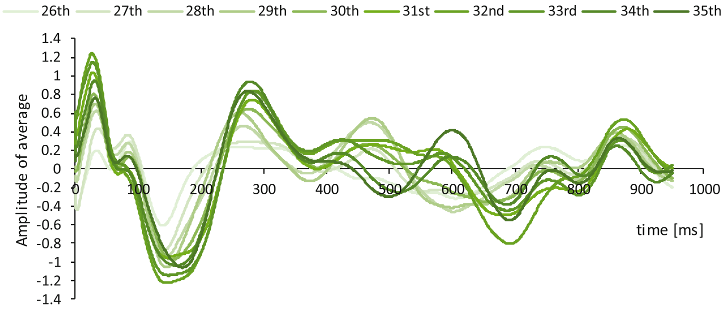


Fig. 7. The last 11-point moving averaged ERPs in the experiment (Position: Cz, letter: え)

Dynamic changes of ERPs for a stimulus “え” (in font MS-Gothic font) are depicted in Figs. 5, 6, and 7. Each figure includes 10 curves. Each curve is obtained by the average of consecutive eleven EEGs. The subject recognized Gestaltzerfall at the 14th repetition. Therefore 9th curve includes one EEG in Gestaltzerfall, 10th curve includes two EEGs in Gestaltzerfall, and so on. By the repetitions, the amplitudes of P3 become smaller and the latencies of P3 become shorter in Fig. 6, while the latencies of P3 become later in Fig. 7 than those of P3 in Fig. 5.

3.3 Parameters of AM and m-DSAM

We calculate the latencies and amplitudes of ERPs obtained by AM, and calculate the parameters (w_Q, T_Q, sd_Q, n_Q) (Q = Cz, Pz, C3, and C4) of ERPs obtained by m-DSAM. The averages and the standard deviations (sd) of amplitudes and latencies obtained by AM and by m-DSAM with Mincho letters are shown in Tables 2 and 3, respectively. Table 3 shows the averages and the standard deviations of those values by m-DSAM with Mincho letters. The variable *n*/*m* denotes the ratio of numbers, where n is the number of selected data by m-DSAM and m is the number of repetition before (after) Gestaltzerfall.

Table 2. Average and sd of amplitudes and latencies of ERPs in Cz obtained by AM

Parameters	Average before Gestaltzerfall average ± sd	Average after Gestltzerfall average ± sd
Cz_*AP*1	0.404 ± 0.185	0.336 ± 0.223
Cz_*AN*1	–0.585 ± 0.191	–0.475 ± 0.220
Cz_*AP*3	0.519 ± 0.168	0.413 ± 0.203
Cz_*P*1	84.9 ± 220.0	89.8 ± 37.6
Cz_*N*1	198.9 ± 27.6	195.5 ± 49.9
Cz_*P*3	381.2 ± 39.4	366.4 ± 61.6

Table 3. Average and sd of parameters in Cz obtained by m-DSAM

Parameters	Average before Gestaltzerfall average ± sd	Average after Gestltzerfall average ± sd
Cz_w_{P1}	54.7 ± 16.7	55.1 ± 22.6
Cz_w_{N1}	101.9 ± 24.3	78.9 ± 25.0
Cz_w_{P3}	103.5 ± 28.3	85.3 ± 28.8
Cz_T_{P1}	98.6 ± 21.0	109.8 ± 28.5
Cz_T_{N1}	200.3 ± 16.7	207.6 ± 31.1
Cz_T_{P3}	402.3 ± 33.0	419.5 ± 42.4
Cz_sd_{P1}	23.2 ± 4.6	24.7 ± 7.1
Cz_sd_{N1}	35.3 ± 4.3	31.6 ± 6.0
Cz_sd_{P3}	38.4 ± 9.0	35.2 ± 10.1
Cz_n_{P1}	15.0 ± 4.4	10.7 ± 3.5
Cz_n_{N1}	16.4 ± 3.5	11.5 ± 3.5
Cz_n_{P3}	15.9 ± 3.9	10.6 ± 3.0
Cz_n/m_{P1}	0.64 ± 0.14	0.68 ± 0.2
Cz_n/m_{N1}	0.71 ± 0.12	0.74 ± 0.31
Cz_n/m_{P3}	0.69 ± 0.13	0.68 ± 0.24

Table 4. Results of *t* test for difference of means of parameters for ERPs by AM (**: $p < 0.01$)

Channels	Cz			Pz			C3			C4		
Parameters	P1	N1	P3	P1	N1	P3	P1	N1	P3	P1	N1	P3
Amplitude		**	**		**	**		**	**		**	**
Latency											**	

Table 5. Results of *t* test for difference of means of parameters for ERP by m-DSAM (*: $p < 0.05$, **: $p < 0.01$)

Channels	Cz			Pz			C3			C4		
Parameters	P1	N1	P3	P1	N1	P3	P1	N1	P3	P1	N1	P3
w_Q		**	**			**		**	**		*	**
T_Q	*		**	*		**	*	*	**			**
sd_Q		**				*		*				
n_Q	**	**	**	**	**	**	**	**	**	**	**	**
n_Q/m_Q												

Since the potentials P1, N1 and P3 appeared clearly in Sect. 3.2, we examine t test for difference of means for paired data, and we obtain Tables 4 and 5. In Table 4, there are significant differences between means of amplitudes of ERPs before Gestaltzerfall and those of ERPs after Gestaltzerfall. In Table 5, there are significant differences regarding to the means of latencies as well as those of amplitudes. By the comparison between Tables 4 and 5, it is suggested that m-DSAM is superior to AM to extract meaningful variables.

3.4 Discriminant Analysis Using the Parameters

Using the parameters before and after Gestaltzerfall in Table 4, we execute discriminant analysis to classify the sets of parameters before and after Gestaltzerfall. As our results, these parameters are discriminated with 84.4% accuracy (shown in Table 6). Similarly, we execute the discriminant analysis for Table 5, we obtain 100% accuracy (shown in Table 7). This result shows that m-DSAM is encouraged for the discrimination analysis [7].

For each position and each font, we also examine the discriminant analysis. Almost all the discriminant functions discriminate with 100% accuracy.

Table 6. Classification results using amplitudes and latencies of P1, N1 and P3 in ERPs by AM

Original membership	Predicted group membership		Total
	Before Gestaltzerfall	Before Gestaltzerfall	
Before Gestaltzerfall	39	7	46
After Gestaltzerfall	7	39	46
Total	46	46	92

Table 7. Classification results using parameters of P1, N1 and P3 in ERPs by m-DSAM

Original membership	Predicted group membership		Total
	Before Gestaltzerfall	Before Gestaltzerfall	
Before Gestaltzerfall	46	0	46
After Gestaltzerfall	0	46	46
Total	46	46	92

3.5 ERPs Obtained by Moving Average and the Values of Discriminant Functions

Using the discriminant function for the position Cz and the font “MS-Gothic”, we calculate the values of the discriminant function for the moving averaged ERPs in Figs. 5, 6 and 7, and obtain Fig. 8. The dotted lines are amplitudes and latencies for the moving averaged ERPs, and the upper solid lines is the moving averaged inputs and the lower solid line is the values of discriminant function. The recognition by the subject corresponds to the discrimination by the discriminant function. The result suggests that the process of Gestaltzerfall is expressed well by discriminant function.

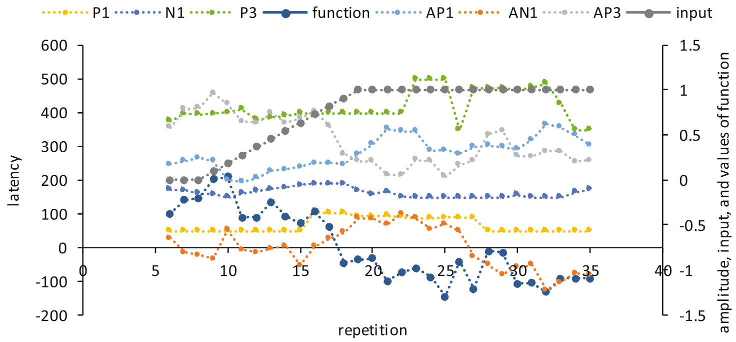


Fig. 8. Changes of amplitudes, latencies, input of moving averaged ERPs and values of the discriminant function (position: Cz)

4 Discussions

4.1 Validity of Experiments and Analysis

In these experiments, we insert a short interval (10 ms) between each stimulus due to the technical issue. During the short interval, a white screen is displayed. To observe Gestaltzerfall, the subject is required to keep watching an object. Even though a short interval is inserted between each stimulus, Gestaltzerfall is observed. Therefore, ERPs are well recorded in these experiments. We consider that the results of analysis for the ERPs are valid.

4.2 Discussion of the Hypothesis Hypo1

In Sect. 3.3, we examine *t* Test for the paired parameters, amplitudes, and latencies of Averaged ERPs. And we also examine *t* Test for the parameters obtained by m-DSAM. There are significant differences between Gestaltzerfall and non-Gestaltzerfall with $p < 0.01$ level for the amplitudes of N1 and P3 for all positions. Furthermore, there are significant differences between Gestaltzerfall and non-Gestaltzerfall with $p < 0.01$ or $p < 0.05$ levels for parameters of latencies and amplitudes obtained by m-DSAM (for all positions). Therefore, it is suggested that our hypothesis "Hypo1: It is possible to measure the occurrence of Gestaltzerfall in the letter recognition by ERPs" is affirmatively confirmed from our experiments and analysis.

4.3 Discussion of the Hypothesis Hypo2

Using the functions for all positions obtained in Sect. 3.3, we calculate the values of discriminant functions. Some of these values are plotted in Fig. 9. The horizontal axis corresponds to the values of discriminant function, and the vertical axis corresponds to moving averaged inputs. The curve of Cz represents the process of Gestaltzerfall precisely. Though other curves include some errors in the indication of Gestaltzerfall, the entire tendency of the curves represents the process of Gestaltzerfall. Similar results are obtained for other stimuli and for all positions. Therefore, our results support affirmatively our hypothesis, "Hypo2: It is possible to grasp the features of Gestaltzerfall and to understand the process getting to Gestaltzerfall in the letter recognition by observing ERPs".

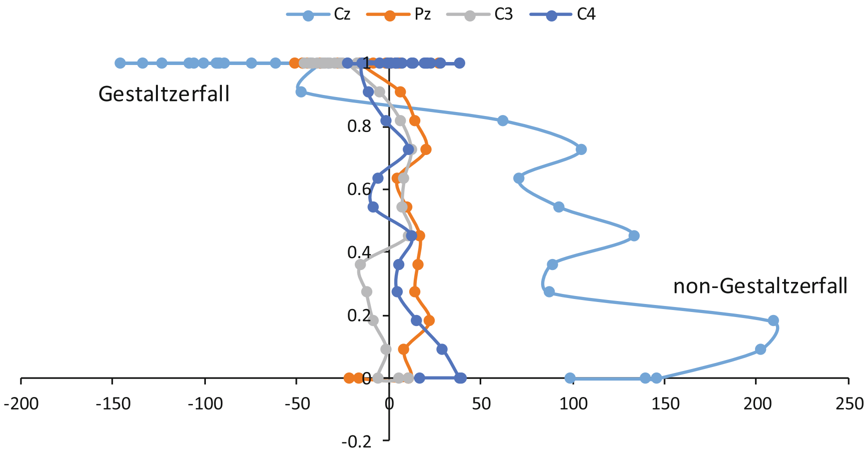


Fig. 9. The process of Gestaltzerfall using the values of discriminant function and moving averaged input of subjects (positions: Cz, Pz, C3, and C4)

5 Summary

Our trials are summarized as follows:

(1) Our results support affirmatively the hypotheses Hypo1 and Hypo2.
(2) In order to examine these hypotheses, we design experiments. Each time of the experiments, the subject was required to keep watching a Japanese Phonetic letter (“Hiragana”) for one minute. These experiments are executed by “single-case experimental design”.
(3) In order to examine Hypo1, we calculate the parameters of ERPs obtained by AM and m-DSAM, and examine *t* Test. As the result, Hypo1 is affirmatively supported by our experiments and analysis.
(4) In order to examine Hypo2, we execute the discriminant analysis for the parameters of ERPs obtained by AM and m-DSAM. As the result, Hype2 is affirmatively supported by our experiments and analysis.
(5) In both examinations for the hypotheses, m-DSAM works effectively.

The results discussed in this paper may suggest something important about understanding the characteristics of human brains. We expect that the research direction discussed in this paper would be able to contribute to the design of comfortable, stress-less, and universal BCI systems. These issues are interesting and worthy for the future investigation.

References

1. Faust, C.: Über Gestaltzerfall als Symptom des parieto-occipitalen Übergangsgebiets bei doppelseitiger Verletzung nach Hirnschuß. Nervenarzt **18**, 103–115 (1947)
2. Ninose, Y., Gyoba, J.: Delays produced by prolonged viewing in the recognition of Kanji characters: analysis of the ‘Gestaltzerfall’ phenomenon. Shinrigaku kenkyu **67**(3), 227–231 (1996)
3. Lee, N.-C.: Perceptual Coherence of Chinese Characters: Orthographic Satiation and Disorganization, University of Edinburgh (2007)
4. Picton, T.W., et al.: Guideline for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology **37**, 128–152 (2000)
5. Barlow, D.H., Hersen, M.: Single case experimental designs: Strategies for studying behavior change, 2nd edn. Pergamon Books, New York (1984)
6. Funada, M., Funada, T., Igarashi, Y.: Discrimination in good-trained brain states for brain computer interface. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) AC 2015. LNCS (LNAI), vol. 9183, pp. 187–198. Springer, Cham (2015). doi:10.1007/978-3-319-20816-9_19
7. Härdle, W., Simar, L.: Applied Multivariate Statistical Analysis, 2nd edn. Springer, Heidelberg (2007)

Decision-Making for Adaptive Digital Escape Route Signage Competing with Environmental Cues: Cognitive Tunneling in High-Stress Evacuation Situations

Sonja Th. Kwee-Meier(✉), Wolfgang Kabuss, Alexander Mertens, and Christopher M. Schlick

Institute of Industrial Engineering and Ergonomics,
RWTH Aachen University, 52062 Aachen, Germany
`{s.meier,w.kabuss,a.mertens}@iaw.rwth-aachen.de`

Abstract. The relationship between stress and cognitive tunneling, i.e. the narrowed attentional focus, is long known. In evacuations, escape route signs have to compete with environmental cues, such as corridor width and lightning. New digital escape route signage can help to guide passengers situation-adequately under dangerous conditions, for instance fires. So far, little is known about the effect of cognitive tunneling on way finding in high-stress evacuation situations and its meaning for the potential to exert influence on direction decisions by digital escape route signage. Therefore, we have conducted an age-differentiated study in a virtual environment of a corridor system with digital escape route signage and competing environmental influences. 60 participants, 30 young (20–30 years) and 30 elderly (60–79 years), made direction decisions in a total of 40 T-intersections in three conditions. There was a low-strain everyday condition and two evacuation conditions with higher mental, emotional and physical demands. Significant differences in direction decisions were found between the everyday and the two evacuation conditions. The participants payed more attention on environmental cues in the everyday situation, while strongly focusing on the visual search for escape route signage in the evacuation conditions, suggesting a pronounced tunneling effect in these high-stress situations.

Keywords: Escape route signage · Digital signs · Stress · Evacuation · Emergency · Cognitive tunneling · Experimental study · Age-differentiated

1 Introduction

1.1 Background

Conditions of emergency situations, such as fire, in large built structures, often necessitate the adaption of escape routes to ensure a safe evacuation. Digital escape route signage allows adaptation to the specific emergency conditions. However, people are not familiar with digital escape route signage and might feel irritated if the displayed escape route direction changes right in front of them. As a result, they might

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 128–140, 2017.
DOI: 10.1007/978-3-319-58472-0_11

follow other factors of influence on direction choices. There are three main categories of such influences, environmental influences, social influences and influences of familiarity with places and routes. Regarding environmental influences, corridor width and lightning, for example, exert an influence on route choice in the way that people tend to take routes that are wider and brighter, first shown by Taylor and Socov [1] and confirmed, for instance, by Vilar et al. [2]. Furthermore, they try to maintain the initial direction of travel, referred to as least-angle strategy by Conroy-Dalton [3] and Hochmair and Frank [4]. Social influences are mostly based on the affiliation theory by Mawson [5, 6], studies of Sime [7, 8] and normative social behaviors, differentiated from informative social behaviors by Deutsch and Gerard [9]. Various investigations have found that people tend to move towards other people, such as Mawson [5, 6], Nilsson et al. [10] and Kinateder et al. [11]. Influences of familiarity with places and routes have been found by Butcher and Parnell [12] and in evacuation field studies by Sime [7, 8] and Shields and Boyce [13]. Moreover, perceived risk, such as in emergency situations, leads to pronounced affective behaviors [14–16]. Analytic thoughts are hindered by time pressure, further leading to affective decisions [17]. Keinan et al. [18] found evidence for rash decisions under stress, supported by the eye-tracking study by Stankovic [19].

1.2 Cognitive Tunneling

Easterbrook [20] was the first to describe what became known as the effect of cognitive tunneling, also referred to by attentional tunneling and attentional bias. Attention was observed to be focused on specific cues, referred to as “central cues”, and less on other cues, referred to by “peripheral cues”, the more people experienced stress resulting in emotional arousal, referred to by Easterbrook [20] as “drive”. Easterbrook’s groundwork [20] was confirmed and refined by various studies. In the field of aviation, Yeh and Wickens [21] reported on attentional tunneling in pilots, missing out on unexpected targets, with the attention overly drawn by cued objects. Briggs et al. [22] reported on the risks of cognitive tunneling due to emotionally involving telephone conversations by drivers, unlikely to be solved by hands-free cell phones [23]. Dixon et al. [24] point to the problem of attentional tunneling in the context of surgery endoscopy with the aid of AR, though not drawing a clear line to inattentional blindness [25] with the famous example of the “invisible gorilla” [26].

1.3 Aim and Scope

Despite the variety of the aforementioned studies on cognitive tunneling, to the best of the authors’ knowledge, there has been no study investigating effects of cognitive tunneling in high-stress evacuation situations with regard to digital escape route signage. Only the study by Vilar et al. [2] referred to this phenomenon though with regard to the established static escape route signage and with the participants not moving in reality. Their findings indicate that the participants decided in compliance with the established static escape route signage, especially in emergency conditions, but payed more attention to environmental influences in their condition without signs. Vilar et al. [27] pointed to the high arousal experienced in emergency conditions. According the Yerkes-Dodson law [28, referred to by 29], the function of performance in dependence has an inverted U-shape. The influence of higher stress levels including physical demands were not investigated in these studies. Moreover, in a prior study, we found that the consistency of decisions in favor of specific digital escape route sign types decreased with mental, emotional and physical stress [30].

Therefore, the scope of the present paper is the investigation on the effect of cognitive tunneling in high-stress emergency situations with mental, emotional and physical stress and its meaning for the use of digital escape route signage. A study in a virtual environment of a corridor system with adaptive digital signage and competing environmental influences, i.e. corridor widths, lightning and angle between the current and the follow-up path, was conducted. In consideration of the ageing society in most western countries, an age-differentiated design was chosen. Furthermore, age effects were expected as we had found differences in preferences of digital escape route sign types in a prior age-differentiated study using paired comparisons [30]. Hence, an age-differentiated study design with 30 young and 30 elderly participants was used. There were three conditions varying in applied stressors, i.e. background noise and simulated fire in the virtual environment (VE), and physical demands in terms of walking speed, which was not only realized in the VE but also matched with the real world requirements using a tread mill. The decisions in favor of the digital escape route signs were analyzed with the focus on differences between low- and high-stress situations.

2 Method

2.1 Participants

60 participants were recruited externally. There were two age groups. The young group consisted of 30 participants (15 f, 15 m) aged between 20 and 30 years ($M = 25.27$, $SD = 2.84$), the old group consisted of 30 participants (15 f, 15 m) aged between 60 and 79 years ($M = 68.20$, $SD = 5.14$) to cover potential age effects. Ethical approval was obtained by the ethics committee of the University Hospital of the RWTH Aachen University (EK 190/16). No participant had to be dropped out of the recruited ones, as exclusion criteria were communicated beforehand. These were defined by severe illnesses, especially coronary heart diseases, cardiac pacemakers, bypasses, artificial arteries and joints, pregnancy as well as mental, mobility, visual and hearing impairments. No participant violated the minimum boundary for visual acuity of 0.8 or suffered from a red-green colour-deficiency. No participant had less than 80% correct responses in a stimulus-response pre-test (see Sect. 2.5).

2.2 Apparatus

The photos in Fig. 1 show the view over the shoulder of a participant onto the virtual environment (VE) and the experimental setup in the laboratory with a participant walking on the treadmill looking at the VE. The h/p cosmos mercury treadmill had a walking surface of 1.5 by 0.5 m. The preinstalled customizable safety harness system prevented participants from injures in case of stumbling, which was of particular importance as persons with an age up to 79 years were included in the study. The VE was displayed on a 65″ monitor in walking direction on the treadmill in a distance of 1.5 m. It was mounted on a height-adjustable table to ensure the best immersion possible. The eyes of the participant were on 2/3 of the monitor's height. For direction choices, i.e. left of right, two Wii remotes, one for each hand, were used. Two external loudspeakers at either side of the monitor were played looped acoustic background noise during the evacuation conditions. There was no natural light to the laboratory to keep light conditions constant. Therefore, all windows were opaquely closed.

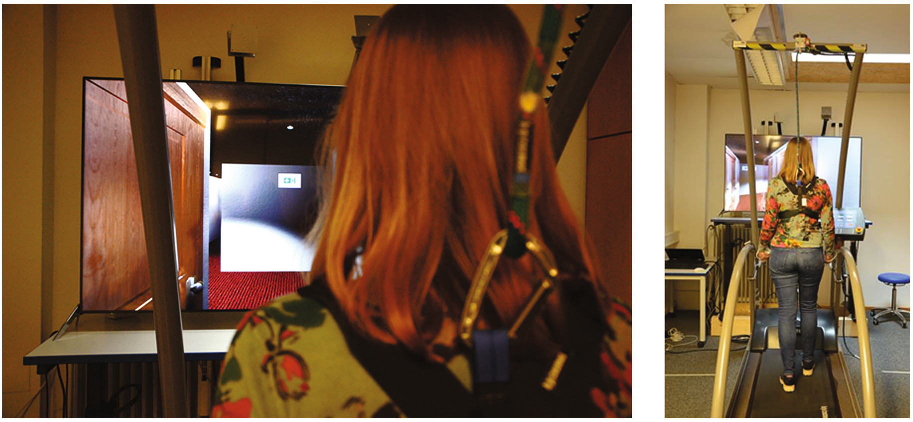


Fig. 1. Left: view over the shoulder of a participant onto a corridor intersection in the virtual environment. Right: view from behind on the participant secured in the harness on the treadmill in front of the 65″ monitor. The opaquely closed windows were slightly opened to improve light conditions for the photos.

2.3 Virtual Environment

The virtual environment (VE) was built in Unity 3D. All potential paths of the participants were predefined by waypoints. The current position was linearly interpolated according to the individual speed of the participant walking on the treadmill. There were 40 models of rooms according to the 40 decision in 40 intersections. The room order was randomized, except for the first two rooms. These were always rooms without signs though with environmental cues potentially influencing direction choices, i.e. differences in lighting, corridor width and angle between the right and the left hand side. Each room model had several inner and outer opaque walls to create corridors and was based on a template room, which contained all possible waypoints, triggers and settings (lighting, sign type, sign direction, sign change, etc.). One corridor led from the starting point of the room to the intersection. The length of this corridor was adaptive to the speed of the participant to offer each participant the same potential decision time of

4000 ms and 2000 ms after sign display change, regardless of his/her individual speed. For the way from the intersection to the starting point of the next room, there was a left and a right path, consisting of several corridor parts. Taking the left and the right path after the T-intersection together, the path through every room model was a closed loop by design. However, the participant was prevented from going backwards by a wall, which was automatically inserted at the end of the path that was not chosen, not noticeable by the participant. On one side, the first part of the corridor after the T-intersection was narrower, i.e. 1.30 m, darker with only one light source at the far end. The path was sharply angled, inhibiting the participant to see how the path would go on. The corridor on the other side was wider, i.e. 2.60 m, and much brighter, as it was equipped with nine light sources. The path was less angled, allowing the participant to see around the corner and that the path almost led in the direction of his current direction of travel. The escape route sign displays, sized 0.37 m by 0.22 m, were placed in the top-middle position in the T-intersection at a height of 1.80 m above the floor in the VE to ensure the highest possible effectiveness of the signage [31]. The background of the sign displays was modelled as emissive light source.

Invisible barriers were used to start and end the decision input option. Each barrier was modeled as a single shot event, which was triggered when the virtual camera, i.e. the virtual position of the participant in the VE, passed the barrier. When crossing the barrier at the beginning of the corridor directly leading towards the next intersection, the participant's decision was triggered by the rumble function, i.e. the Wii remotes in both hands started to vibrate. As soon as the participant decided for a direction, only the Wii remote in the corresponding hand continued to vibrate, while the other one turned still as feedback for the participant. For example, the participant pressed the button on the Wii remote in the left hand, deciding to turn left at the intersection, this Wii remote continued to vibrate, while the Wii remote in the right hand turned still. Technically, after the direction decision, the waypoints of the left or right path were added to the queue according to the choice of the participant, followed by adding the waypoints of the first path of the next room, leading towards the next intersection, to the queue. The conditions were highly controlled and the same for each participant. Hence, participants could not stand still or turn, not in the VE and, obviously, not on the treadmill. Hence, if the virtual camera reached an intersection but no input, i.e. direction decision, had been entered, the movement was continued "against the wall". The whole monitor turned black for 3000 ms. Then, the camera was placed at the starting point of the next room. This procedure was shown in the VE training to avoid potential surprises and unintended behavioral responses by participants, such as stopping on the treadmill and stumbling during the actual experimental conditions.

2.4 Experimental Design

A mixed design was used in the present study. All participants completed 40 randomized decision-making tasks for directions at 40 T-intersections in each of three fully permuted experimental conditions varying in applied stressors and walking speed, individually matched between the VE and the treadmill (see Table 1).

Table 1. Experimental conditions varying in walking speed and applied stressors.

	Condition 40noS	Condition 40 + S	Condition 65 + S
Situation	Everyday	Evacuation	Evacuation
Walking speed[a]	40%	40%	65%
Stressors	–	Fire and background noise	Fire and background noise

[a] Relative to the achieved walking speed in the 10-meter Fast Walk Test (10 mFWT).

Each intersection led onto a lighter, wider and less angled route on one side offering the view into the following corridors in the direction of initial travel, and onto a darker, narrower and sharper angled path on the other side. 32 of the 40 intersections were equipped with a digital escape route sign based on ISO 7010 [32] pointing to the right or the left. Four sign types were used, each one at eight intersections: *Standard* without further elements, *Updated* with temporal update information, *Dynamic* with a green flashing frame and *Crossed* with a cross for the opposite direction indicating that it is blocked. The direction of the displayed escape route sign was changed half of the times while the participant was approaching. The distance to the next T-intersection was always the same and independent of the direction decision of the participant, facilitated by the automatically adapting VE, not noticeable for the participant (see Sect. 2.3).

Each participant completed the walk through the automatically adapting corridor system three times under three experimental conditions varying in the situation, i.e. everyday vs. emergency, applied stressors and physical demands (see Table 1). The least demanding condition was the 40noS condition, defined as everyday condition. The walking speed was 40% of the maximum walking speed assessed in a 10-meter Fast Walk Test (10 mFWT [33], see Sect. 2.5) corresponding to a comfortable speed, calculated on the data basis provided by Bohannon [34]. The walking speed was realized on the treadmill and as well within the virtual environment. The speed was the same in the evacuation condition 40 + S. However, stressors in form of fire, coming from underneath the doors in the corridors, and continuous background noise, i.e. a looped sequence of echoed incomprehensible crowd mumbling, played at about 68 dB (A), were applied. The highest demanding condition was the evacuation condition 65 + S with the same stressors as in 40 + S but at 65% of the maximum walking speed in the 10mFWT, calculated as the ratio of the maximum evacuation speeds according to the MSC Circ. 1238 [35] and maximum walking speeds [34]. The dependent variable is the percentage of direction decisions in compliance with the escape route direction recommended by the digital signage.

2.5 Procedure

Upon arrival, participants were orally informed about the procedure, before they signed an informed consent form. They answered several questionnaires, for instance on demographics. The participants were tested for visual acuity, red-green colour-deficiency and reactions in a computer-based arrow stimulus-response (S-R) test. A 10-meter Fast Walk Test (10 mFWT [33]) was performed to assess the maximum walking speed of participants to adapt the treadmill speed to the participants' gait abilities in the trials and to receive comparable results from persons with different characteristics, such as age and gender. The participants were instructed to walk as fast as they safely could without running from one marking on the floor to another, as this approach has established since Bohannon et al. [36]. The markings were three meters before and after the measuring distance to inhibit effects of acceleration and deceleration within the measuring distance of the ten meters. Furthermore, Watson [37] recommended the destination marking not to be the ending point of the measuring distance to avoid premature deceleration and also distraction from the walking test, while time is still measured. Participants were asked to perform the test twice. If the slower speed of the two tests deviated by more than 10% from the faster, the test was repeated and the results compared again. The average of the two tests served as reference value for the following parts.

As a treadmill was used in the experimental conditions, each participant completed a five-minute training on the treadmill. Thereby, every participant became used to walking on a treadmill and between-subjects effects of potential prior experiences were minimized. Safety was ensured by a harness system (see Fig. 1 right). It prevented participants from injuries in case of stumbling. Before the training, the harness was adjusted to the participant's girth and body height. This adaptation not only offered optimal safety for the participant but also ensured that no body weight was loaded on the harness during the trials. The maximum speed during the experimental conditions was 65% of the walking speed achieved in the 10 mFWT. Thus, the walking speed was gradually increased to this maximum experimental speed reaching 25% of this speed after one minute, 50% after two minutes, 75% after three minutes and 100% of the maximum experimental speed after four minutes, i.e. 65% of the speed in the 10 mFWT, which was kept constant during the last minute. Afterwards, decision-making with the input devices, i.e. the Wii remotes, was practiced. The training in the VE through four intersections without environmental influences nor escape route signage ensured understanding and allowed a stronger focus on the task during the experimental conditions rather than on the operation.

Prior to the experimental conditions and in between them participants had ten-minute breaks in sitting position to minimize influences from the previous part on the following condition. The situation, i.e. everyday vs. evacuation, was described to the participants upon the start of each condition with a standardized text with the differentiation between an everyday and an evacuation situation. Participants were told to imagine that they were in corridor of passenger cabins inside a cruise ship. They were instructed to try get out. In the everyday scenario, this instruction was described as a general ship announcement, clearly indicating that there is no emergency, but as result of a fire alarm in the evacuation scenarios. Further instructions were that the participants could not stand still nor turn around in the VE and that they were supposed to make their decisions as they would in reality. Each experimental condition was concluded by a short semi-structured interview on strain and decision-making.

2.6 Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics 23. The level of significance was set to $\alpha = .05$. All tests were two-tailed. The data was not normally distributed. Therefore, the non-parametric Friedman's ANOVA for repeated measures was used. Pairwise comparisons between the conditions were performed by the non-parametric Wilcoxon signed-rank test for dependent samples, reported according to Field [38] by the test statistic T as the lowest value of the two types of ranks, i.e. positive and negative ranks, and the effect size $r = Z/\sqrt{N}$, with N for the number of observations. Bonferroni correction was applied to pairwise comparisons for evaluating significance, $\alpha_c = .05/c$, with c for the number of comparisons. Marginally significant was defined as $\alpha > \alpha_{marg} > \alpha_c$. Exact statistics instead of asymptotic statistics were used for all tests.

3 Results

At intersections with congruent environmental cues and displayed escape route directions, the compliance rate with the signage was very high, i.e. 89% $(SE = 1.92)$. This is little surprising. However, the decisions at the intersections with conflicting cues are by far more meaningful. These were the intersections with escape route signage pointing in one direction but the environmental cues such as corridor width and illumination suggesting a decision for the opposite direction. A change in decision-making behavior, depicted in Fig. 2, was observed between the everyday condition and the emergency conditions with applied stressors in form of fire from underneath the doors in the VE and continuous background noise. The decisions at intersections with matched environmental cues and escape route sign directions as well as the comparison with these are not further discussed in this article.

There was a main effect of condition, indicated by Friedman's ANOVA, $\chi^2(2) = 22.34$, $p < .001$. The depicted data in Fig. 2 suggests a stronger effect of condition on young than on elderly people, confirmed by separated Friedman's ANOVAs, young group: $\chi^2(2) = 20.18$, $p < .001$; old group: $\chi^2(2) = 4.89$, $p = .089$. Hence, the influence of condition on decisions in favor of escape route sign recommended directions was highly significant in young participants, but not quite significant in elderly participants. Bonferroni corrected pairwise comparisons between the everyday condition and both emergency conditions revealed significant increases in decisions in compliance with escape route signage in the young group for both emergency conditions, condition 40 + S: $T = 3$, $p < .001$, $r = -.51$, condition 65 + S: $T = 4$, $p < .001$, $r = -.48$, indicating a large and a medium to large effect size. Regarding the old group, the increase in decisions in favor of the escape route sign direction from the 40noS condition, without any stressors at a comfortable walking speed, to the 65 + S condition, with fire in the VE and background noise at evacuation speed, was significant, $T = 8$, $p = .021$, $r = -.29$, while the increase to the 40 + S, with applied stressors but without increased speed, was only marginally significant, $T = 8$, $p = .034$, $r = -.27$, both indicating small effect sizes.

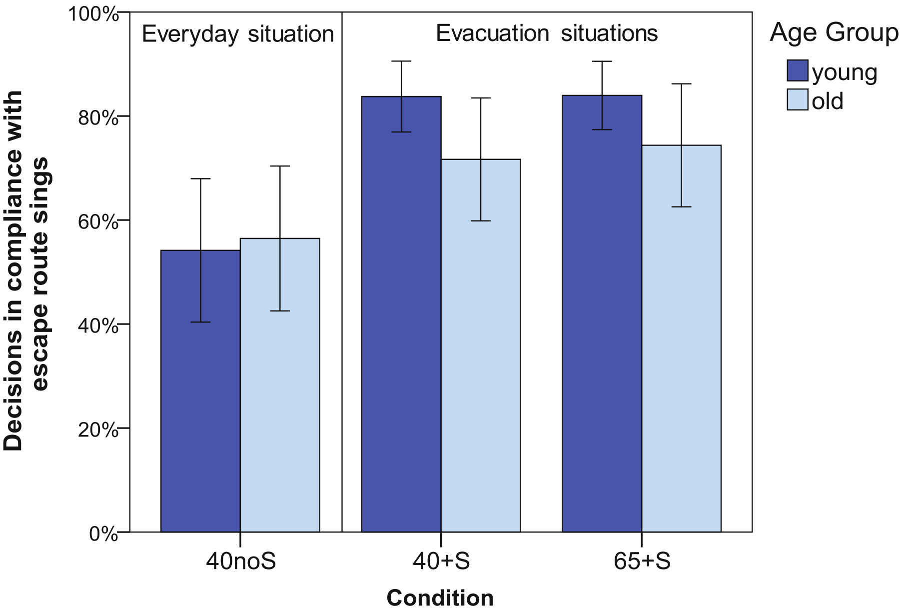


Fig. 2. Decisions for directions in compliance with the digital escape route signs at intersections with the displayed direction competing with the direction suggested by the environmental influences, over the three experimental conditions varying in the walking speed between 40% and 65% of the maximum individual speed and the application of stressors in the two evacuation conditions 40 + S and 65 + S.

4 Discussion

In summary, we have conducted an age-differentiated study in a virtual environment with digital escape route signage and competing environmental influences, such as corridor width and lightning. The two age groups consisted each of 30 participants with the young group aged from 20–30 years and the old group aged from 60–79 years. Their decisions varied between the three conditions, i.e. a low-demanding everyday condition without stressors at a comfortable walking speed (40noS) and two evacuation conditions with simulated fire and background noise with the condition 40 + S of these two evacuation conditions at a comfortable walking speed and the other at a more demanding typical evacuation speed.

Significant differences in direction decisions were found between the everyday and the two evacuation conditions when sign directions conflict with environmental behavior, suggesting a pronounced tunneling effect in the high-stress situations. A high compliance of direction choices with the displayed directions on digital signage was observed in the evacuation conditions, especially in the young group. It is noteworthy that the effect was stable despite the randomized adaptation of the displayed escape route direction to – in reality – potentially changing emergency conditions, such as fires. Participants reported in the semi-structured interviews to have felt more stressed in the evacuation conditions and to have, hence, focused more strongly on the visual search for the digital escape route signs followed by the complying decisions. In contrast, in the everyday situation, participants payed more attention to the other environmental cues and made their decisions often in favor of the corridor brightness and width. This is particularly interesting, when taking into account that there would have been good reason to hypothesize the opposite as the influence of affect and limitation of logical reasoning had been found to increase, for example, with risk perception [14, 16] and perceived stress [39]. However, the attentional tunneling in favor of escape route signage is in line with the escape study by Vilar et al. [2], using the established static signs and finding high compliance rates in emergency conditions.

The observed, strong focus on digital escape route sings, especially in evacuation conditions emphasize the importance of up-to-date escape route signage. People are likely to miss out on information provided from other sources in evacuation situations, as they narrowly focus on escape route signs during escape, like the pilots in the study by Yeh and Wickens [21] who missed out on unexpected targets when focused on cued targets.

Nevertheless, the findings should be interpreted in light of the limitations of the conducted research. There are limitations regarding the sample as well as the virtual environment (VE). Strict exclusion criteria were defined. On the one hand, they ensured participants' health and safety. On the other hand, these exclusion criteria might have biased the sample resulting in a healthier and fitter sample than the average population, especially regarding the elderly participants. Regarding the VE, real world conditions would, of course, have been more desirable, but were discarded for ethical and safety reasons.

5 Conclusion

Evacuation conditions were found to lead to cognitive tunneling and a strong attentional focus on digital escape route signs, especially in young persons. These findings emphasize the potential and importance of up-to-date digital escape route signage to support people's way finding during evacuations. Obsolete direction information provided by traditional static signage is not only useless but dangerous because of the strong focus on signage, especially under evacuation conditions, in which correct direction choices are most critical.

Acknowledgements. The study was conducted within the project SIREVA in the context of the national program "Research for Civil Security" and the call "Maritime Security" (FKZ: 13N12954). The findings of cognitive tunneling and their meaning for the design of human-machine interfaces used in high-stress situations are further of relevance for the project DURCHBLICK in the context of the national program "Research for Civil Security" and the call "Civil Security – Aspects and Measures of Coping with Terrorism", in which the investigations will be continued in the direction of human-machine interaction under stress. Both projects were/are funded by the German Federal Ministry of Education and Research (BMBF).

References

1. Taylor, L., Socov, E.: The movement of people toward lights. J. Illum. Eng. Soc. **3**, 237–241 (1974). doi:10.1080/00994480.1974.10732257
2. Vilar, E., Rebelo, F., Noriega, P., Duarte, E., Mayhorn, C.B.: Effects of competing environmental variables and signage on route-choices in simulated everyday and emergency way finding situations. Ergonomics **57**(4), 511–524 (2014). doi:10.1080/00140139.2014.895054
3. Conroy-Dalton, R.: The secret is to follow your nose: route path selection and angularity. Environ. Behav. **35**(1), 107–131 (2003). doi:10.1177/0013916502238867
4. Hochmair, H., Frank, A.: Influence of estimation errors on way finding decisions in unknown street networks – analyzing the least-angle strategy. Spatial Cogn. Comput. **2**, 283–313 (2000). doi:10.1023/A:1015566423907
5. Mawson, A.R.: Panic behavior: a review and a new hypothesis. In: Paper Presented at the 9th World Congress of Sociology. Doubleday, Doran and Company, York (1978)
6. Mawson, A.R.: Is the concept of panic useful for scientific purposes? Behavior in fires. In: Proceedings of the Second International Seminar on Behavior in Fire Emergencies. NBS Report NBSIR-802070. U.S. Department of Commerce, Washington, DC (1980)
7. Sime, J.D.: Affiliative behaviour during escape to building exits. J. Environ. Psychol. **3**(1), 21–41 (1983). doi:10.1016/S0272-4944(83)80019-X
8. Sime, J.D.: Movement toward the familiar person and place affiliation in a fire entrapment setting. Environ. Behav. **17**(6), 697–724 (1985). doi:10.1177/0013916585176003
9. Deutsch, M., Gerard, H.B.: A study of normative and informational social influences upon individual judgment. J. Abnorm. Soc. Psychol. **51**(3), 629–636 (1955). doi:10.1037/h0046408
10. Nilsson, D., Johansson, M., Frantzich, H.: Evacuation experiment in a road tunnel: a study of human behaviour and technical installations. Fire Saf. J. **44**(4), 458–468 (2009). doi:10.1016/j.firesaf.2008.09.009
11. Kinateder, M., Ronchi, E., Gromer, D., Müller, M., Jost, M., Nehfischer, M., Mühlberger, A., Pauli, P.: Social influence on route choice in a virtual reality tunnel fire. Transp. Res. Part F Traffic Psychol. Behav. **6**, 116–125 (2014). doi:10.1016/j.trf.2014.06.003
12. Butcher, E.G., Parnell, A.C.: Designing for Fire Safety. Wiley, Hoboken (1983)
13. Shields, T.J., Boyce, K.: A study of evacuation from large retail stores. Fire Saf. J. **35**(1), 25–49 (2000). doi:10.1016/S0379-7112(00)00013-8
14. Alhakami, A.S., Slovic, P.: A psychological study of the inverse relationship between perceived risk and perceived benefit. Risk Anal. **14**(6), 1085–1096 (1994). doi:10.1111/j.1539-6924.1994.tb00080.x
15. Peters, E., Slovic, P.: The role of affect and worldviews as orienting dispositions in the perception and acceptance of nuclear Power1. J. Appl. Soc. Psychol. **26**(16), 1427–1453 (1996). doi:10.1111/j.1559-1816.1996.tb00079.x
16. Slovic, P.: Trust, emotion, sex, politics, and science: surveying the risk assessment battlefield. Risk Anal. **19**(4), 689–701 (1999). doi:10.1111/j.1539-6924.1999.tb00439.x
17. Finucane, M., Alhakami, A., Slovic, P., Johnson, S.: The affect heuristic in judgments of risks and benefits. J. Behav. Decis. Making **13**(1), 1–17 (2000)
18. Keinan, G., Friedland, N., Ben-Porath, Y.: Decision making under stress: scanning of alternatives under physical threat. Acta Psychol. **64**(3), 219–228 (1987). doi:10.1016/0001-6918(87)90008-4

19. Stankovic, A., Aitken, M.R., Clark, L.: An eye-tracking study of information sampling and decision-making under stress implications for alarms in aviation emergencies. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting 58, pp. 125–129 (2014)
20. Easterbrook, J.A.: The effect of emotion on cue utilization and the organization of behavior. Psychol. Rev. **66**(3), 183–201 (1959). doi:10.1037/h0047707
21. Yeh, M., Wickens, C.D.: Display signaling in augmented reality: effects of cue reliability and image realism on attention allocation and trust calibration. Hum. Factors **43**(3), 355–365 (2001). doi:10.1518/001872001775898269
22. Briggs, G.F., Hole, G.J., Land, M.F.: Emotionally involving telephone conversations lead to driver error and visual tunneling. Transp. Res. Part F Traffic Psychol. Behav. **14**(4), 313–323 (2011). doi:10.1016/j.trf.2011.02.004
23. Horrey, W.J., Wickens, C.D., Consalus, K.P.: Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. J. Exp. Psychol. **12**(2), 67–78 (2006). doi:10.1037/1076-898X.12.2.67
24. Dixon, B.J., Daly, M.J., Chan, H., Vescan, A.D., Witterick, I.J., Irish, J.C.: Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. Surg. Endosc. **27**(2), 454–461 (2013). doi:10.1007/s00464-012-2457-3
25. Mack, A., Rock, I.: Inattentional Blindness. MIT Press, Cambridge, Massachusetts (1998)
26. Simons, D.J., Chabris, C.F.: Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception **28**(9), 1059–1074 (1999). doi:10.1068/p281059
27. Vilar, E., Rebelo, F., Noriega, P., Teles, J., Mayhorn, C.: Signage versus environmental affordances: is the explicit information strong enough to guide human behavior during a way finding task? Hum. Factors Ergon. Manufact. Serv. Indus. **25**(4), 439–452 (2015). doi:10.1002/hfm.20557
28. Yerkes, R.M., Dodson, J.: The relation of strength of stimulus to rapidity of habit-formation. J. Comp. Neurol. Psychol. **18**(5), 459–482 (1908). doi:10.1002/cne.920180503
29. Wickens, C.D., Hollands, J.G., Banbury, S., Parasuraman, R.: Engineering Psychology and Human Performance. Pearson Education, Upper Saddle River (2013)
30. Kwee-Meier, S., Mertens, A., Schlick, C.: Age-related differences in decision-making for digital escape route signage under strenuous emergency conditions of tilted passenger ships. Appl. Ergon. **59**(A), 264–273 (2017). doi:10.1016/j.apergo.2016.09.001
31. Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., Polich, J., Reinvang, I., Petten, C.V.: Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin. Neurophysiol. **120**(11), 1883–1908 (2009). doi:10.1016/j.clinph.2009.07.045
32. ISO 7010:2011(E) Graphical symbols – Safety colours and safety signs – Registered safety signs
33. Thompson, P., Beath, T., Bell, J., Jacobson, G., Phair, T., Salbach, N.M., Wright, F.V.: Test-retest reliability of the 10-metre fast walk test and 6-minute walk test in ambulatory school-aged children with cerebral palsy. Dev. Med. Child Neurol. **50**(5), 370–376 (2008). doi:10.1111/j.1469-8749.2008.02048.x
34. Bohannon, R.W.: Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. Age Ageing **26**(1), 15–19 (1997). doi:10.1093/ageing/26.1.15
35. International Maritime Organization (IMO): Guidelines for a evacuation analysis for new and existing passenger ships. MSC.1/Circ.1238, 30th October 2007, London (2007)
36. Bohannon, R.W., Andrews, A.W., Thomas, M.W.: Walking speed: reference values and correlates for older adults. J. Orthop. Sports Phys. Ther. **24**(2), 86–90 (1996). doi:10.2519/jospt.1996.24.2.86

37. Watson, M.J.: Refining the Ten-metre walking test for use with neurologically impaired people. Physiotherapy **88**(7), 386–397 (2002). doi:10.1016/S0031-9406(05)61264-3
38. Field, A.: Discovering Statistics Using IBM SPSS Statistics. Sage, Los Angeles, London, New Delhi, Singapur und Washington, D.C (2009)
39. Porcelli, A.J., Delgado, M.R.: Acute stress modulates risk taking in financial decision making. Psychol. Sci. **20**(3), 278–283 (2009). doi:10.1111/j.1467-9280.2009.02288.x

Factors Research on EEG Signal Analysis of the Willingness of Error Reporting

Hongxia Li(✉) and Nan Zhou

Xi'an University of Science and Technology, Xi'an, People's Republic of China
406144519@qq.com, zhounanpaopao@gmail.com

Abstract. Error events and near-miss events often appear in our daily life and are featured by strong concealment, which may become a hot-bed of major accidents. From the perspective of error control, enterprises need to integrate relevant information to deal with those events so as to reduce their negative effects to the enterprise. The source of error information is error report. In order to find the measures of improving the willingness for report errors, mixed experiments designed for both within the subject and in between two subjects have been applied, together with related potentials (ERPs), to explore the effect of different motivation mechanism on the mechanism of self error report. The research shows that as far as the willingness of error report is concerned, the positive monetary incentive (bonus) often leads to greater FRN and P300 amplitudes, compared to the negative stimulus (penalty), whereas positive motivation is more effective than negative motivation when provoking extrinsic motivation. Meanwhile, when increasing the self-selection right of error report, i.e., when voluntary error report system is applied, the inner motivation of error report will increase accordingly and more attention will be paid to the incentive. Hence the effect of incentive will be improved.

Keywords: Error reporting intention · Event related potentials · Extrinsic motivation · Intrinsic motivation

1 Introduction

With the rapid economic development, the errors problem of also exist in the enterprise. From the beginning, the enterprise in order to eliminate the loss caused by the error, often used to deal with the error to avoid: that is, when the error event signs, from the source to kill all the possibility of errors occur, thereby reducing the occurrence of errors. However, due to various reasons, the error event is always inevitable. There are some limitations to error avoidance, such as the inability to eliminate or predict all errors, and the invariance of rules. In fact, the error is not only for the enterprise losses. According to modern error management theory, when the error event occurs, the error analysis and analysis to get relevant information, and organizational learning, sharing, can reduce the probability of the same error event recurrence; and the relevant error information to learn, not only Help to open the train of thought, and in the future work will stimulate his desire for innovation [2].

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 141–154, 2017.
DOI: 10.1007/978-3-319-58472-0_12

Therefore, companies need to establish some incentive mechanism to improve the staff to report the proportion of errors. From the external motivation, reward and punishment are powerful motivating factors, pre-researchers to conceal the incentive mechanism used in the more concentrated in the punishment of this level, that is, the person responsible for the incident and related events to conceal. False positives, late reports, false negative behavior punishment, increase the cost of such acts, and thus encourage individual initiative to report the error [3], but the effectiveness of this approach still need further empirical, experimental test. There are some organizations trying to volunteer to report the event of the incident exemption incentives. Even some studies suggest that reward is more effective than punishment [4].

In 2014, Shui-cheng Tian et al. and other research on the coal mine risk trillion events, risk management and risk trillion events reported voluntarily willing to report the relevant content of the study, which is the earlier report on the field of coal mine will study [5–8]. in the coal mine risk trillion incident voluntarily report the factors affecting factors, “reporting the consequences” of the dimensional indicators, the fear of punishment is the impact of miners reported dangerous events of the enthusiasm of the important factors, so the risk of trillion incident Retribution to promote “active reporting, non-punitive” principle, and the establishment of reporting reward system [9].

From the motivation of error reporting to individuals, the incentive mechanism used by the researchers from the previous period mostly focused on the punishment level. Recently, the researchers advocated rewarding the error reporting behavior, and believed that the reward and punishment of the incentive factors are effectiveness, and even more effective way to reward. However, most of these studies are based on the organization to study the willingness to report the error, from the level of individual neural mechanisms of error self-report motivation intensity is relatively small; from the research method, But also rarely from people’s physiological and psychological mechanisms into the study to prove that incentive incentives than punishment more effective, and given experimental data to confirm. Therefore, use the neural management to research on the individual level of error report willingness. By studying the mechanism of cognitive activity in the brain, it is more accurate to record the EEG data generated by the externally applied influence, quantitatively analyze the data, and obtain the conclusion from the measured data, which is relatively objective [10, 11]. Thus, the FRN component and the P300 component were selected for the study.

In the process of studying on the FRN components, the FRN component is also induced by the fact that the feedback is not the same as expected. This shows that, even if the feedback is rewarded, but if they do not meet the psychological expectations of the subjects, will still send a clear FRN components. In the research of 2007, Oliveira et al. [12] designed an experimental task to verify the rationality of the expected violation of the hypothesis. The experiment is divided into two experiments, the experimental materials are required to respond to a constant movement of the light point, the difference between the two experiments is that after the completion of the experiment in the subjects need to respond to their own just after the results of the behavior do an estimate, and then give the results of feedback procedures, and experimental two direct skip their assessment of the stage, the program directly showing the feedback results. According to the experimental results, it can be seen that FRN is induced only when the predicted results do not match the feedback results, regardless of the task completion results. Previous studies have confirmed that FRN components can reflect the degree of satisfaction with the results and the size of the expected gap, and are not sensitive to the true response. The FRN component can therefore be used as a measure of the intensity of motivations induced by different stimuli.

In the course of studying the P300 component, many experimental results can show that the P300 component is associated with the attention resource allocation. In 1986, studies by Donchin et al. [13] have shown that the P300 component is related to the allocation of attention resources. The study shows that the P300 component and attention to the relationship between the allocation of resources, designed a dual-task model of the experimental program, and many of the results in the subsequent proof of P300 component amplitude and Attention assigned to the size of. The study of Leng and Zhou's in 2010, it was also a pattern of simultaneous multi-tasking: a pair of friends and a stranger were simultaneously tested as gamblers, and one of the friends used to take the EEG caps for experiments, The subject was told before the experiment in addition to the need to complete their own tasks, but also need to pay attention to their friends there is another stranger in the performance of the task. The results are discussed, giving the view that: subjects and their friends a total of emotional stronger, so give friends more attention [15]. Therefore, it also shows that the P300 component can well reflect the individual's attention resource allocation. The same studies of Itagaki and Katayama [17] and Fukushima [16] and Hiraki (2009) also demonstrate this cognitive phenomenon associated with the allocation of attention resources.

In this paper, event-related potential techniques in neuroscience are used to analyze the factors that affect the motivation of individual error reporting. The purpose of this study is to find out whether the incentive of reward is more effective than punishment and whether the right of self-determination affects the intrinsic motivation of error reporting. Stronger, and provide physical evidence of the data.

2 Method

2.1 Experimental Design Framework

This research use three different stimulus methods were selected: positive, negative, and non-stimulating. Individuals were rewarded, punished or unresponsive for random error reporting. The experiment was divided into two groups. Group: experimental group and control group, the experimental group for the voluntary reporting group, the control group for the mandatory reporting group, the individual's self-determination as independent variables, self-determined behavior of error reporting behavior of intrinsic motivation.

2.2 Experimental Subjects

In this study, 20 postgraduates were recruited from Xi'an University of Science and Technology as subjects aged 20–27 years (mean age 24.9 years, mean SD 1.86). In order to reduce the physiological factors caused by the brain on the EEG, the right eye;

Data accuracy. So the subjects cannot stay up late, cannot drink high-caffeine beverages.

2.3 Process of Experiment

This research program is realized through E-Basic language and Script language programming, the language is written in E-Prime2.0 in the Inline object. Experiments using the software implementation program (E-Run) presentation of stimulus material, while recording the relevant behavior data.

The experimental procedure is divided into two blocks. The first Block, which is the first stage of the experiment showed 200 trials, the subjects cannot be free to end the experiment; 200 trial will pop up after an interface to remind the subjects can take a break, and then free to choose willing to continue the experiment. If the subjects are going to the second phase. The participants can finish the experiment at any time (the maximum number of trials in the second stage is 200) (Fig. 1).

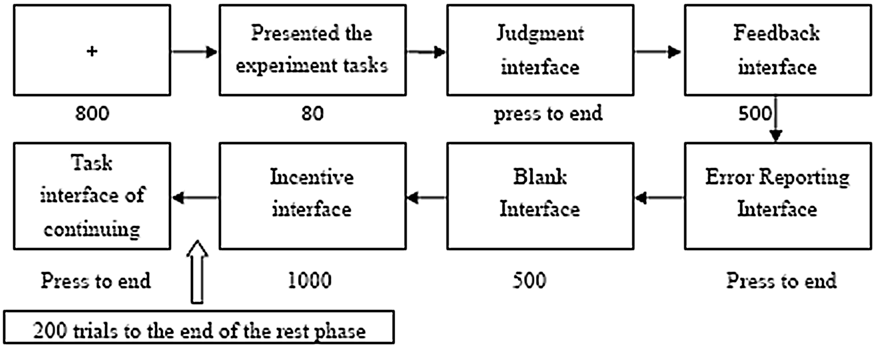


Fig. 1. The experimental stimulus flow chart (unit: ms)

The Neuron EEG/ERP event-related potential system was used to collect EEG signals.

3 Analysis of Willingness to Influence on Experimental Data

According to statistical principles: for the behavioral data, the t test between groups of experimental data analysis. For EEG data, the three groups of experiments in the use of stimulation methods and electrode points ANOVA analysis of variance analysis of data processing. The variance analysis method was used to analyze the data, and t-test was used to analyze the difference amplitude data.

3.1 Behavioral Data Analysis

The number of completed experiments in different experimental groups in the second stage of the experiment (the subjects could finish the experiment voluntarily) was statistically analyzed. The results showed that 48.60 experiments were carried out in the experimental group, and the variance was 15.01. In the control group, 20.40 experiments were conducted on average, and the variance was 11.02.

3.2 EEG Data Analysis

The Analysis of FRN Components. In the forehead and central regions of the brain, the FRN component appears in the ACC region, and the amplitude of the FRN component in this region is relatively large. Therefore, we selected the ACC region of the Fl, FZ, F2, FCl, FC2, C1, CZ and C2 a total of eight electrode points for analysis. According to the analysis of EEG data, the average amplitude of 200–250 ms is selected as the amplitude of the FRN component, and the data are analyzed statistically.

(8 feedback points: F1, FZ, F2, FC1, FC2, C1, and C2) for the average amplitudes of the FRN components (between 200 and 250 ms) 3 (3 feedbacks: bonus, penalty and no response) CZ and C2) × 2 (two experimental groups: experimental group and control group) to do mixed measurement of variance analysis grouping factors were two groups of experimental subjects (experimental group and control group), repeated measurements of the level of the electrode (9) And feedback (three). The results show that the amplitude of the FRN component has a significant main effect on the three feedbacks: $F(1, 19) = 3.54$, $p < 0.05$; no major effect is found between the experimental groups and the electrodes.

It was found that the FRN amplitudes ($M = 5.51$, $SD = 0.52$) of the positive excitation in the experimental group were significantly larger than those of the negative excitation ($M = 3.22$, $SD = 0.67$), $p = 0.000 < 0.05$; The FRN amplitudes of unstimulated FRN amplitudes were significantly larger than those of unstimulated FRN amplitudes ($M = 4.47$, $SD = 0.95$), $p = 0.04 < 0.05$. ($M = 5.44$, $SD = 0.49$) was significantly greater than the FRN amplitude of negative excitation ($M = 3.10$, $SD = 0.58$), $p = 0.000 < 0.05$ in the control group, FRN of positive excitation $P = 0.024 < 0.05$. The FRN amplitude of unstimulated FRN amplitude is significantly larger than that of negative excitation, $p = 0.023 < 0.05$, and the amplitude of unstimulated FRN is larger than that of unstirred FRN ($M = 4.32$, $SD = 1.23$).

The FN, FC1, and CZ, which are the most prominent of the eight FRN components selected for the study, are plotted as shown in Figs. 2 and 3. Among them, FRN components induced by different excitation conditions are boxed out.

The Analysis of the P300 Composition. From the literature, we found that the P300 component in the posterior region of the brain was more pronounced and the amplitudes were greater. Therefore, we selected nine central electrodes (C1, CZ, C2, CP1, CPZ, CP2, P1, PZ, and P2). The time domain of the P300 component was selected, and the waveforms of the P300 were selected and observed and analyzed. The time window was 300–350 ms, and the voltage value of the P300 component in this time period was

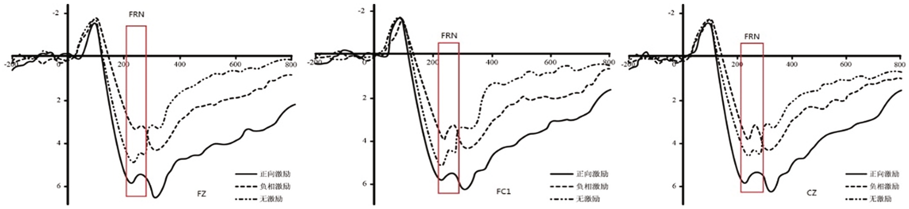


Fig. 2. FRN waveforms of the experimental group FZ, FC1 and C points

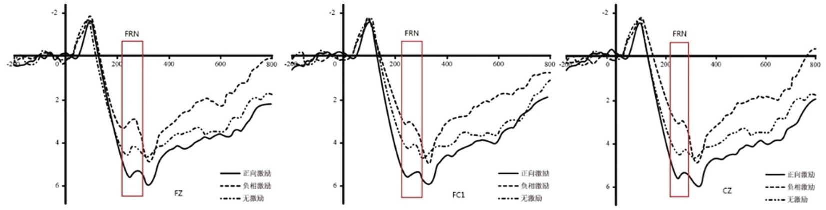


Fig. 3. FRN waveforms of FZ, FC1 and C points in the control group

superimposed and averaged to obtain the component The average amplitude value, and use this data for statistical analysis.

In the same way, the average amplitude of the P300 component in the selected time window is 3 (3 feedbacks: bonus, penalty, and no response) × 8 (8 electrode points: C1, CZ, C2, CP1, CPZ, CP2, P1, PZ and P2) × 2 (two experimental groups: experimental group and control group) to do mixed measurement of variance analysis. Similar to the FRN component analysis model, the grouping factors were two groups of subjects (with and without option), and the repeated group level was electrode (nine) and feedback (three). The results showed that the amplitude of P300 had a significant main effect on the three feedbacks: F (1, 19) = 2.88, $p < 0.05$; there was no main effect at the electrode point; there was interaction between the feedback type and the experimental group, 19) = 3.39, $p < 0.05$.

In order to further analyze the relationship of the three feedback types on the 300 components, paired T-test analysis of the three feedback results was done. The control group experimental group, the results of three kinds of feedback and electrode pairs does T test analysis.

The results showed that the amplitudes of P300 components induced by positive excitation (M = 6.58, SD = 0.37) were significantly larger than those of P300 (M = 4.11, SD = 0.35), $p = 0.000 < 0.05$; ($P = 0.000 < 0.05$). The P300 amplitude of the unstimulated P300 amplitude is significantly smaller than the P300 amplitude of the negative excitation ($p = 0.001 < 0.05$). In the control group, the P300 amplitude of positive excitation (M = 5.97, SD = 0.87) was significantly greater than that of negative excitation (P = 0.001, P = 0.001; ($P = 0.007 < 0.05$). There was no significant difference between the amplitude of P300 and the amplitude of negative excitation ($p = 0.28 > 0.05$), and the amplitude of P300 was not significantly different from that of no excitation (M = 4.38, SD = 0.69).

The P300 component of the nine electrode points selected for the study was plotted with three distinct CZ, CPZ, and PZ electrode points, as shown in Figs. 4 and 5, where the author box out the P300 components induced by different excitation conditions.

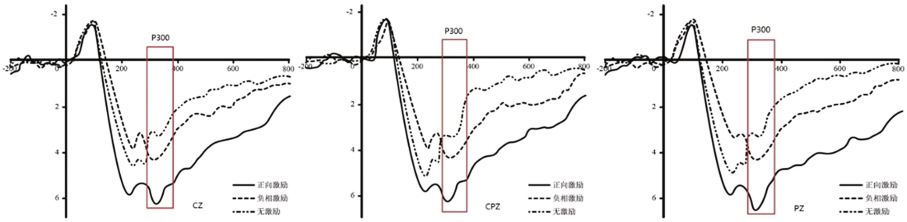


Fig. 4. P300 waveforms of the CZ, CPZ and P points in the experimental group

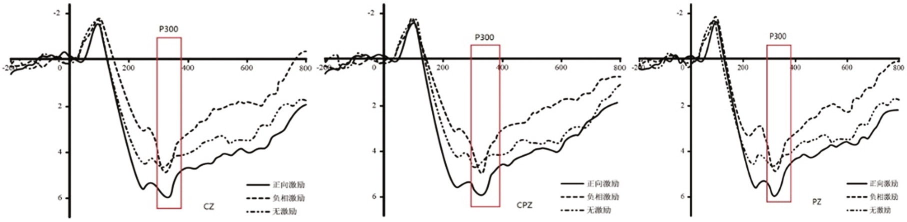


Fig. 5. The P300 waveform of the CZ, CPZ and P points in the control group

There is an interaction effect between the feedback type and the experimental group. Therefore, the p300 amplitudes of positive stimulus feedback P300 of the experimental group were significantly larger than that of the positive stimulus feedback P300 ($p = 0.04 < 0.05$), but for the experimental group The P300 amplitude of the negative stimulus feedback and P300 amplitude of the non-excitation feedback were significantly smaller than those of the negative stimulus feedback P300 ($p = 0.01 < 0.05$) and P300 amplitude ($p = 0.017 < 0.05$). Similarly, the experimental group in the three feedback excitation amplitude difference is greater than the control group of three feedback excitation amplitude difference.

4 Discussion

4.1 Behavioral Data Results Discussion

The experimental results showed that the number of tasks completed in the second stage of the experimental group was significantly greater than the number of tasks completed in the control group. In the experiment group, the intrinsic motivation intensity was greater than the mandatory report the intrinsic motivation of a task. It can also be argued that giving the subject a high degree of self-determination, for example, giving the participant the option of reporting his own error will produce a stronger task intrinsic motivation than mandatory error reporting. Thus, the report shows that voluntary reporting mechanisms are more effective and more developmental than mandatory reporting.

4.2 EEG Data to Discuss the Results

Conclusions Related to FRN Composition. FRN amplitude has significant effect on three kinds of feedback results, but there is no main effect between different experimental groups and between electrode points. In experiment group and control group, the mean value of negative excitation results is higher than that of positive excitation and the result of non-excitation feedback is small. The FRN component is a negative trend-specific wave in the feedback phase of the results, so this paper discusses the motivation and cognitive evaluation. A large number of studies have shown that FRN components are more active in the anterior cingulated gurus (ACC) region, whereas ACC detects whether the resulting feedback is inconsistent with expectations, so the FRN component is also induced by feedback and expectation. In a 2007 study, Oliveira et al. designed an experimental task to verify the rationality of the expected violation of the hypothesis. According to the experimental results, it can be seen that the FRN components are induced only when the predicted results are not consistent with the given feedback results, even if the feedback results are positive feedback, so the FRN component reflects. The result of the feedback is different from the expected procedure [-]. For the theory of emotional motivation, Gehrig et al. showed that the amplitude of the FRN component induced by the feedback at the time of losing money was significantly larger than that of the winning FRN component. Researchers to discuss the conclusion, FRN component is actually the performance of the brain on different stimuli reflect different motives and emotions, or measure the intensity of motivation indicators. In contrast to negative feedback, positive feedback will induce a smaller FRN component amplitude, in fact, also proved that FRN components subject to the same subjective (such as Zhou, Z., etc.), relative to negative feedback, Emotional impact.

Thus, based on the experimental results, “the negative-excitation feedback results in a larger FRN-component amplitude than in the forward and no-excitation feedback results.” It can be seen that the subjects were more willing than the negative stimulus (penalty) (Reward), but in the feedback phase there is not the same as he expected feedback, there will be more significant negative wave waveform. The positive incentive (reward) is higher than the negative subjective incentive (punishment) and no incentive. It also shows that FRN can stably reflect the subjective evaluation of feedback results.

Beside,there was no significant difference in FRN amplitude between the two groups. However, in both experiments, the FRN amplitudes induced by the negative excitation (penalty) were significantly larger than those without excitation and positive excitation (reward).

Conclusions Related to P300 Composition. The method of analyzing the composition of P300 is the same as that of FRN. It is concluded that the amplitude of the P300 component has a significant main effect on the three kinds of feedback results, and the feedback type has an interaction effect with the experimental group. The mean P300 voltage of positive excitation is larger than that of negative excitation and no excitation, which indicates that the feedback result of positive excitation results in a larger P300 amplitude than non-excitation feedback and negative feedback. The result of the pairing test shows that the relationship between the amplitude and the amplitude is positive excitation > negative excitation > no excitation, both in experiment group and in control group. But the difference is that the P300 amplitude of the unstimulated P300 is significantly less than the P300 amplitude of the negative stimulus in the experimental group, while there is no significant difference between the P300 amplitude and the negative excitation in the control group.

P300 components can reflect the cognitive function of the brain, attention to how much the allocation of resources. The data show different degrees of personal cognition [22–24]. In the literature review, it is also suggested that many of the experimental results can reflect the evidence that the P300 component is associated with the attention resource allocation. Studies by Donchin et al. [13] have shown that the P300 component is related to the allocation of attention resources (during cognitive processing of stimuli). In this study, Donchin and other researchers in order to explore the P300 component and attention to the relationship between the allocation of resources, and after a lot of research results have proved that the P300 component amplitude and distribution of attention to the size of the results show that The more attention is paid to the current stimulus, the greater the amplitude of the P300 [25–27]. There is also an interpretation of the P300, that is, it is related to the effect of forward, is the target desire to achieve the target [28, 29].

In the present research results, the feedback of the positive stimulus (reward) induced a greater P300 amplitude than the negative (punish) and non-stimulated, and the P300 amplitude of the negative stimulus (penalty) was greater than the P300 amplitude. In contrast to the non-incentive, the negative incentive (punishment) is also effective, but the incentive effect is not as good as the incentive to be a positive incentive (reward); Positive incentive (reward). Because of the organization of the employee error reporting behavior to reward or punishment which are the behavior and report the error message of a way of attention.

4.3 Behavioral Data the EEG Data Structure Joint Analysis

In the experimental group, the P300 amplitude of no excitation was significantly smaller than that of the negative excitation, while the P300 amplitude of no excitation and the amplitude of negative excitation did not differ significantly in the control group. According to the results of behavioral data analysis, we can see that self-determination of the task to bring a stronger intrinsic motivation. High incentive reward motivation in 2011 is found by Murayama. K. and stimulate the individual more intense sense of happiness. Moreover, the magnitude of the amplitude of the P300 component is related to how much attention is drawn to the feedback result and the amplitude of the P300 increases as the excitation mode becomes more interesting. Therefore, it can be concluded that the participant will have more attention to the experiment with the option, and less attention will be paid to the control group experiment with less attention. So it can be said that the intrinsic motivation to improve, can enhance the external incentives to the impact of subjects.

5 Discussion on Incentive Countermeasures Based on Error Report

Positive incentives (incentives) are more effective than negative incentives (punishments) in motivating individuals to report error motivations for reporting errors. This is also the recent researchers to promote the reporting behavior of the error reward, that reward and punishment are a powerful incentive mechanism. This study provides a mechanism for rewarding more efficient physiological experimental data. The inherent motivation for error reporting is to increase the individual's self - determination power for error reporting and to increase the individual's intrinsic motivation for error reporting.

Therefore, this paper proposes to make some management recommendations as follow:

5.1 Build Voluntary Errors Reporting System

Improve the error reporting incentive mechanism can be strengthened from two aspects of the staff will report the error. In increasing the motivation of employee error reporting, it is more effective to reward the error reporting behavior than punish the error concealment. Separate incentives for employees reporting errors can lead to more employees joining the ranks of reporting errors. The organization's reporting of errors to employees, whether rewarding or penalizing employees for errors, is an acknowledgment of error messages reported by employees, and this emphasis can also lead to employee risk of error But it is more effective to reward error reporting behavior, because the organization's attitude towards error reporting can lead to errors in the employee's sense of error. More attention to the occurrence of the error itself, rather than the first thought of this error will not bring me losses, to give employees a relatively open environment, can make everyone more willing to exchange error information; Reward can also increase the individual's Expected benefits, when employees believe that reporting errors have substantial benefits, tend to report the error [30], this change by changing the behavior of the employee's error reporting behavior is feasible.

According to the results of this paper, we can put forward some points needing attention:

1. Voluntary principles of error reporting.

According to the research conclusion, giving employees the right to self-determination of error incident reports can help to improve the intrinsic motivation of employee's error reporting behavior. Error reporting system is also designed to enable enterprises to get more comprehensive and accurate error information, voluntary reporting model will enable enterprises to collect less of the useless information, greatly reducing the error information processing efficiency. But even so, companies cannot guarantee timely and effective collection of all the information, after all, reporting behavior is voluntary, you can report or not report. Therefore, another incentive mechanism is proposed.

2. Embedded in a reasonable incentive mechanism into the error reporting system.

According to the results of the study, when the individual in the report of the error, the more hope to see the incentive incentives, so employees can complete an error reporting task after its reward. But this reward is the need to control; otherwise it will only violate our intention to set up the error reporting system. Therefore, we can use the error reporting integral system to determine the extent of reward or punishment.

For example, when you report your errors in the Voluntary Reporting System, your reports are scored against your report's importance, validity, completeness, and operability, Points, the more points, the more bonuses; the same time, the system will make mistakes based on the size of the impact of your business to deduct the response points, but you can submit information to contribute to offset the knowledge, which is equivalent to Your information is valuable, then you add the score will be greater than the deduction to the score. In addition, the error event duplication of information, if there is no new educational significance, no extra points.

3. The availability of error reporting system.

When the voluntary principles and incentives are completed, the error reporting system must be established to consider its availability. Too complex, difficult to understand the system will reduce the enthusiasm of the staff error report: Imagine, when we finally determined to upload their experience, they encountered a system of how the uncertain, only our excitement poured Off, and then later do not want to use this system, because the reporting behavior is voluntary. At the same time, building an error reporting system requires enterprises to have the infrastructure with the performance matching to reduce the probability of system problems. Therefore, improving the availability of error reporting systems is also a way to keep employees reporting errors.

5.2 Interpreters Establish a Positive Error Management Atmosphere

A positive error management atmosphere is "the freedom of the organization to discuss errors, share knowledge of the error, but also in the face of error problems when dealing with and solve" [31]. Van Dyck C believes that the error management culture (EMC), like corporate safety culture, is also an important component of organizational culture, and the establishment of error management culture encourages employees to discuss their own mistakes without the burden of experience, All have a positive impact on the establishment of organizational error reporting systems [32].

Gold et al. (2014) also found that an open error management environment facilitates error reporting [33]. On the domestic front, Qunyang Xie et al. (2015) argue that a positive culture of error management will enhance employee willingness to report errors [32].

From a practical point of view, organizations often influence the behavior of employees through culture. At present, the transmission of error messages relies more on the degree of freedom that the organization provides to employees. Positive errors in the atmosphere, the higher the degree of freedom of employees for error reporting behavior of the intrinsic motivation is also higher. Creating a positive error management environment can be done in the following ways:

1. To build up establishment of easy discussion of errors and experience of the business environment.

Obstacles to employee error reporting behavior of the reasons, more because of their own psychological barriers, such as fear of being detained because of mistakes wages, fear of losing face is not easy to open, which is now the enterprise rules and regulations. Now the enterprise to establish a positive atmosphere of error management less, the rules and regulations is relatively harsh, so that individual employees are more willing to share their own error messages. But if everyone is hiding, then the same mistakes will continue to occur, so companies need to allow employees the freedom to explore the communication of an error event in the event of the problem, timely communication with others or errors Information can be reported, then the error event can be processed in time to prevent its continued deterioration; the same time can also reduce the recurrence of the same error event probability, in order to reduce the occurrence of security incidents. At this time a good organizational atmosphere will help improve the staff's willingness to report the error.

2. To carry out error culture training lectures.

Through the training of error culture, it can guide the employee's attribution of errors. This way not only to employees to express the importance of error events, as well as the degree of tolerance, but also to promote employees from a variety of aspects to study the causes of errors from the staring at the consequences of errors can be studied more Methods to solve this problem; on the other hand, it can also promote the innovation behavior of enterprise employees.

References

1. Frese, M., Keith, N.: Action errors, error management, and learning in organizations. Annu. Rev. Psychol. **66**, 661–687 (2015)
2. Zhao, B.: Learning from errors: the role of context, emotion, and personality. J. Organ. Behav. **32**(3), 435–463 (2011)
3. Li, X., Yang, J., Xie, Y.: Effects of the willingness to report employees' mistakes - a study on the motivation, behavioral value and organizational climate. Cent. Univ. Finan. Econ. **01**, 99–106 (2016)
4. Greitemeyer, T., Weiner, B.: The asymmetrical consequences of reward and punishment on attributional judgments. Pers. Soc. Psychol. Bull. **29**, 1371–1382 (2003)
5. Tian, S., Qian, X., Li, H., et al.: Study on safety management incentive mechanism and safety de-incentive factors. J. Xi'an Univ. Sci. Technol. **22**(1), 15–17 (2002)

6. Yuan, X., Li, H., Tian, S.: Application of catastrophe theory in early warning of industrial accident. J. Xi'an Univ. Sci. Technol. **31**(4), 482–488 (2011)
7. Li, L., Tan, S., Deng, J., et al.: Analysis of influencing factors of unsafe behavior of miners and control countermeasures. Xi'an Univ. Sci. Technol. **31**(6), 794–798 (2011)
8. Yu, G.: Study on management of dangerous events based on three kinds of hazard sources. Xi'an University of Science and Technology (2013)
9. Tian, S., Sun, S., Yu, G., et al.: Study on factors affecting willingness to voluntary willingness of coal mines. Xi'an Univ. Sci. Technol. **34**(5), 517–522 (2014)
10. Ma, Q., Fu, H., Bian, J.: Neural industrial engineering: a new stage of industrial engineering development. Manag. World **6**, 015 (2012)
11. Ma, Q., Wang, X.: From neuroeconomics and neuromarketing to neural management. Chin. J. Manag. **20**(3), 129–132 (2006)
12. Oliveira, F.T.P., McDonald, J.J., Goodman, D.F.M.: Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. J. Cogn. Neurosci. **19**(12), 1994–2004 (2007)
13. Donchin, E., Kramer, A., Wickens, C.: Applications of event-related brain potentials to problems in engineering psychology. In: Psychophysiology: Systems, Processes, and Applications. Guilford Press, New York (1986)
14. Kok, A.: On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology **38**(3), 557–577 (2001)
15. Leng, Y., Zhou, X.: Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. Neuropsychologia **48**(2), 448–455 (2010)
16. Fukushima, H., Hiraki, K.: Perceiving an opponent's loss: gender-related differences in the medial-frontal negativity. Soc. Cogn. Affect. Neurosci. **1**(2), 149–157 (2006)
17. Itagaki, S., Katayama, J.: Self-relevant criteria determine the evaluation of outcomes induced by others. NeuroReport **19**(3), 383–387 (2008)
18. Holroyd, C.B., Hajcak, G., Larsen, J.T.: The good, the bad and the neutral: electrophysiological responses to feedback stimuli. Brain Res. **1105**(1), 93–101 (2006)
19. Jessup, R.K., Busemeyer, J.R., Brown, J.W.: Error effects in anterior cingulate cortex reverse when error likelihood is high. J. Neurosci. **30**(9), 3467–3472 (2010)
20. Alexander, W.H., Brown, J.W.: Medial prefrontal cortex as an action-outcome predictor. Nat. Neurosci. **14**(10), 1338–1344 (2011)
21. Zhou, Z., Yu, R., Zhou, X.: To do or not to do? Action enlarges the FRN and P300 effects in outcome evaluation. Neuropsychologia **48**(12), 3606–3613 (2010)
22. Hansenne, M.: The p300 cognitive event-related potential. I. Theoretical and psychobiologic perspectives. Neurophysiologie clinique = Clin. Neurophysiol. **30**(4), 191–210 (2000)
23. McDowell, K., Kerick, S.E., Santa Maria, D.L., et al.: Aging, physical activity, and cognitive processing: an examination of P300. Neurobiol. Aging **24**(4), 597–606 (2003)
24. Polich, J.: Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. **118** (10), 2128–2148 (2007)
25. Furdea, A., Halder, S., Krusienski, D.J., et al.: An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology **46**(3), 617–625 (2009)
26. Gray, H.M., Ambady, N., Lowenthal, W.T., et al.: P300 as an index of attention to self-relevant stimuli. J. Exper. Soc. Psychol. **40**(2), 216–224 (2004)
27. Watter, S., Geffen, G.M., Geffen, L.B.: The n-back as a dual-task: P300 morphology under divided attention. Psychophysiology **38**(06), 998–1003 (2001)
28. Wu, Y., Zhou, X.: The P300 and reward valence, magnitude, and expectancy in outcome evaluation. Brain Res. **1286**, 114–122 (2009)
29. Ferdinand, N.K., Kray, J.: Age-related changes in processing positive and negative feedback: is there a positivity effect for older adults? Biol. Psychol. **94**(2), 235–241 (2013)

30. Zhao, B., Olivera, F.: Error reporting in organizations. Acad. Manag. Rev. **31**(4), 1012–1030 (2006)
31. Van Dyck, C., Frese, M., Baer, M., et al.: Organizational error management culture and its impact on performance: a two-study replication. J. Appl. Psychol. **90**(6), 1228 (2005)
32. Dyck, C.: Putting errors to good use: error management culture in organizations. In: KLI (2000)
33. Gold, A., Gronewold, U., Salterio, S.E.: Error management in audit firms: error climate, type, and originator. Acc. Rev. **89**(1), 303–330 (2013)

Mentally Imagined Item Captures Attention During Visual Search

Haifeng Li(✉) and Xiaomei Li

Department of Psychology,
Fujian Normal University, Fuzhou, People's Republic of China
haifeng.li@fjnu.edu.cn

Abstract. This study aims to explore whether mental imagery can capture attention during visual search in a combined working memory and visual search task. The study compared a working memory condition that participants required to maintain a presented item then perform a visual search task with a mental imagery condition that participants mentally generated an item then searching. In both conditions, the item that they maintained can be a valid or invalid cue for the visual search. The result showed that, search reaction time was delayed by the invalid cue compared with the valid cue in the mental imagery condition, indicating mental imagery may share the same visual representation with attention. This result may have implications on imagery training in different ways which helps people detect malfunctions or dangerous targets more efficiently in the interactive environments.

Keywords: Mental imagery · Working memory · Visual attention

1 Introduction

Working memory serves to temporarily maintain and manipulation information for further use [1]. Recently, studies revealed that attention can be guided or biased by the contents in working memory [2–5], a phenomenon termed as memory-driven attentional capture. In these studies, participants were required to maintain an item while they performed a visual search task. Search reaction time was delayed by a working memory related distractor relative to a working memory unrelated distractor, or was accelerated by a working memory related target relative to a working memory unrelated target. The *biased competition model* of visual attention was used to explain why content in working memory captures attention [6]. It proposed that items maintained in working memory can pre-activate the representations of these items. When the visual search display appears, visual attention gives preference to those items that contain the same or similar representations in working memory, thus they can survive from the competition. Based on this, a great deal of psychological and neural studies has revealed that working memory and attention (or perception) activate a striking degree of overlap in common neural resources, indicating they may share the same representation [7–10].

If working memory can be regarded as a process of maintaining recently encountered stimuli, mental imagery can be seen as a process of mentally generating a

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 155–163, 2017.
DOI: 10.1007/978-3-319-58472-0_13

stimulus from short- or long-term memory. Mental imagery is the representation in a person's mind of the physical world but the world is not actually being perceived [11]. Visual imagery is accompanied by the experience of "seeing with the mind's eye". Likewise, many studies supported that mental imagery also activates the representations in the visual system. In a pilot study from Farah [12], participants were instructed to imagine either an *H* or a *T*. Then in a temporal two-interval forced-choice detection task, two successive observation intervals faintly presented one of these letters. Participants should report which interval contained the imagined letter. The letters were better detected when they matched the imagined letter, suggests a shared visual representation between mental imagery and attention (or perception). In a follow-up study, Farah instructed participants to mentally image an *H* or a *T* into a grid of empty squares [13]. After that, participants were asked to detect a probe dot that fell on or off the image. Results found that, relative to the dots falling off the image, dots falling on the image were better detected. Pashler and Shiu also showed the connection between mental imagery and visual representation in an attentional blink task [14]. Participants were asked to image a specified object (e.g., *tiger*), and then they should search for a target digit from a series of rapid sequential presented pictures. Digit detection was impaired when the imaged object was presented before the digit.

Mental imagery and visual attention also activate a great deal of overlapped brain regions [15, 16]. These common cortical regions involved in occipitoparietal and occipitotemporal visual association areas [17–19], the primary visual cortex [20–22] and the lateral geniculate nucleus [23]. Besides, studies found that, kinesthetic imagery can activate the primary motor cortex [24], tactile imagery can activate the somatic cortex [25]. In a meta-analytic study, McNorgan compared the activation between mental imagery and perception, and revealed that in most studies, mental imagery from different sensory channels can activate the corresponding primary sensory cortex [26]. But there were still some studies showed that the primary visual cortex was not activated in the imagery condition [27, 28], especially when imaging a moving object or scene [29, 30].

Based on the studies above, it is still unclear that whether mental imagery and attention share the same visual representation. In the present study, we examined whether a mentally imagined item can capture attention during visual search. We compared a working memory condition that participants maintained a shape then search for a tilted line target with a mental imagery condition that they generated a shape according to the instructions then searching. In the mental imagery condition, a big shape and a small shape were presented at the beginning of the trial. During a 2,000 ms blank, participants were required to mentally subtract the smaller shape from the bigger one, then maintained the new shape. After a 100 ms mask, a search display of four shapes was presented. One of the shapes was the same shape as the participants generated. There was a line in each shape. One was a tilted line, and others were vertical lines. Participants were asked to discriminate the orientation of a tilted line in one of the shapes; therefore, the shape could be valid or invalid cues for the discrimination task. At the end of the trial, a probe shape was presented to test their working memory. In the working memory condition, all the stimuli were identical as in the mental imagery condition, except that they just maintained the shape that was presented at the beginning of the trial. If the visual search is delayed by the mentally imagined item when it is an invalid cue for the search target, then we can infer that mental imagery also utilizes the similar visual representation as visual attention.

2 Method

2.1 Participants

Twenty undergraduate and graduate students were paid for their participation. Two were exclude due to a low memory accuracy (55% and 66% respectively). Therefore, a total of 18 participants (15 females; 17–24 years) were included in the final analyses. All participants were right-handed, had normal or corrected-to-normal visual and none reported color blindness.

2.2 Materials

The to-be-memorized or to-be-generated items were transformed from 5 original shapes: a hexagon (radius 1° visual angle), a parallelogram (length 2.5°, height 2° visual angle), a pentagon (radius 1° visual angle), a rhombus (length 2.2°, height 2.2° visual angle), and a square (length 2°, height 2° visual angle). All shapes appeared in gray (RGB: 85, 85, 85) and were presented on a black background. In the mental imagery condition, a shape and a part of the shape (a small triangle) were presented. Participants were required to mentally generate a shape via subtracting the smaller one from the bigger one. In the working memory condition, the generated shape was directly presented. The example stimuli were shown in Fig. 1. In the formal test, each original shape had 4 variations.

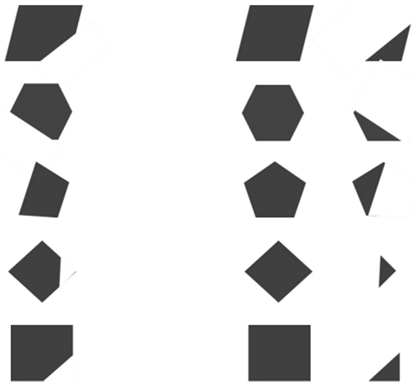


Fig. 1. Example stimuli for this study. Left and right shapes were presented at the beginning of the working memory condition and the mental imagery condition respectively.

In the search display, most of the shapes contained a black vertical line (length 0.8° visual angle, width 2 lb). One of the shapes contained a black tilted line which was rotated left or right 15° around the center of the vertical line.

2.3 Apparatus

The experiment was programmed using E-prime 2.0, and was run on a 17-in. LCD at a viewing distance of approximately 60 cm without a chin rest. The monitor was set to a 1024 × 768 resolution with a 85 Hz refresh rate and 32-bit colors.

2.4 Procedure

The procedure was illustrated in Fig. 2. The background of the experiment was black. At the beginning of each trial, in the working memory condition, a shape was presented in the screen's center for 2,000 ms. Participants were asked to remember this shape until the end of the trial. After the shape disappeared, a 2,000 fixation was inserted during which participant could consolidate the shape they maintained. In the mental imagery condition, an original shape and a part of this shape were presented. During the fixation presenting phase, participants were required to mentally subtract the smaller shape from the bigger one, then maintained the new shape they generated. Next, a 100 ms mask (length 14.82°, height 11.11° visual angle) was presented to prevent participants from refreshing the item.

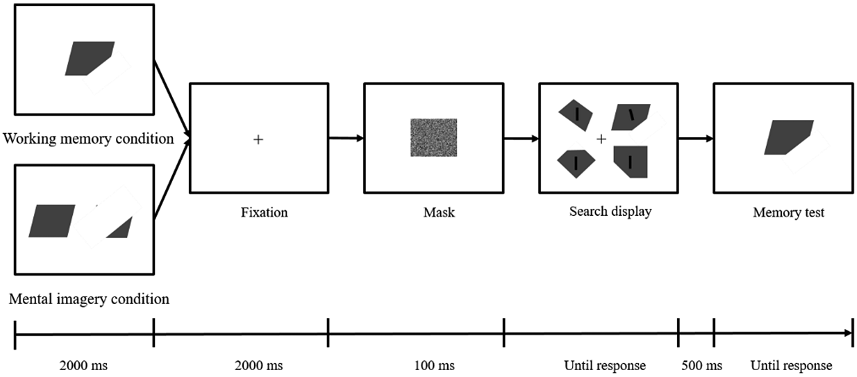


Fig. 2. Experimental procedure and example stimuli for this study.

The search display appeared after the mask disappeared. This display consisted of 4 shapes. They were placed on the vertices of an imaginary square (length 12°, height 12° visual angle), which was centered on the fixation. One of the shapes was the same shape as they maintained. Each shape contained a line in the shape center. Three of them were vertical lines, only one was a left-tilted or right tilted line. Participants were asked to find the tilted line and respond with the left arrow key or the right arrow key when the line was left-tilted or right-tilted. Therefore, the shape could be valid or invalid cues for the search task.

After another 500 ms blank screen, participants were instructed to match a probe shape that was presented in the screen's center. If the probe shape matched the shape they maintained, participants responded with the left arrow key; otherwise, press the right arrow key. The probe shape matched on 1/2 of the trials. In each trial, the probe shape and the maintained shape were always transformed from the same original shape.

2.5 Design

This was a 2 (Experiment type: working memory condition vs. mental imagery condition) × 2 (Cue type: valid vs. invalid) design experiment. There was a total of 160 trials with 40 trials per treatment combination. All the trials were randomly intermixed across the whole experiment. Each participant first received 16 practice trials, then they completed 4 blocks with 40 trials each, with a 1-min break between blocks. The experiment will last less than 30 min.

3 Results

The accuracy of the visual search task and the memory test were 99.55% and 92.95% respectively. A repeated measures ANOVA with experiment type and cue type as within factors revealed that, memory accuracy was higher in the working memory condition (94.14%) than in the mental imagery condition (91.72%), as confirmed by a main effect of experiment type, $F(1, 17) = 7.28$, $p = .015$, $\eta_p^2 = 0.300$. There were no main effect of cue type, $F(1, 17) = 0.001$, $p = .978$, and interaction between experiment type and cue type, $F(1, 17) = 0.019$, $p = .891$.

The result of search reaction time was illustrated in Fig. 3. A repeated measures ANOVA with experiment type and cue type as within factors revealed that, the main effect of experiment type was not significant, $F(1, 17) = 0.04$, $p = .842$. There was no reaction time difference between the mental imagery condition and the working memory condition. But the main effect of cue type was significant, $F(1, 17) = 7.85$, $p = .012$, $\eta_p^2 = 0.316$, indicating the valid cue condition had faster reaction time than the invalid cue condition. Importantly, there was a significant interaction between experiment type and cue type, $F(1, 17) = 5.28$, $p = .035$, $\eta_p^2 = 0.237$. Post hoc analysis using the LSD test revealed that the valid cue condition had faster reaction time than the invalid cue condition in the mental imagery experiment type ($p = .002$), but in the working memory experiment type, the valid and invalid cue condition had the similar reaction time ($p = .621$).

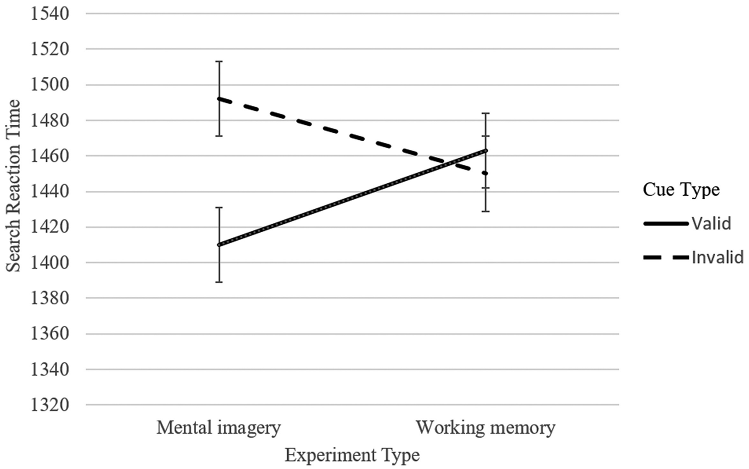


Fig. 3. Search reaction time as a function of experiment type and cue type. Error bars represent 95% within-participants confidence intervals with Masson and Loftus's method [31].

4 Discussion and Conclusion

This result revealed that a mentally imagined shape can capture attention during visual search, indicating that mental imagery and attention may also share the same visual representation. But surprisingly, there was no memory-driven attentional capture was found.

This result was consistent with the notion proposed by other researchers. Thomas assumed that there is only quantity difference, but no quality difference, between imagery and perception or attention [32]. Pearson, Naselaris, Holmes and Kosslyn proposed that, imagery is a weakening perception [33]. To some extent, imagery can replace perception and plays the similar role as perception. A series studies provided the evidences to support these viewpoints. Using the mental scanning paradigm, Borst and Kosslyn found that participants scanned the imagery and perception content at the same rate, indicating visual imagery and perception had the similar representation at the early stage of information processing [34]. Using the binocular rivalry paradigm, other studies found that, the beforehand formed visual imagery could interfere participants' perceptual pattern at the later stage [35, 36]. Also during the stage of generating imagery, binocular rivalry can be disturbed by an inserted visual distractor [37].

With respect to the relationship between mental imagery and attention, the *perceptual anticipation theory* proposed that mechanisms that used to generate mental imagery involve processes that used to anticipate perceiving stimuli [38, 39]. Imagery can lead to retinotopic activity in early visual areas, which is similar with the perceptual process. In the present study, once the mental imagery was generated, it provided a cue for the perceptual task. When performing the visual search task, the imagery could be regarded as a top-down priming cue for the search task, hence increased the search efficacy when it was a valid cue.

However, there was no memory-driven attentional capture found in this study, which was contradicted with a series of relative studies [2–5]. Several factors played crucial role on determining whether the memory-driven attentional capture occurs or not. Studies found that there will be observed less or no memory-driven attentional capture when the memory item is easy to verbalize [4, 40, 41], the time course between the onset of the memory item and the search display is long (longer than 3,500 ms) [40], and the visual search task is perceptually easy [2]. In is concluded that, maintaining a sufficiently strong visual representation then sustaining it during retention is necessary for the memory-driven attentional capture. In the present study, to keep consistence between the working memory condition and the mental imagery condition, and to make sure participants generate a stable imagery, the time course between the onset of the memory item and the visual display was 4,000 ms. Such a long time course was unfavorable for participants to maintain a strong visual representation in the working memory condition, so it was possible that memory-driven attentional capture was not observed.

It should be noted that, more and more researches started to think that mental imagery, perception (or attention) and memory are integrated, interrelated and intertwined, and may be affiliated to the same cognitive system [42]. When the formation of mental imagery relies on the perceptual information, mental imagery could guide attention to the representation of that object as a priming cue. When the formation of mental imagery relies on the information from memory, it can be regarded as a memory cue to lead people recall imagery-related information, or disturb memory at the memory retrieval phase [44]. Thus Tong proposed that mental imagery was a dynamic element of visual working memory [44]. Therefore, mental imagery connects perception and memory. It contains two different processes, one emphasizes the influence of outside input information on imagery, which could be regarded as a bottom-up processing. The other emphasizes the retrieval of past experience, which is a top-down processing.

In conclusion, this study provided the behavior evidence that mental imagery captures attention during visual search. The result may have implications on imagery training in different ways. For example, according to the imagery training on meter display and fault display, novice drivers could better adapt to and identify display problems in their further driving. Imagery training on vigilance could help the air traffic controller monitor the to and from aircrafts more efficiently. Also imagery training on different forms of dangerous goods helps the securities detect the dangerous targets faster and more accuracy.

Acknowledgement. This research was supported by 2015 Fujian Province Education Research Project for Young and Middle-aged Teachers (JAS150193).

References

1. Baddeley, A.D., Hitch, G.: Working memory. In: Bower, G.A. (ed.) Recent Advances in Learning and Motivation, pp. 47–90. Academic Press, New York (1974)

2. Han, S.W., Kim, M.S.: Do the contents of working memory capture attention? Yes, but cognitive control matters. J. Exp. Psychol. Percept. Perform. **35**(5), 1292–1302 (2009)
3. Lavie, N., De Fockert, J.: The role of working memory in attentional capture. Psychon. Bull. Rev. **12**(4), 669–674 (2005)
4. Olivers, C.N.L., Meijer, F., Theeuwes, J.: Feature-based memory-driven attentional capture: visual working memory content affects visual attention. J. Exp. Psychol. Hum. Percept. Perform. **32**(5), 1243–1265 (2006)
5. Soto, D., Heinke, D., Humphreys, G.W., Blanco, M.J.: Early, involuntary top-down guidance of attention from working memory. J. Exp. Psychol. Hum. Percept. Perform. **31**(2), 248–261 (2005)
6. Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. **18**, 193–222 (1995)
7. Awh, E., Jonides, J.: Overlapping mechanisms of attention and spatial working memory. Trends Cogn. Sci. **5**(3), 119–126 (2001)
8. Corbetta, M., Kincade, J.M., Shulman, G.L.: Neural systems for visual orienting and their relationships to spatial working memory. J. Cogn. Neurosci. **14**(3), 508–523 (2002)
9. Mayer, J.S., Bittner, R.A., Nikolic, D., Bledowski, C., Goebel, R., Linden, D.E.: Common neural substrates for visual working memory and attention. Neuroimage **36**, 441–453 (2007)
10. Pessoa, L., Ungerleider, L.G.: Top–down mechanisms for working memory and attentional processes. In: Gazzaniga, M.S. (ed.) The New Cognitive Neurosciences, pp. 919–930. MIT Press, Cambridge (2004)
11. Kosslyn, S.M.: Aspects of a cognitive neuroscience of mental imagery. Science **240**(4859), 1621–1626 (1988)
12. Farah, M.J.: Psychophysical evidence for shared representational medium for mental images and percepts. J. Exp. Psychol. Gen. **114**, 91–103 (1985)
13. Farah, M.J.: Mechanisms of imagery-perception interaction. J. Exp. Psychol. Hum. Percept. Perform. **15**, 203–211 (1989)
14. Pashler, H., Shiu, L.P.: Do images involuntarily trigger search? A test of Pillsbury's hypothesis. Psychon. Bull. Rev. **6**, 445–448 (1999)
15. Ganis, G., Thompson, W.L., Kosslyn, S.M.: Brain areas underlying visual mental imagery and visual perception: an fMRI study. Cogn. Brain. Res. **20**, 226–241 (2004)
16. Kosslyn, S.M., Thompson, W.L., Alpert, N.M.: Neural systems shared by visual imagery and visual perception: a positron emission tomography study. Neuroimage **6**, 320–334 (1997)
17. D'Esposito, M., Detre, J.A., Aguirre, G.K., Stallcup, M., Alsop, D.C., Tippet, L.J., Farah, M.J.: A functional MRI study of mental image generation. Neuropsychologia **35**(5), 725–730 (1997)
18. Mellet, E., Tzourio, M., Denis, M., Mazoyer, B.: A Positron emission tomography study of visual and mental spatial exploration. J. Cogn. Neurosci. **7**, 433–445 (1995)
19. Roland, P.E., Gulyas, B.: Visual memory, visual imagery, and visual recognition of large field patterns by the human brain: functional anatomy by positron emission tomography. Cereb. Cortex **5**, 79–93 (1995)
20. Klein, I., Paradis, A.-L., Poline, J.-B., Kosslyn, S.M., Le Bihan, D.: Transient activity in the human calcarine cortex during visual-mental imagery: an event-related fMRI study. J. Cogn. Neurosci. **12**(S2), 15–23 (2000)
21. Kosslyn, S.M., Alpert, N.M., Thompson, W.L., Maljkovic, V., Weise, S.B., Chabris, C.F., Hamilton, S.E., Rauch, S.L., Buonanno, F.S.: Visual mental imagery activates topographically organized visual cortex: PET investigations. J. Cogn. Neurosci. **5**, 263–287 (1993)
22. Le Bihan, D., Turner, R., Zeffiro, T.A., Cuënod, C.A., Jezzard, P., Bonnerot, V.: Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. Proc. Nat. Acad. Sci. **90**, 11802–11805 (1993)

23. Chen, Y., Martinez-Conde, S., Macknik, S.L., Bereshpolova, Y., Swadlow, H.A., Alonso, J.M.: Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. Nat. Neurosci. **11**, 974–982 (2008)
24. Meulen, M.V.D., Allali, G., Rieger, S.W., Assal, F., Vuilleumier, P.: The influence of individual motor imagery ability on cerebral recruitment during gait imagery. Hum. Brain Mapp. **35**(2), 455–470 (2014)
25. Schmidt, T.T., Ostwald, D., Blankenburg, F.: Imaging tactile imagery: changes in brain connectivity support perceptual grounding of mental images in primary sensory cortices. Neuroimage **98**(9), 216–224 (2014)
26. McNorgan, C.: A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. Front. Hum. Neurosci. **6**(42), 285 (2012)
27. Ishaï, A., Haxby, J.V., Ungerleider, G.: Visual imagery of famous faces: effects of memory and attention revealed by fMRI. NeuroImage **17**, 1729–1741 (2002)
28. Knauff, M., Kassubek, J., Mulack, T., Greenlee, M.W.: Cortical activation evoked by visual mental imagery as measured by fMRI. NeuroReport **11**, 3957–3962 (2000)
29. Kaas, A., Weigelt, S., Roebroeck, A., Kohler, A., Muckli, L.: Imagery of a moving object: the role of occipital cortex and human MT/V5+. Neuroimage **49**, 794–804 (2010)
30. Roland, P.E., Eriksson, L., Stone-Elander, S., Widen, L.: Does mental activity change the oxidative metabolism of the brain? J. Neurosci. **7**, 2373–2389 (1987)
31. Masson, M.E.J., Loftus, G.R.: Using confidence intervals for graphically based data interpretation. Can. J. Exp. Psychol. **57**, 203–220 (2003)
32. Thomas, N.J.T.: The multidimensional spectrum of imagination: images, dreams, hallucinations, and active, imaginative perception. Humanities **3**(2), 132–184 (2014)
33. Pearson, J., Naselaris, T., Holmes, E.A., Kosslyn, S.M.: Mental imagery: functional mechanisms and clinical applications. Trends Cogn. Sci. **19**(10), 590–602 (2015)
34. Borst, G., Kosslyn, S.M.: Visual mental imagery and visual perception: structural equivalence revealed by scanning processes. Mem. Cogn. **36**(4), 849–862 (2008)
35. Keogh, R., Pearson, J.: The sensory strength of voluntary visual imagery predicts visual working memory capacity. J. Vis. **14**(12), 1–13 (2014)
36. Pearson, J.: New directions in mental-imagery research: the binocular-rivalry technique and decoding fMRI patterns. Curr. Dir. Psychol. Sci. **23**(3), 178–183 (2014)
37. Chang, S., Lewis, D.E., Pearson, J.: The functional effects of color perception and color imagery. J. Vis. **13**(13), 227–232 (2013)
38. Kosslyn, S.M.: Mental images and the brain. Cogn. Neuropsychol. **22**(3–4), 333–347 (2005)
39. Kosslyn, S.M., Thompson, W.L., Sukel, K.E., Alpert, N.M.: Two types of image generation: evidence from PET. Cogn. Affect. Behav. Neurosci. **5**(1), 41–53 (2005)
40. Dombrowe, I., Olivers, C.N.L., Donk, M.: The time course of working memory effects on visual attention. Vis. Cogn. **18**(8), 1089–1112 (2010)
41. Olivers, C.N.L.: What drives memory-driven attentional capture? The effects of memory type, display type, and search type. J. Exp. Psychol. Hum. Percept. Perform. **35**(5), 1275–1291 (2009)
42. Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., de Lange, F.P.: Shared representations for working memory and mental imagery in early visual cortex. Curr. Biol. **23**, 1–5 (2013)
43. Kosslyn, S.M.: Mental images and the brain. Cogn. Neuropsychol. **22**, 333–347 (2005)
44. Tong, F.: Imagery and visual working memory: one and the same? Trends Cogn. Sci. **17**(10), 489–490 (2013)

Evaluation of the Usability and Playability of an Exergame for Executive Functions Stimulation and Its Development Process

João Batista Mossmann[1,4](✉), Eliseo Berni Reategui[1],
Débora Nice Ferrari Barbosa[2], Rochele Paz Fonseca[3],
Caroline de Oliveira Cardoso[3], and Vitor Caetano Silveira Valadares[4]

[1] PGIE, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
`mossmann@gmail.com`
[2] Doctored Cultural Diversity and Social Inclusion, Feevale University, Novo Hamburgo, Brazil
[3] Post-Graduate Program of Psychology - Human Cognition, Pontifical Catholic University of Rio Grande Do Sul (PUCRS), Porto Alegre, Brazil
[4] Digital Games, Feevale University, Novo Hamburgo, Brazil

Abstract. The present article shows the process of development of an active-play video game, also known as exergame, focused on stimulating Executive Functions (EF), more specifically inhibitory control in children in the early years of elementary school. In addition to presenting EF and inhibitory control concepts, this article also approaches the creation process of the game. Furthermore, it also analyzes data obtained from a pilot trial that was carried out for a month with a boy from the targeted group. The purpose of this trial was to evaluate if the project was suitable to be tested in an experimental group.

Keywords: Exergame · Educational games · Executive functions · Inhibitory control · Cognitive development

1 Introduction

This article evaluates the usability and playability of an exergame, whose development was part of the project, created for stimulating Executive Functions (EF), more specifically inhibitory control in primary schoolers. The main characteristic of the project is the use of an active-play video game, also called exergame, in which the players body is used as the game controller.

Diamond (Diamond 2013), states that it is possible to work on a subjects EF (including children) and thereby stimulate and improve their actions before different situations of daily life and in school. Amongst some tools used to stimulate EF there is a digital game entitled Cogmed, originally researched by Klingberk et al. (2005). The main goal of Cogmed is to promote the stimulation of working memory (one of the components of EF). For such purpose, however, the game must be played along with another specific program and the player must be

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 164–179, 2017.
DOI: 10.1007/978-3-319-58472-0_14

supervised by a trained person. Experiments with this game were carried out and results indicated academic improvement of students who have played it, since it exercises the working memory (Klingberk et al. 2005).

Thinking about exergames, it is possible to find practical applications in the school environment mainly regarding physical activities (Finco et al. 2015, Ahn et al. 2009). Moreover, Mossmann et al. (2015) investigated the occurrences of Cognitive Load Theory guidelines, which are proposed by the Australian psychologist John Sweller (Sweller et al. 2011), in an Exergame. Thus, they demonstrated the possible use of these games as instructional material and pedagogical support for different activities and school contexts. Furthermore, there are research articles that indicate that players learn new skills, perceptions, attitudes or behaviors from games that challenge them to think, explore and react (Glass et al. 2013).

Subjects of this research, who are elementary schoolers, presented difficulty in concentrating on tasks or in writing cohesively (Barbosa et al. 2016). Currently, it is possible to help children develop and improve executive skills through playful activities that work with reasoning, planning, as well as with inhibitory control (Diamond 2013). This research field is still very new in Brazil, however, there is a strong potential for innovation in the area, thus becoming more affordable for schools, specially for the public ones.

In this context, an exergame was developed to stimulate EF, focusing on inhibitory control, in elementary schoolers. The project, entitled The Incredible Adventures of Apollo and Rosetta in Space, was produced by a multidisciplinary working group composed by students, volunteers and professionals from psychology, computer science, pedagogy, and educational game design. The playfulness of the game is based on the narrative that was created based on an opinion poll carried out with the targeted audience. The game is composed of seven different mini-games, each one designed to stimulate different areas related to inhibitory control and to EF. The game is played through physical movements, and not controllers, which are captured by Kinect, a motion capture sensor.

Thus, this article presents the development process of an exergame project and the results obtained through a preliminary test performed with a child from the targeted audience. The child was a 10-year-old boy with whom the test sessions took place on the same day and time every week, in a reserved space at the Integrated Psychology Center (CIP) at Feevale (a university in Rio Grande do Sul, southern Brazil). All sessions were accompanied by members of the development team and recorded so that they could analyze the data obtained.

During each session, questions were raised about what should be corrected and improved in the game. Hence, different elements of the game were corrected, such as problems with the difficulty curve, with the gameplay or with the game instructions. The main changes were based on feedback received, such as the need for more explicit indicators to point out what action is expected from the player. Other issues mentioned by the team were related to technicalities of the system that were not relevant to the experiment and/or did not impact the performance of the test subject.

There is, currently, an intervention in which the previously described exergame is used for stimulating components of EF in elementary schoolers. The results allowed the team to evaluate and correct issues related to usability, playability and understanding of the rules of the game.

This article is organized in 5 sections. In Sect. 2 all the concepts necessary to understanding the topics related to EF and inhibitory control are provided. Afterwards, in Sect. 3, the exergame and the development process for stimulating inhibitory control are presented. Subsequently, Sect. 4 discusses the pilot test that was carried out, in addition to the description of the methodology used and the analysis of the results obtained. Finally, Sect. 5 addresses final considerations as well as possibilities for future works.

2 Executive Functions

Executive functions refer to a set of skills that regulate and control human behavior to achieve specific goals. There is a consensus among neuroscientists that divides EF into three main components: (1) inhibition control (Inhibitory Control); (2) working memory; and (3) cognitive flexibility (Diamond 2013). Recent studies have demonstrated that high levels of EF during childhood are directly linked to greater creativity, self-control and flexibility. Such skills are considered essential for physical and mental health, as well as for cognitive, social and psychological development, and for school and professional success. Nowadays, it is known that it is possible to help children develop and perfect executive skills through playful activities that work with reasoning, planning, and inhibitory control. Furthermore, there is evidence in literature that shows that physical activity also improves EF.

Adele Diamond (Diamond 2015) states that, as a consequence of the use of the cerebral prefrontal cortex, physical activities can lead to improvement in cognitive and brain functions. This area of the brain is associated with EF. So, when we perform some activity that is related to EF, such as planning, decision-making, predicting actions and social behaviors in general, the prefrontal cortex of the brain is activated.

Inhibitory control, one of the components of EF, is the ability to control behaviors, this being the capacity we have to stop inappropriate action/behavior. Inhibitory control allows us to modify and choose how to react and behave when facing particular situations or attitudes, thus allowing the person to postpone, inhibit or change a certain response according to the situation (Diamond 2013).

Inhibitory control is, for example, when two children are playing and eventually one of them bumps into the other and the child who was hit must control the impulse to not strike back by pushing the other child. Such act of inhibiting, controlling and resisting temptations by not acting impulsively is what we call inhibitory control.

As previously mentioned, studies have shown that inhibitory control skills, as well as EF, develop through childhood, adolescence and adulthood. In old age a decrease in inhibitory control can be seen (Diamond 2013, Luna 2009). According to Carlson and Moses (2001), during the first months of life, elementary forms of inhibitory control can be perceived in babies, and, in their first year, babies are already capable of inhibiting motor responses. However, the authors also point out that emotional and self-control are only developed in the end of preschool years, between the ages of zero and five years.

Clinically, it is known that children who have attention deficit and hyperactivity disorder (ADHD) show some deficit in inhibitory control (Salum et al. 2014). Therefore, stimulation of the EF component can help impaired children, as well as perfecting inhibitory control in typically developed children.

Inhibitory control is related to students academic performance (Brock et al. 2009, Visu-Petra et al. 2011), since it allows them to develop a series of abilities related to problem solving, reasoning and improving self-control. Thus, schools are expected to adapt their activities accordingly, providing the students with playful activities associated with practicing executive functions.

3 Exergame to Stimulate Inhibitory Control

For the development and creation of the exergame, entitled The Incredible Adventures of Apollo and Rosetta in Space, a multidisciplinary team was composed. There were researchers and students from several courses: Psychology, Computer Science, Pedagogy, Computer Science applied to education, and Digital Games. The digital game that was produced is composed by 7 activities (mini-games) that let the player deal with inhibitory control events. Each activity starts at the most basic level and continuously grows in difficulty.

During the development of the game, a fictional universe that was conceived having in mind the targeted audience was created. In order to construct this narrative that was designed to add playfulness to the game, Campbell (2008) process for developing narratives was used. The storytelling of the game was employed as a way to provide the student (player) with a pleasant and involving fictional universe. The Incredible Adventures of Apollo and Rosetta in Space has 3 main characters: (1) Apollo; (2) Rosetta; and (3) Master. Apollo and Rosetta are siblings who love space and Astronomy. Master is an elder alien, teacher at the School for Spatial Explorers and the character who guides the two siblings through the story.

The game has a space theme. The story starts with two siblings (Apollo and Rosetta) camping out in their yard. As it happens every Friday night, both of them wear astronaut costumes, camp out in the garden and, with a telescope, watch the sky to see the stars. When both brother and sister fall asleep inside the camping tent they start to dream. In the dream, Master appears and invites them to start studying at this school which graduates space explorers. Each of the 7 activities in the game has a different gameplay that was developed to stimulate the child's inhibitory control. In the game these activities are presented as the courses the protagonists must pass to complete the training at school.

The next part of the study shows the development process of Exergame and its activities. Following, data of the target group and the 7 activities that compose the game are presented.

3.1 Development Process

Creating digital games can be done through different approaches. Some of these come from well-established processes in Software Engineering, a course which major concern is to research and formalize such methods. This way, through such formalization of methods, a more repetitive methodology for developing software could be set, thus decreasing the number of errors (Sommerville 2004).

A study by Murphy-Hill (Murphy et al. 2014) shows that games represent a significant share in the software development industry. Nevertheless, this same study points out the little effort that Software Engineering researchers have made to study the specificities of the development process despite its significant differences.

The present work aims at understanding the singularities between the different development processes, keeping in mind, when concerning digital games, the necessity of managing individual and collective abilities throughout several stages of the project (Tschang 2005).

It is also necessary to add that digital games are creative products derived from a multidisciplinary team effort. Currently, the game industry is formed by professionals of many abilities and areas, such as computer science, design, communication, music, games (specialization) and many others, as well as the specialists which are required in the development of educational games.

This work had to consider the technical needs for a game development and the determination and formalities required so the goal of creating a game capable of stimulating inhibitory control could be reached. Baba and Tschang (2001) describe an outward spiral software development model (that works with recurrent revisions of the material produced), that was used to describe the different stages necessary for the development of the game described in this work.

Therefore, prototyping was used as means of allowing programming, testing and evaluating. It may also influence all new tasks in a cyclic and spiral form, thus, including the participation of an EF specialist during the whole development process. This way, there are 5 stages in the process: (1) Inspiration; (2) Concept; (3) Project; (4) Development; and (5) Testing and evaluating.

Figure 1 shows the different steps originally proposed as a model by Boehm (1986). In this work the spiral was inspired by Baba and Tschang (2001) work. Nonetheless, it was adapted so it fitted the development process of an educational digital game, which has no need to consider questions such as budget and market.

In the first stage, Inspiration, the idea of the game is described in a few words to all the participants. In this step there was the direct participation of a research group in clinical and experimental neuropsychology. From the beginning, meetings were carried out with the participation of the research group and a specialist in EF. The first meetings established a group of requirements that oriented the development of the game. Still during this step the team carried out an opinion poll with the targeted group. A group of 156 children in the 3rd and 4th year of elementary school answered a questionnaire with questions concerning their age, sex, favorite cartoons, characters and games, as well as where those games were played.

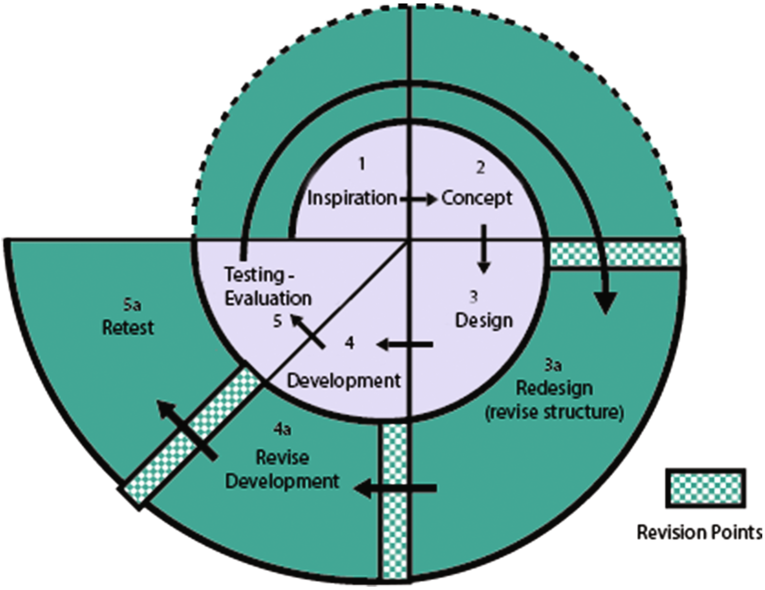


Fig. 1. Spiral software development model adapted from Baba and Tschang (2001).

In the 2nd stage, Concept, a representation of the subjects discussed previously in the first stage (Inspiration) was built. Visual diagrams were developed to show all characteristics necessary. Figure 2 is a model that shows the playability of one of the activities analyzed. During this stage the narrative and the aesthetics of the game were also discussed. Figure 3(A) shows the initial conceptual model of the activity Jumping Asteroids and in Fig. 3(B) the evolution of this activity after production can be seen.

During the third stage, Project. the development team created the basic structure of the game and planned how the software would be developed. The architecture of the system modules and the communication among the computer artifacts, the art and the narrative were also discussed and defined. Still in this stage, assignments and schedules for the next stages were specified and assigned to the team members.

In stage 4, Development, the art, characters' animation, narrative, sounds and interface of the game were programmed into what, after evolving in the spiral of development, would be the game. Figure 4(A) shows an activity called Deciphering Codes still in the initial phase of a playable prototype as opposed to Fig. 4(B) that shows the game finished and integrated with narrative, sound and art.

Finally, in the 5th stage, Evaluation, tests were performed by the team members. Problems of different sources were identified, such as art, programming, playability and specificity of the game (inhibitory control). In the aforementioned Fig. 3 the evolution of the activity Jumping Asteroids was seen. Initially,

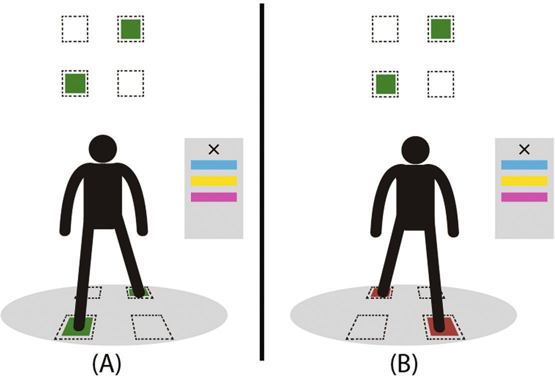


Fig. 2. An example of one of the activities playability.

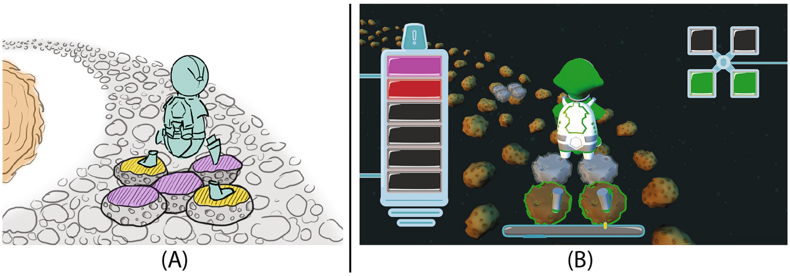
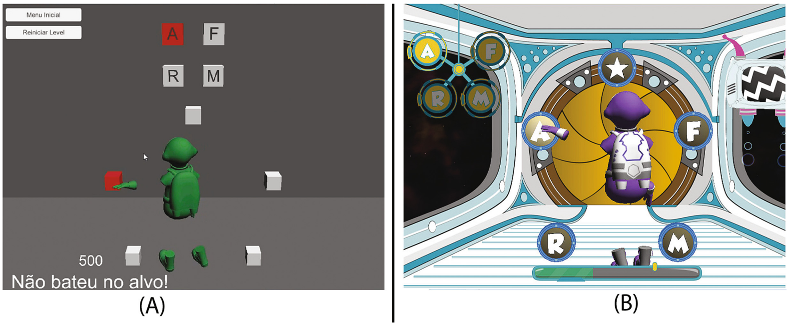


Fig. 3. Comparison between the initial concept and the final result of an activity.



Não bateu no alvo!

Fig. 4. Comparison between the initial prototype and the final version of the activity.

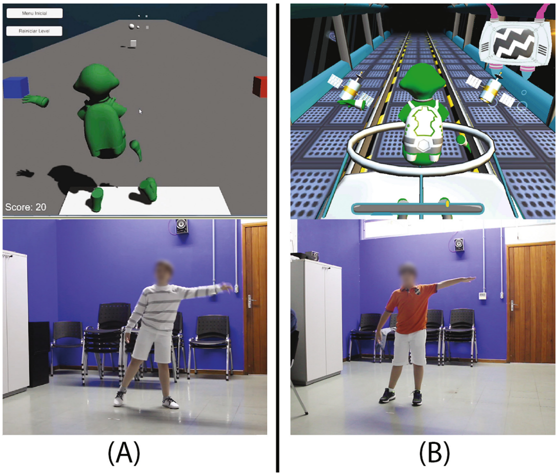


Fig. 5. Test of the prototype and the final version with a voluntary

this activity was conceived to have 5 asteroids (located underneath the main character), however it had to be altered to 4 due to its playability and stimulation of inhibitory control. In this context, Fig. 5 shows a test with a voluntary being (A) with the prototype of the game and (B) with the final version of the game. It is also important to stress that tests were trying the playability of the game and not its psychometric properties.

3.2 Mini-games

Activities in the game The Incredible Adventures of Apollo and Rosetta in Space were developed accordingly to the projects requirements listed during the development process. Therefore, there was a concern for the stimulation of inhibitory control, which was carried out through verbal and visual-spatial stimuli among others. Thus, this section aims at presenting all the 7 activities or mini games, as shown in Fig. 6, demonstrating their playability and how they stimulate the inhibitory control during gameplay.

In the mini-game called Explorer, Fig. 6(A), the player needs to catalog different items that are floating in the scenario. These items are listed on the left side of the screen. In order to win the game the player must collect only the correct items and bypass the obstacles. This mini-game demands that the child catalogue a series of specific elements in the beginning of each activity, this way the child tends to memorize them and respond automatically. Nonetheless, the child must restrain him/herself not to collect other objects, besides keeping focused on avoiding the obstacles (by jumping and crouching down) what stops the player from automatizing responses and keeps him/her attentive to what to do, thus not acting impulsively.

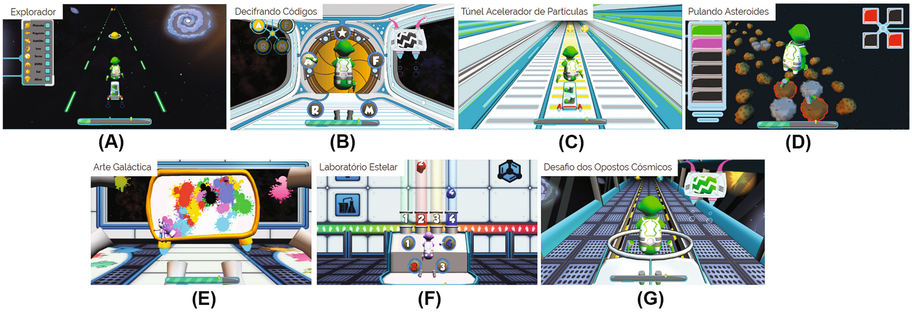


Fig. 6. Mini-games in the exergame The Incredible Adventures of Apollo and Rosetta in Space.

Figure 6(B) shows the mini-game Deciphering Codes in which the player has to press buttons that correspond to the letters drawn on his left or press the star button above his head. Throughout the game the player will listen to some words, and every time that the beginning letter of this word corresponds to a drawn letter the player has to press the star button. When playing the child cannot automatize any response, since there is no pattern in the game, thus inhibiting the tendency to repeat responses and keeping the child focused on the next command of the activity. Here, inhibitory control is also seen when the player restrains him/herself from touching the letters drawn when the word pronounced starts with the same letter.

The main goal of the Particle Accelerator, the mini-game shown in Fig. 6(C), is to guide the character by moving right and left and avoiding the obstacles which comes his way. Throughout the game the camera switches between two different angles, front and back view of the character. Therefore, the player has to react according to the perspective of the character instead of his/her own. This way, the game stimulates the inhibitory control of the child by inhibiting the natural tendency of the player to play according to his/her perspective. This activity exercises the cognitive flexibility of the player, demanding that the player stays focused on the character. In addition to that, the player must inhibit the predisposed response of playing with his/her perspective of sides instead of the character's when the camera changes between back and front view, just as the rules of the game state.

Another exergame mini-game is Jumping Asteroids, Fig. 6(D), which consists of making the player jump from one asteroid to another. On the top right of the screen there are 4 squares that represent the asteroids the player must be on after jumping. On the left side of the screen it is possible to see a list of colored squared, when one of the colors squares lights up the player must step on the asteroids that are not lit. Jumping Asteroids demands quick and different responses from the player. The game has in its programming a code that restricts some asteroid colors, this way invalidating the possibility of generalization of the responses. As a result, the child has to control and block responses that do not follow the rules of the game, hence stimulating the inhibitory control. The child also exercises the working memory, since he/she needs to take under consideration the colors of the asteroids that need to be inhibited.

In Galactic Art, the mini-game shown in Fig. 6(E), the player has to hit paint balls. These balls will pass through the screen and the player has to hit all of them in order to create a painting, avoiding the black and the white ones. During the game, a fly (computer controlled character) shoots black paint balls at the screen hindering the playability and making the player lose points in the final score. This game stimulates the inhibitory control because the player has to inhibit his/her response tendency suppressing the action of hitting all the paint balls and, as the rule of the game states, avoid hitting the black and the white ones. Moreover, the activity promotes a mechanic that can be classified as hot (emotional) when the fly appears and interferes with the game, ruining the painting and leading the player to lose points.

In the mini-game Stellar Laboratory, shown in Fig. 6(F), elements (alien vitamins) of different colors and numbers are presented. They slide into 4 transparent tubes, which are also marked with specific colors and numbers, and when the elements matching these colors and numbers reach a specific mark on the tubes the player has to press the button that represents that tube. The child who plays this game cannot automate the responses, thus exercising his/her inhibitory control. The player has to pay attention to the numbers and colors of the elements and the tubes present in the scenario.

Finally, in the mini-game Challenge of the Cosmic Opposites, in Fig. 6(G), the player has to collect different objects. There are two characters, Tivo and Ovit, who tell the player which objects to collect. When Tivo commands the player has to do what is asked. However, when Ovit commands the player has to do the exact opposite of what is asked. Challenge of the Cosmic Opposites stimulates inhibitory control and verbal comprehension, since during the game the child has to pay attention to the command spoken by the characters and, depending on the character, has to inhibit the tendency to do it and do the opposite of what is asked. Consequently, the game also exercises the player's cognitive flexibility.

4 Playability Evalutation

As already mentioned, an adaptation of the spiral process of development described by Baba and Tschang (2001) was used in this work. In the spiral there is the stage of evaluation, which was carried out as previously proposed. Still, in a more advanced stage of the development of the game the team thought to be important that another evaluation with a different volunteer was conducted.

Hence, during the month of June of the year 2016 the pilot project was subjected to tests. Sessions of gameplay with the volunteer were carried out with the objective of evaluating the playability of the game by a child in the targeted audience (Fig. 7). In this regard, an empiric evaluation of the difficulty curve was carried out, focusing on how complex each activity was as the player

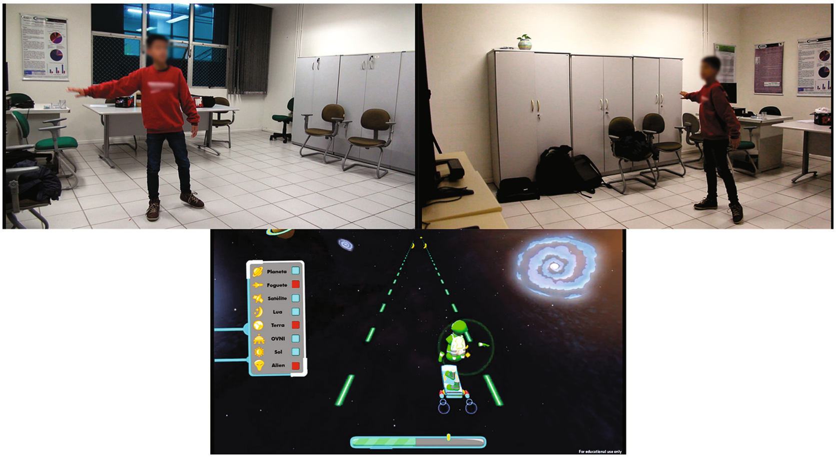


Fig. 7. Pilot test.

advanced levels. The rules and the narrative of the game were also observed in order to see if they were well understood by the child. In addition to that, technical problems that had to be corrected were also detected,

To this end, a child that had never got in touch with the project before was chosen among others in the targeted group. This child is a 10-year-old boy 5th grader. Tests were carried out in 5 one-hour-long sessions every Wednesdays at 6:30 p.m. at the Integrated Psychology Center (CIP) at Feevale. All sections were recorded in video and accompanied by at least two members of the development team who were responsible for organizing the room, orienting the child and recording everything that could be relevant for the final goal of the project. Videos were recorded by two cameras (Fig. 8): the first one (CAM 1) was used for capturing more open shots of the entire room and had the function of registering the complete movements made by the child; and the second camera (CAM 2) which function was to focus on the details of the movements made by the child, even if sometimes it was also used with a more broad view of the room.

In order to standardize and oversee the test sessions, checklists were created with a series of tasks the members of the development team should fulfill before, during and after the sessions. Such checklists contained the order in which the instructional videos should be shown to the child as well as which and in which order the mini-games should be played and for how long the child should play them in each session. There were also other technicalities such as the Kinect calibration, storing of the game records generated at the end of each session and the videos recorded. These lists had a space destined to general and specific comments about each session, such as suggestions, criticisms and technical problems detected as well.

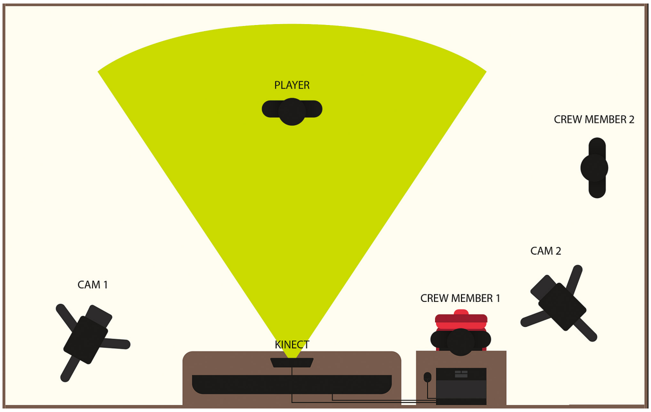


Fig. 8. Test room organization

4.1 Evaluation Sessions

During the 5 test sessions of the game the team members explained briefly to the child the objective of the project, the activities expected to be fulfilled in that session and the schedule for the next meetings. After that, the child signed in the game and two scenes of narrative were showed to the player. Initially, the introductory scene was shown to the child, in which the fictional universe of the game was contextualized. Afterwards, the video on psychoeducation, where the player saw topics such as inhibitory control and the main motto of the game (STD - stop, think, do), was presented. The team members questioned the child about the videos shown to be sure that the message was understood. The child said that he understood, mentioned the motto STD and even explained its meaning.

Before starting the activities the Kinect was calibrated according to the child's height so that he could have the proper experience. Due to the short time available in each session, the child could not play all the 7 activities, being able only to watch the cutscenes that contextualized those activities. The activities that could not be played at the time were alternated between the ones that were actually played on the day. The activities that were chosen to be played on the first day were Explorer, Jumping Asteroids, and Stellar Laboratory. For these, before actually playing them, a tutorial explaining how to play the mini-games was shown to the child. The team decided to use activities that demanded more physical movements in order to make the child more attentive to the narratives.

In the following sessions, the child could play all the 7 activities for at least 5 min each. Tutorials of the mini-games that had not been played on the first day were shown in the second session or whenever the boy felt necessary. Before starting each game the team members asked the child about the rules of the game or if he had any doubts about that specific activity. During the second and third sessions some of the instructions had to be recalled, since the child seemed insecure about how to proceed. Such problems were perceived specially concerning Challenge of the Cosmic Opposites and Jumping Asteroids.

In both cases the child had difficulty in perceiving changes of action during inhibitory control exercises. In the first activity he could not see the change in the button as well as thinking that the rule of the word said would stay the same for the next round. In the second, the child could not identify that his jump had not been validated and, as a consequence, the whole move would not be valid. In order to circumvent these problems, the team members established a fixed time between activities in Challenge of the Cosmic Opposites and reinforced the visual feedback of the button being pressed, thus avoiding any confusion in this aspect. For Jumping Asteroids the team added a board on the screen that informs the player if the jump has been validated or not. In addition to both of these activities, alterations were also necessary for Particle Accelerator in which the team noticed that, even if the child had the correct notion of space and moved to the right to move the character to the left as demanded, he could not perform the moves in time. In this case, signs that show the location of the obstacles off screen were added. Additional notes were related to technical problems with the system that were not relevant to the experiment and/or did not impact on the child's performance.

In the 5th and last session of the pilot test all 7 activities were played by the child. The same standard procedures used before were repeated, such as Kinect calibration and video recording. In the end, after all activities being played for at least 5 min, an ending cutscene to finish the game was shown to the player in order to provide the child with a sense of closure and accomplishment. This was the last session of the pilot test of the game. Over the five weeks of tests the team could see improvement in the performance of the child when playing, what was seen through the decreasing number of wrong moves even if the difficulty of the game had been progressively increasing.

4.2 Results

In the end of each of the aforementioned sessions, with the aid of the recorded videos, the team analyzed the different occurrences detected during the exergame playtime. Afterwards, such incidents were registered so that the development team could make the changes necessary.

In the first activity, Explorer, the list that indicated what items should be collected was altered to facilitate the player's understanding about which items could and should be collected and which should not. At first the list presented empty checkboxes when an item should not be collected and a checkbox filled in blue when the player had to collect the item. In order to circumvent this problem, the checkbox of the item that should not be collected was filled in with red. It was also difficult for the player to understand that the obstacles were objects that should not be collected, what was solved by adding a description in the tutorial of the activity. In addition to that, a jumping obstacle presented in this mini-game had to be modified, as it made the obstacle too difficult in the beginning levels since the speed of the track and the character was slower than the obstacle, thus making the anticipation of the movement much harder than in advanced levels of the game. After being modified, the character moved forward more rapidly when the player jumped.

In Deciphering Codes, because of the fact that when the player touched more than one button at the same time (not in an intentional manner) all of them were accepted, the game was modified so that the player could press only one button at a time. Visually the buttons were also changed, so that the player could see more clearly the difference between an active and an inactive button. When the button is inactive it looks faded and a smaller in contrast to when it is active, when it looks bigger and more colorful. Another change was carried out after the team noticed that the child had difficulties realizing when a gameplay had changed, especially in moments when the game included a new word. Due to this, a established delay between the ending and the beginning of a gameplay was added.

As for the activity Particles Accelerator the main difficulty was related to the inverted point of view (when the player has a front view of the character). However the biggest problem was not related exactly to the child not being able to perform the correct movement, but to the lack of time to take action. Obstacles appeared on the screen too close to the avatar, even though they were generated far from it, following the difficulty parameters for each level. To aid the player, visual indicators that indicate the next obstacle were added to the screen.

One of the main problems registered by the development team during Jumping Asteroids was related to the lack of feedback when the player jumped and that jump was validated by the game. For times the player jumped too low and missed the jump, or even continually jumped during the game hoping that this could result in an accepted play eventually. To solve this problem a visual indicator that confirms the move was added to the screen.

In the mini-game Galactic Art some problems related to balancing and controlling the game were identified. The player had trouble controlling the game, consequently ended up getting frustrated and getting a negative outcome as well. Furthermore, hitting wrong items combined with high speed in the first levels led the scores of the first tests to suffer big losses. The initial speed of the game was reduced, as well as improvements in the calibration of the movements and in the scoring of the game, in order to avoid loss of interest on the game by the player.

During the tests of Stellar Laboratory an exploit in which all the buttons could be activated at the same time and thereby all the plays were accepted by the game was discovered. In order to solve this problem a system update was implemented in which the game verifies if the player is pressing the correct button as well as verifying if the other buttons are not pressed, and if they are the move is considered invalid.

In the last activity, Challenge of the Cosmic Opposites, no modifications in the game were necessary. There were not any problems related to the understanding of the rules by the player and the incorrect plays were as expected for an inhibitory control activity.

Finally, it is important to point out some general technical questions that are not related specifically to any activity. One of these problems was with the registration screen, in which, once a player was detected, the game prevented the input of new data by dislocating the cursor that was controlled by Kinect. To control this problem points in the game where Kinect is deactivated were created.

Moreover, still thinking of technical questions, modifications in the system that interpret the Kinect capture to prevent system freezing, what led to delays in the sessions, since the console had to be turned off and on again. The calibration interface was also modified to avoid unintentional calibration caused by the player.

5 Conclusion

Keeping in mind that early stimulation of EF in children can be beneficial in short and long-term to the child's development, more research is necessary to invest in stimulation of cognitive processes. This way, the present work presents an exergame that focuses on exercising EF, more specifically inhibitory control in children.

This project makes it possible for this game to go from an academic laboratory to benefit the community, through a proposal of a game for cognitive stimulation supported by applied neuropsychology in a clinical and educational environment. Such proposal is justified by the necessity of an innovative program with enhanced techniques to support cognitive development during childhood and adolescence. The project enables the use of activities compatible with the range of cognitive development expected from each school stage and age range of the children.

The achievement of a project that approaches neuropsychology, clinical psychology and education is a very important alternative to fulfill gaps that properly met can contribute in a crucial manner to educational practice. Thus, the authors aim at stimulating the academic community by, along with schools, implementing a program for cognitive stimulation by using this exergame created by the team,

References

Ahn, M., Kwon, S., Park, B., Cho, K., Choe, S.P., Hwang, I., Jang, H., Park, J., Rhee, Y., Song, J.: Running or gaming. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology, pp. 345–348 (2009)

Baba, Y., Tschang, F.T.: Product development in Japanese TV game software: the case of an innovative game. Int. J. Innov. Manage. **05**(04), 487–515 (2001)

Barbosa, D.N.F., Bassani, P.B.S., Martins, R.L., Mossmann, J.B., Barbosa, J.L.V.: Using mobile learning in formal and non-formal educational settings. In: Zaphiris, P., Ioannou, A. (eds.) LCT 2016. LNCS, vol. 9753, pp. 269–280. Springer, Cham (2016). doi:10.1007/978-3-319-39483-1_25

Boehm, B.W.: A spiral model of software development and enhancement. SIGSOFT Softw. Eng. Not. **11**(4), 14–24 (1986)

Brock, L.L., Rimm-Kaufman, S.E., Nathanson, L., Grimm, K.J.: The contributions of hotand cool executive function to childrens academic achievement, learning-related behaviors, and engagement in kindergarten. Early Childhood Res. Q. **24**(3), 337–349 (2009)

Campbell, J.: The Hero with a Thousand Faces, vol. 17. New World Library, New York (2008)

Carlson, S.M., Moses, L.J.: Individual differences in inhibitory control and childrens theory of mind. Child Devel. **72**(4), 1032–1053 (2001)

Diamond, A.: Executive functions. Ann. Rev. Psychol. **64**, 135–168 (2013)

Diamond, A.: Effects of physical exercise on executive functions: going beyond simply moving to moving with thought. Ann. Sports Med. Res. **2**(1), 1011 (2015)

Finco, M.D., Reategui, E., Zaro, M.A., Sheehan, D.D., Katz, L.: Exergaming as an alternative for students unmotivated to participate in regular physical education classes. Int. J. Game-Based Learn. (IJGBL) **5**(3), 1–10 (2015)

Glass, B.D., Maddox, W.T., Love, B.C.: Real-time strategy game training: emergence of a cognitive flexibility trait. PLoS One **8**(8), e70350 (2013)

Klingberg, T., Fernell, E., Olesen, P.J., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C.G., Forssberg, H., Westerberg, H.: Computerized training of working memory in children with ADHD-a randomized, controlled trial. J. Am. Acad. Child Adolesc. Psychiatry **44**(2), 177–186 (2005)

Luna, B.: Developmental changes in cognitive control through adolescence. Adv. Child Devel. Behav. **37**, 233–278 (2009)

Mossmann, J.B., Barbosa, D.N.F., Barros, P.R., Silva, R.N., Valadares, V.C.S., Goncalves, B.T., Fischer, R.E.C., Brochetto, V.: Busca de evidências da aplicação da teoria da carga cognitiva em um exergame de dança. In: SBC - Proceedings of SBGames 2015, pp. 368–375 (2015)

Murphy-Hill, E., Zimmermann, T., Nagappan, N.: Cowboys, ankle sprains, and keepers of quality: how is video game development different from software development? In: Proceedings of the 36th International Conference on Software Engineering, ICSE 2014, New York, NY, USA, pp. 1–11 (2014)

Salum, G., Sergeant, J., Sonuga-Barke, E., Vandekerckhove, J., Gadelha, A., Pan, P., Moriyama, T., Graeff-Martins, A., de Alvarenga, P.G., do Rosario, M., et al.: Specificity of basic information processing and inhibitory control in attention deficit hyperactivity disorder. Psychol. Med. **44**(03), 617–631 (2014)

Sommerville, I.: Software Engineering, 7th edn. Addison-Wesley, Reading (2004)

Sweller, J., Ayres, P., Kalyuga, S.: Cognitive Load Theory. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies. Springer, New York (2011)

Tschang, F.T.: Videogames as interactive experiential products and their manner of development. Int. J. Innov. Manage. **9**(01), 103–131 (2005)

Visu-Petra, L., Cheie, L., Benga, O., Miclea, M.: Cognitive control goes to school: the impact of executive functions on academic performance. Procedia-Soc. Behav. Sci. **11**, 240–244 (2011)

Understanding the Relations Between Self-concept and Causal Attributions Regarding Computer Use

Adelka Niels(✉) and Monique Janneck

Faculty of Electrical Engineering and Computer Science,
Luebeck University of Applied Sciences, Luebeck, Germany
{adelka.niels,monique.janneck}@fh-luebeck.de

Abstract. This study examines the correlations between self-concept and causal attributions regarding computer use. The computer-related self-concept is a psychological construct that describes computer-related experiences, interests, motivations, attitudes, and competencies. Computer-related causal attributions are subjective causal explanations for successful and unsuccessful computer-related outcomes that are known to influence users' behavior, motivation, and emotions. The goal is to shed more light on computer-related attitudes, behaviors, and cognitive characteristics of computer users. Two standardized questionnaires were used to measure the participants' computer-related self-concept and their computer-related causal attributions. The tests were conducted in a laboratory study with N = 64 participants. Relations between computer-related self-concept and attributional dimensions (locus, stability, controllability, globality) as well as attributional patterns (ability, effort, luck, and task difficulty) were examined. In contrast to other research domains (e.g., education), it was found that users with a positive computer-related self-concept attribute causes for successes and failures predominantly to external reasons. With respect to the attributional patterns in situations of failure, correlation and post-hoc analyses showed results compared to findings in other research domains: Users that attribute causes of failures to their own abilities (internal-stable) showed a significantly more negative computer-related self-concept than users that attribute causes of failures to factors such as effort, task difficulty, or luck. In situations of success, no significant correlations were found.

Keywords: Cognitive psychology · Attribution theory · Self-concept · User experience · Personality traits · User types

1 Introduction and Theoretical Background

It is well-known that personal characteristics like experience, expertise, but also self-confidence and attitudes towards computer technology play an important role in how people interact with computers. In this regard, cognitive theories are promising to better understand user behavior and to design computer technology that better fits the user needs.

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 180–199, 2017.
DOI: 10.1007/978-3-319-58472-0_15

This paper is devoted to two well-known cognitive theories, namely self-concept and attribution theory, and investigates their connections in the field of human-computer interaction (HCI).

The *self-concept* is defined as the sum of an individual's self-related beliefs, attitudes, and expectations [1]. These perceptions refer to a number of areas, for example, intellectual status, physical and physical appearance, personality traits, and emotional tendencies. Theorists consider the self-concept as generated from lifelong experiences, which are developed by individuals' interactions with their surroundings [2].

Attribution theory, in turn, strives to explain how causal explanations influence individuals' motivation, emotions, and behavior [3]. Attribution theorists assume that people try to understand themselves and the world around them in order to achieve cognitive mastery of their environment's causal structure [4].

In prior research, these two theories have been independently investigated in the field of computer-related self-perceptions and both provided valuable insights into the understanding of learning and behavior of computer users. However, the direct link between these two theories has not been investigated yet, especially in the field of HCI, which is the contribution of the current study. The goal is to shed more light on computer-related attitudes, behaviors, and cognitive characteristics of computer users.

This paper is structured as follows: Sect. 1 introduces the two constructs of self-concept and attribution theory in general as well as with regard to HCI, followed by a brief description of the connections between these theories. Section 2 provides a description of our research questions and methodology. The description of the analysis and the results obtained in this study are outlined in Sect. 3. Finally, we conclude with a discussion of implications for practice and research, limitations, as well as an outlook for future research.

1.1 Self-concept

The *self-concept*, in general, denotes the total of all cognitive representations which an individual has stored in his/her memory [5–7]. These self-referred representations relate to different areas, for example, one's own intellectual abilities, aspects of one's body, or one's experiences and behavior. Hence, the self-concept is multifaceted and can be subdivided, for example, into an academic, a social, an emotional and a physical self-concept, which in turn can be differentiated into further sub-areas [8].

In the field of social sciences, the self-concept is one of the best-explored constructs (Hattie 1992) and many theories about the "self" have been proposed in the last 100 years. In this regard, an early distinction has been made between two global aspects of the self: the *self-knowing self* (self as knower, I, pure ego) and the *recognized self* (self as known, me, empirical ego). Moreover, in describing the self, a distinction can be made between the material self (body, family, and possessions), the social self (views others hold of the individual), and the spiritual self (emotions) [9]. Another example is the *looking-glass self*, which describes the self-concept as an developing result of the perceived impressions and evaluations in social interaction [10].

There are two main theoretical approaches in self-concept research: the *unifactorial construct* and the *multifaceted hierarchical construct*.

Representatives of the *unifactorial approach* hold the opinion that the self-concept is dominated by a single general or global self-concept and that individual or specific factors can not be adequately differentiated [e.g., 11, 12]. For example, they claim that children are too immature to make distinctions between different facets of self-concept.

The proponents of the *multifaceted hierarchical construct* claim that the global self-concept is built up from a number of different facets. The global self-concept is at the top of the hierarchical tree, which then splits into smaller components such as academic, social, emotional and physical self-concepts [e.g., 8, 13].

Hansford and Hattie [14] examined the correlations between different self-concept variables and performance measures. They found that the correlation became significantly larger when they merely considered studies in which only the academic self-concept was measured. **Other researchers discovered that school performance is highly correlated with the domain-specific self-concept, but is only moderately related to the general self-concept [15].** These findings illustrate that the self-concept is domain-specific and thus today's self-concept research assumes that self-related cognitions mainly refer to specific sub-areas. For this reason, recent research mostly does not refer to a general self-concept but is narrowed to domain-specific self-concepts. This study is conducted within the multi-faceted hierarchical paradigm and focuses on assessing individuals' levels of specific computer-related self-concept.

Computer-Related Self-concept. The *computer-related self-concept* is a quite new psychological construct that is especially related to computer-related experiences, interests, motivations, attitudes, and competencies [16]. The construct is based on the Three-Components-Model of Attitudes [17] and comprises:

- The *conative component* representing the concrete actions, behaviors, or specific experiences with regard to dealing with computers (experiences in childhood, youth, and adulthood);
- The *motivational component* describing emotional and content-specific motifs for using computers (fun, joy, fascination, computer anxiety, and individual motives in using computers);
- The *cognitive component* including the subjectively perceived competence and self-efficacy regarding the handling of computers, attribution processes as well as strategies for dealing with information technology (computer-related self-perception of competencies, self-efficacy, attribution processes, and strategies).

1.2 Attribution Theory

Attribution theory originates in the 1950s and was developed when psychologists began to be more interested in studying the human's perception of causes for their behavior rather than their perception of the behavior itself. Attributions are subjective causal explanations for successful and unsuccessful outcomes that are known to influence individuals' behavior, motivation, and emotions [18].

Heider [4] was one of the first who investigated how people try to understand the causes of their own actions. He argued that individuals are trying to analyze, as a kind of 'naive scientist', the causes of an action. According to Heider, the cause of an action is attributed either to *internal* or *external* factors. For example, a person may either feel responsible for a positive or negative outcome (internal) or relate it to external circumstances [4]. This first dimension is called the *locus* (or *locus of control*) dimension and served as the basis for subsequent attribution theories as well as further attributional dimensions [19].

The subsequent theory that we built upon in this paper is called the *achievement-related attribution theory*. It states that individuals generally attribute their failure and success in performance-related situations to one of four *attributional patterns*. These attributional patterns are *ability*, *effort*, *task difficulty*, and *luck* [18]. *Ability* is defined as the knowledge and abilities a person believes to have to perform a task. *Effort* is defined as the physical and mental energy a person exercises in performing a task. *Task difficulty* is defined as how easy or difficult a task is perceived. Finally, *luck* is defined as that role that opportunity plays in performing a task [18]. These attributions, in turn, can be determined by means of the four attributional dimensions of *locus*, *stability*, *controllability*, and *globality* [20]. Table 1 illustrates the relationships.

Table 1. Attributional classification scheme: relations between attributional dimensions and attributional patterns for failure and success situations [20].

		Locus	
		Internal	External
Stability	Stable	Ability	Task difficulty
	Unstable	Effort	Luck

Ability and *effort* are considered as *internal* attributions because they relate directly to the person who makes the attribution. On the contrary, *task difficulty* and *luck* are regarded as *external* attributions because the person ascribes the cause to external circumstances.

The second dimension to classify attributions is the *stability* dimension. In this regard, causes are considered as *stable* over time (recurring) or as *unstable* (singular event) [18]. *Ability* and *task difficulty* are regarded as *stable* attributions. Conversely, *effort* and *luck* are considered as *unstable* attributions.

Two more dimensions are used to describe individuals' attributions: The *controllability* dimension describes whether a cause is perceived as *controllable* (easy to change) or as *uncontrollable* (hard to change or even unchangeable) [18].

Finally, the *globality* dimension distinguishes between causes perceived as *global* (generally valid) or only valid for *specific* (certain) situations [21].

Computer-Related Attribution Theory. The investigation of attributions is fairly young in the field of HCI. However, distinct attributional patterns with favorable as well as unfavorable attribution styles were observed by Niels and Janneck [22] in the information systems context. Moreover, demographic factors (e.g., age, gender, self-assessed computer skills) [23], as well as the impact of different attribution patterns on the assessment of computer systems were examined [24]. However, to our knowledge, relations between computer-related self-concept and causal attributions have not been investigated yet.

1.3 Connection Between Self-concept and Attribution Theory

In other research domains, particularly in the educational context, research on attribution theory and self-concept has already spawned important implications for research and practice. Attribution theory allowed researchers to gain insights into what causes students identify as the reason for their performance. They also found that these perceived causes are crucial because they influence students' future expectations, motivation, and subsequent behavior [25]. Furthermore, the self-concept also affects learning behavior: Those with a positive self-concept performed better than those with a negative self-concept [2].

First connections between self-concept and attributions were discovered rather by chance when spontaneous remarks by participants were additionally recorded in a self-concept study [26]. It turned out that those participants with a positive self-concept externalized causes for insoluble tasks while participants with negative self-concept rather attributed their failure to internal causes. Subsequent research was able to replicate these results [e.g., 18, 27–30]. This connection was then further investigated by other researchers [e.g., 31–33] and especially by Weiner [18]. Weiner found that individuals with a positive self-concept tend to attribute success to their own *ability* and individuals with a negative self-concept tend to attribute failure to unstable causes such as *effort* or *luck*. This was also found in an educational context. For example, students with a positive self-concept were more likely to attribute failures to external-unstable factors (luck), while students with a negative self-concept tend to attribute failures to internal-stable factors (ability). In situations of success, the opposite is the case. In summary, more correlations in situations of failure than in situations of success were found [34–38].

In this regard, a number of studies focused on the attributional dimensions (i.e., *locus*, *stability*, *controllability*, *globality*), but there is also a series of studies that considered the four causal attributional patterns (*ability*, *effort*, *task difficulty*, and *luck*). However, both methods yield very similar results.

In contrast to the other components of the self-concept, attributions are considered as influenceable, for example, through measures such as the so-called reattribution training. The aim of the training is to transform unfavorable attribution patterns into favorable attribution patterns. This transformation, in turn, has a positive effect on the self-concept [39–41].

2 Research Questions and Methodology

So far, research has shown a fairly consistent pattern: individuals with a positive self-concept mainly attribute success to internal and failures to external causes. For individuals with a negative self-concept, the opposite is the case. However, most of these prior studies were conducted with children or students in educational context. Thus, in order to draw conclusions in the field of HCI, it is important to examine these relations in this specific context and with "average" computer users since the self-concept [42] as well as attributions [43] are deemed to be domain-specific.

The present study differs from previous studies by exploring relationships between self-concept and causal attributions by focusing on computer-related situations; in particular, regarding computer-related failure and success situations experienced in "real-life". The aim is to shed more light on computer-related attitudes, behaviors, and cognitive characteristics of computer users. For this purpose, we investigated causal attributions as well as the computer-related self-concept in a laboratory setting by conducting usability tests. We investigated correlations between the attributional dimensions (locus, stability, controllability, and globality) as well as the attributional patterns (ability, effort, task difficulty, and luck) and the computer-related self-concept. This is one of the first studies combining these two cognitive theories with regard to computer use, thus making a unique contribution to the body of knowledge in the field of HCI.

2.1 Measurements

Next to the assessment of general demographic aspects (age, gender, education, general computer use and skills), two validated questionnaires were used to cover the aspects of computer-related self-concept and computer-related causal attributions.

Computer-Related Self-concept Questionnaire (CSC). The standardized *Computer-Related Self-concept Questionnaire* developed by Janneck, Vincent-Höper and Ehrhardt [16] was used to measure the participants' level of computer-related self-concept. The measurement focuses on the individual's self-perceptions rather than attempting to infer their self-concept by observing the behavior or the attributions of others. The questionnaire consists of 11 subscales with a total of 27 items, col-lecting conative, motivational, and cognitive computer-related self-ratings using a five-point Likert scale ("1 = strongly disagree" to "5 = strongly agree"). The *Computer-Related Self-Concept Questionnaire* has proven to have a satisfactory reliability and validity [16]. Table 2 shows the English version of the questionnaire. This questionnaire was chosen because, to the best of our knowledge, it is the only questionnaire that relates to computer-related situations. Moreover, the questionnaire is simple to evaluate and provides quantitative values that allow easy comparison.

Computer-Related Attribution Questionnaire. The standardized *Attribution Questionnaire* developed by Guczka and Janneck [44] was used to measure computer-related attributions. The questionnaire distinguishes between failure and success situations, as usually done in attribution research, and contains four items to measure the attributional dimensions of *locus*, *stability*, *controllability* and *globality*. The questionnaire is based on the Sport Attributional Style Scale, SASS [45]. Table 3 shows the English version of the questionnaire relating to situations of failure. Items measuring attributions of success are worded analogously. This questionnaire was chosen because it allows examining the attributional dimensions separately, but also to determine the attributional patterns.

Table 2. The Computer-Related Self-Concept Questionnaire [16].

	Subscales and Items
Conative	*Practical experiences*
	I gained practical experiences using computers at an early age
	I did not deal with computers much in my childhood and youth*
	In my free time, I use computers a lot (e.g. Internet, picture editing, games…)
Motivational	*Positive emotions*
	I have a lot of fun using computers
	Computer technology fascinates me
	Computer anxiety
	When dealing with computers, I am afraid of doing something wrong*
	When using computers, I am scared of screwing something up or deleting data*
	I feel anxious when using computers*
	Understanding
	I want to understand how computers work
	When computer technology does not work, I want to understand the reason why
	Creating
	I like to create things with computers (e.g. programming, graphics processing…)
	I deal with computers because I can create things with them
	Tool perspective
	For me, computers are only a means to an end*
	I simply want computers to work alright, I'm not interested in technical details*
Cognitive	*Computer-related competencies*
	I consider myself very competent at handling computer devices
	I'm more confident than the average person when using computers
	I have comprehensive computer knowledge
	Computer-related self-efficacy
	I keep cool when confronted with computer problems because I can always rely on my computer skills
	I'm up for most computer-related challenges
	When I'm confronted with computer problems, I find ways and means to solve them
	When I try hard, I generally succeed in finding solutions to computer problems
	Computer-related internal attribution
	When a computer device doesn't work right, generally it is because I have done something wrong*
	When a computer problem occurs, usually I have caused it*
	Computer-related external control beliefs
	I have no control over computer difficulties that occur*
	Software functionality often seems random to me*
	Computer-related strategies
	I am not shy about simply trying out new software
	I usually try out new computer applications intuitively at first

*Note: Items ranging from "1 = strongly disagree" to "5 = strongly agree". Items denoted by * are inverse coded. Translation of the German questionnaire.*

Table 3. Excerpt from the Attribution Questionnaire for failure situations [44].

What caused the breakdown?		
I would locate the cause of the breakdown… internally (I am to blame)	1 2 3 4 5 6 7	Externally (the system is to blame)
The cause of the breakdown is…a singular event	1 2 3 4 5 6 7	Recurring
The cause of the breakdown is…Controllable	1 2 3 4 5 6 7	Uncontrollable
The cause of the breakdown is likely to promote other breakdowns…just in this situation	1 2 3 4 5 6 7	In other situations as well

Note: Items for success situations are worded analogously. Translation of the German questionnaire.

Overall, evaluative statements made by individuals about themselves are considered to be valid and reliable data sources and self-reports provide valuable information about the individual [e.g., 46–48].

2.2 Procedure

Data was collected in a laboratory setting, conducting usability tests. First, the participants filled out a questionnaire containing the demographic questions as mentioned above, as well as the *Computer-Related Self-Concept Questionnaire* (Table 2).

Afterwards, the participants were asked to edit three task pairings on three different applications or devices, whereby one task of each pairing was easy to solve (situation of success – e.g.: Search for the district office opening hours on a municipal home page) and one task was hard or even unsolvable (situation of failure – e.g.: Search for the building regulations of a certain district, which did not exist on the homepage). The tests included two tasks with a web application (website), two tasks with a desktop application (spreadsheet program), and two tasks with a mobile application (nutrition application).

To slightly increase the pressure, the participants were told that they have five minutes only to solve each task. In fact, the timeframe was not strictly adhered to. If necessary the participants had little more time to solve the success tasks and less time to solve the failure tasks. However, the processing time varied one minute at maximum.

Before the participants started with the first task, they were explained with the aid of an example how to answer the questions that would follow the tasks. Then the participants started working on the tasks. After each task, the participants completed the *Computer-Related Attribution Questionnaire* (Table 3).

In all, a total of 340 situations (163 failure and 177 success situations) were recorded. This imbalance was due to individual perceptions of the outcome of the task: For example, some participants also succeeded in the hard task condition, while others were not successful in the easy task condition. Moreover, the participants decided by themselves whether they successfully solved a task or not. For example, some participants failed objectively but nevertheless believed they had been successful.

On average, the participants needed one hour to complete the test. They were not paid for their participation.

2.3 Sample

In order to obtain a balanced sample, the participants were selected according to the following criteria: Approximately the same number of female and male persons, about one-third of them aged between 14 and 25 years, one-third between 26 and 45 years and one-third aged 46 years and above.

In all, 64 persons participated in the study (46,9% female, 53,1% male). Mean age was 38.20 years (Median = 31, SD = 16.97, range: 17–75 years). The general level of education was quite high (75% with a high school or university degree). Participants were rather experienced computer users. On average they had 13.88 years (Median = 14, SD = 7.17, range: 3–32 years) of experience in computer use and they used computers on average 5.77 h a day (Median = 6, range: 3–14 h). Participants self-rated their computer skills on a Likert scale ranging from 1 (low) to 7 (expert) on average at 4.27 (Median = 4.33, SD = 1.61, range: 1–7).

3 Data Analysis and Results

In order to examine the relations between computer-related self-concept and computer-related causal attributions, we first analyzed the computer-related self-concept mean values for each subscale as well as the mean value of the total scale. Secondly, we analyzed the attribution questionnaire and calculated the means for each attributional dimension (locus, stability, controllability, and globality), separately for situations of failure and success. Furthermore, the attributional patterns (ability, effort, task difficulty, and luck) were determined based on Weiner's attribution matrix [3].

Finally, correlations between self-concept scales and attributional dimensions as well as correlations between self-concept scales and attributional patterns were calculated. In the following sections, methods and results are explained.

3.1 Computer-Related Self-concept

In a first step, inverse coded item values were transformed, as the order of positive and negative terms is randomized in the questionnaire. Secondly, means (M), standard deviations (SD), and internal consistencies (Cronbach's α) for the subscales as well as for the overall self-concept (total scale) were calculated (Table 4). Higher values indicate a more positive computer-related self-concept.

In general, participants reported a quite positive computer-related self-concept (M = 3.12). The subscales with the high-est means were *computer anxiety* (M = 4.11), *strategies* (M = 3.77), and *external control beliefs* (M = 3.30). The scales indicating a negative or unfavorable self-concept were *tool perspective* (M = 2.41), *internal attribution* (M = 2.67), and *creating* (M = 2.77). The subscales show acceptable to good reliability coefficients.

Table 4. Results Computer-Related Self-Concept Questionnaire subscales and total scale.

Self-concept scales	M	SD	α
Practical experiences	2.97	1.09	.568
Positive emotions	2.86	1.30	.794
Computer anxiety	4.11	1.09	.900
Understanding	3.11	1.31	.848
Creating	2.77	1.49	.899
Tool perspective	2.41	1.21	.752
Competencies	2.82	1.17	.908
Self-efficacy	3.21	1.08	.889
Internal attributions	2.67	1.09	.830
External control beliefs	3.30	1.03	.684
Strategies	3.77	1.33	.772
Total scale	3.12	0.90	.953

Note: M = Mean, SD = standard deviation, α = internal consistency (Cronbach's α).

3.2 Computer-Related Causal Attributions

In our analysis, we distinguished between situations of failure and success, as it is usually done in attribution research. First, the mean value for each attributional dimension was calculated (Table 5). Results show that the locus dimension is nearly equally distributed. The participants see internal as well as external reasons for their failure and success. Furthermore, the causes are perceived to be stable over time, persist in different situations (global), and are perceived as controllable. Overall, this represents more positive attributional patterns.

Table 5. Results Attribution Questionnaire dimensions for success and failure situations.

Attributional dimension	Failure		Success	
	M	SD	M	SD
Locus	4.18	1.71	3.73	1.45
Stability	5.61	1.12	5.04	1.47
Controllability	3.55	1.69	2.14	1.02
Globality	4.69	1.39	4.69	1.39

Note: M = Mean, SD = standard deviation.

Secondly, the attributional patterns were determined for each participant separately for failure and success situations, based on Weiner's attribution matrix [3]. This was done by combining the attributional dimensions of *locus* and *stability* and clustering the combinations into the four attributional patterns of *ability*, *effort*, *task difficulty*, and *luck*. Participants who attributed the reasons for success or failure, respectively, to internal/stable causes were assigned to the attributional pattern *ability*. Participants who attributed the causes to external/stable reasons were assigned to the *task difficulty* pattern, participants who attributed internal/unstable to the *effort* pattern, and participants who attributed the reasons to external/unstable causes to the attributional pattern of *luck*. The attributional pattern matrix, as well as the number of participants in each pattern, separately for situations of failure and success, is shown in Table 6.

Table 6. Attributional pattern classification matrix based on Weiner's [3] attribution theory and number of participants for each pattern in situations of failure and success.

	Internal			External		
Stable		Failure	Success		Failure	Success
	Ability	n = 11	n = 16	Task difficulty	n = 24	n = 13
Unstable	Effort	n = 19	n = 18	Luck	n = 10	n = 17

3.3 Relationship Between Self-concept and Attributional Dimensions

Correlations (Spearmen's Rho) between the computer-related self-concept total scale as well as the subscales and the attributional dimensions (locus, stability, controllability, and globality) were calculated separately for failure and success (Table 7).

Regarding the self-concept total scale, analyses revealed significant correlations for the *locus* dimension: Users with a positive computer-related self-concept attributed causes for failure ($r = 0.550$, $p = 0.000$) and success ($r = 0.266$, $p = 0.034$) predominantly to external reasons. The latter is interesting, as this does not correspond to the findings from other domains where individuals with positive self-concept usually attribute success to internal reasons [34–38]. For the other attributional dimensions, fewer correlations were found.

With regard to the self-concept subscales in *failure situations*, significant positive correlations between the attributional dimension of *locus* and all self-concept subscales were found. Furthermore, a positive correlation between the self-concept subscale *internal attributions* and the attributional dimension of *controllability* was found ($r = 0.265$, $p = 0.034$). Individuals who generally tend to attribute failures to external causes are more likely to perceive the cause to be uncontrollable. Moreover, a negative correlation between the self-concept subscale of *internal attributions* and *globality* could be established ($r = -0.304$, $p = 0.015$). Individuals who attribute failures to external causes believe that the cause only affects a specific situation and is not valid in other computer-related situations.

In *success situations*, fewer significant correlations were found. First, significant positive correlations between the attitudinal dimension of *locus* and the self-concept subscales *practical experiences* ($r = 0.328$, $p = 0.008$), *positive emotions* ($r = 0.262$, $p = 0.036$), *competencies* ($r = 0.256$, $p = 0.041$), *self-efficacy* ($r = 0.301$, $p = 0.016$), *internal attributions* ($r = 0.256$, $p = 0.041$) and *strategies* ($r = 0.358$, $p = 0.004$), were found. Hence, individuals who have a lot of practical experience in dealing with computer technology, have fun with computer technology, consider themselves to be competent in handling computer devices, have a high degree of computer self-efficacy, and have positive computer-related learning strategies, rather attribute success to the system (external) than to their own abilities (internal). Furthermore, there is a positive correlation

Table 7. Correlations (Spearmen's Rho) between computer-related self-concept subscales (CSC) and computer-related attributional dimensions in situations of failure and success.

CSC		Failure				Success			
		LOC	STA	CON	GLO	LOC	STA	CON	GLO
PX	r	**0.366**	−0.037	−0.012	−0.201	**0.328**	−0.014	0.212	0.185
	p	**0.003**	0.772	0.923	0.111	**0.008**	0.910	0.093	0.144
PE	r	**0.389**	0.003	0.137	−0.105	**0.262**	0.182	0.195	0.123
	p	**0.001**	0.983	0.280	0.407	**0.036**	0.151	0.122	0.332
CA	r	**0.384**	−0.138	0.018	0.008	0.059	0.087	0.154	0.134
	p	**0.002**	0.277	0.886	0.949	0.645	0.493	0.225	0.290
UD	r	**0.294**	−0.059	0.010	−0.078	0.174	0.044	0.125	0.085
	p	**0.018**	0.642	0.937	0.540	0.169	0.728	0.324	0.507
CR	r	**0.278**	0.084	0.059	−0.099	0.194	0.064	**0.274**	−0.012
	p	**0.026**	0.508	0.643	0.437	0.125	0.613	**0.029**	0.924
TP	r	**0.263**	−0.032	0.126	−0.192	0.010	0.111	0.201	0.014
	p	**0.035**	0.801	0.323	0.129	0.935	0.380	0.111	0.910
CT	r	**0.622**	0.115	0.130	−0.201	**0.256**	0.110	0.152	0.022
	p	**0.000**	0.367	0.307	0.111	**0.041**	0.387	0.232	0.864
SE	r	**0.524**	0.034	0.065	−0.075	**0.301**	0.167	0.031	0.148
	p	**0.000**	0.788	0.609	0.553	**0.016**	0.188	0.808	0.244
IA	r	**0.554**	0.080	**0.265**	**−0.304**	**0.256**	0.133	0.208	−0.040
	p	**0.000**	0.529	**0.034**	**0.015**	**0.041**	0.295	0.099	0.752
EC	r	**0.314**	−0.055	0.030	−0.226	0.172	0.073	0.208	0.055
	p	**0.012**	0.664	0.817	0.072	0.174	0.568	0.100	0.665
ST	r	**0.468**	0.132	−0.002	0.046	**0.358**	0.117	−0.093	**0.273**
	p	**0.000**	0.299	0.989	0.717	**0.004**	0.358	0.465	**0.029**
TS	r	**0.550**	0.021	0.106	−0.163	**0.266**	0.114	0.202	0.101
	p	**0.000**	0.869	0.406	0.197	**0.034**	0.369	0.109	0.426

Note: Bold font indicates significant differences with p < 0.05; PX = practical experiences, PE = positive emotions, CA = computer anxiety, UD = understanding, CR = creating, TP = tool perspective, CT = competencies, SE = self-efficacy, IA = internal attributions, EC = external control beliefs, ST = strategies, TS = total scale, LOC = locus, STA = stability, CON = controllability, GLO = globality.

between the self-concept scale *creating* and the attributional dimension of *controllability* ($r = 0.274$, $p = 0.029$). Accordingly, people who like to use computers as a design tool tend to perceive the cause of success as uncontrollable. In addition, a positive correlation between the self-concept scale *strategies* and *globality* emerged ($r = 0.273$, $p = 0.029$). People who believe that the success refers to a global cause tend to explore systems more intuitively than persons attributing the cause to a specific situation.

3.4 Relationship Between Self-concept and Attributional Patterns

Attributional patterns and self-concept scores were tested globally for differences followed by post-hoc tests for pairwise comparison. Because of partially non-normally distributed data Kruskal-Wallis test was used instead of analyses of variance.

Regarding *situations of failure*, tests yielded significant differences for the total self-concept score ($p = 0.001$) as well as for the subscales *practical experiences* ($p = 0.033$), *positive emotions* ($p = 0.024$), *computer anxiety* ($p = 0.001$), *competencies* ($p = 0.000$), *self-efficacy* ($p = 0.001$), *internal attributions* ($p = 0.001$), *external control beliefs* ($p = 0.043$), and *strategies* ($p = 0.005$). Merely for *understanding*, *creating*, and *tool perspective* no differences were found. In situations of success, tests showed no significant differences for any of the self-concept scales (Table 8 and Fig. 1).

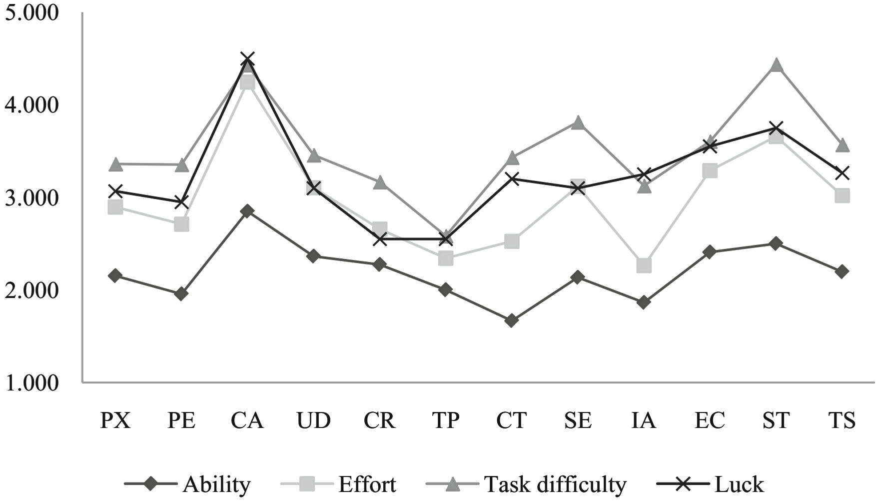


Fig. 1. Mean values of computer-related self-concept scale and subscales for each attributional pattern in situations of failure. *Note: PX = practical experiences, PE = positive emotions, CA = computer anxiety, UD = understanding, CR = creating, TP = tool perspective, CT = competencies, SE = self-efficacy, IA = internal attributions, EC = external control beliefs, ST = strategies, TS = total scale.*

Post-hoc tests were calculated to identify the relations between the individual attribution patterns and self-concept scales. In *situations of failure*, the analysis showed significant differences between the attributional patterns and the self-concept scales (Table 9). Results for *situations of success* are not reported since prior Kruskal-Wallis tests showed no significant differences.

In all, individuals that attributed failures to the *ability* pattern showed the lowest total self-concept score ($M = 2.195$) while individuals that attributed failures to *task difficulty* showed the highest score ($M = 3.569$).

Table 8. Relations between self-concept total scale as well as subscales (CSC) and attributional patterns in situations of failure and success – results of Kruskal-Wallis tests.

Self-concept scales	Failure		Success	
	Chi^2	p	Chi^2	p
Practical experiences	8.718	**0.033**	5.368	0.147
Positive emotions	9.430	**0.024**	5.015	0.171
Computer anxiety	16.245	**0.001**	0.296	0.961
Understanding	5.335	0.149	1.352	0.717
Creating	3.931	0.269	1.898	0.594
Tool perspective	2.201	0.532	0.961	0.811
Competencies	19.590	**0.000**	3.085	0.379
Self-efficacy	16.123	**0.001**	4.322	0.229
Internal attributions	15.473	**0.001**	5.071	0.167
External control beliefs	8.176	**0.043**	0.536	0.911
Strategies	12.874	**0.005**	3.544	0.315
Total scale	16.525	**0.001**	3.095	0.377

Note: Bold font means significant difference with $p < 0.05$; $df = 3$.

Regarding the self-concept subscales in *situations of failure*, the following significant differences were found: Individuals who attributed their performance to *ability* ($M = 2.152$) scored their *practical experience* significantly lower than those who attributed their performance to *task difficulty* ($M = 3.361$) or *luck* ($M = 3.067$). They also showed fewer *positive emotions* towards computer technology than those who attributed their performance to *task difficulty* ($M = 1.955$ vs. $M = 3.354$).

Moreover, individuals who attributed their performance to *ability* ($M = 2.849$) showed higher values of *computer anxiety* than those who attributed their performance to *effort* ($M = 4.246$), *task difficulty* ($M = 4.431$) or luck ($M = 4.500$).

Individuals who attributed their performance to *ability* ($M = 1.667$) also considered themselves to have fewer computer *competencies* than those who attributed their performance to *effort* ($M = 2.526$), *task difficulty* ($M = 3.431$) or luck ($M = 3.200$). Furthermore, individuals who attributed their performance to *effort* ($M = 2.526$) consider themselves to have fewer computer *competencies* than those who attributed their performance to *task difficulty* ($M = 3.431$).

Individuals who attributed their performance to *ability* ($M = 2.136$) showed lower values of *self-efficacy* than those who attributed their performance to *effort* ($M = 3.118$), *task difficulty* ($M = 3.813$) or *luck* ($M = 3.100$). Moreover, individuals who attributed their performance to *effort* ($M = 3.118$) showed lower values than those who attributed their performance to *task difficulty* ($M = 3.813$) and those, in turn, showed lower values than individuals who attributed their performance to *luck* ($M = 3.100$).

Individuals who attributed their performance to *ability* ($M = 1.864$) or *effort* ($M = 2.263$) showed lower values of *internal attributions* than those who attributed their performance to *task difficulty* ($M = 3.125$) or *luck* ($M = 3.250$). This is

Table 9. Relations between computer-related self-concept total scale as well as subscales (CSC) and computer-related attributional patterns in situations of failure – Post-hoc test (LSD).

Self-concept scales	Ability	Effort	Task difficulty	Luck
Practical experiences	**2.152**^{a,b}	2.895	**3.361**^{a}	**3.067**^{b}
Positive emotions	**1.955**^{c}	2.711	**3.354**^{c}	2.950
Computer anxiety	**2.849**^{d,e,f}	**4.246**^{d}	**4.431**^{e}	**4.500**^{f}
Understanding	2.364$^{(g)}$	3.105	3.458$^{(g)}$	3.100
Creating	2.273	2.658	3.167	2.550
Tool perspective	2.000	2.342	2.583	2.550
Competencies	**1.667**^{h,i,j}	**2.526**^{h,k}	**3.431**^{i,k}	**3.200**^{j}
Self-efficacy	**2.136**^{s,t,u}	**3.118**^{s,v}	**3.813**^{t,v,w}	**3.100**^{u,w}
Internal attributions	**1.864**^{l,m}	**2.263**^{n,o}	**3.125**^{l,n}	**3.250**^{m,o}
External control beliefs	**2.409**^{p,q,r}	**3.290**^{p}	**3.604**^{q}	**3.550**^{r}
Strategies	**2.500**^{x,y,z}	**3.658**$^{x,\alpha}$	**4.438**$^{y,\alpha-}$	**3.750**^{z}
Total scale	**2.195**^{1,2,3}	**3.020**^{1,4}	**3.569**^{2,4}	**3.263**^{3}

Note: Superscript letters show significant differences with $p < 0.05$.

theoretically plausible since lower values on this scale refer to internal attributions and this, in turn, corresponds to the attributional pattern of *ability* and *effort*. Conversely, high values refer to external attributions, which correspond to the patterns of *task difficulty* and *luck*. These findings also support the construct validity of our measurement.

Similarly, individuals who attributed their performance to *ability* ($M = 2.409$) showed lower values of *external control believe* than those who attributed their performance to *effort* ($M = 3.290$), *task difficulty* ($M = 3.604$) or *luck* ($M = 3.550$). They believe to have little control over computer problems that occur due to their abilities.

Finally, individuals who attributed their performance to *ability* ($M = 2.500$) scored their *computer-related strategies* significantly lower than those who attributed their performance to *effort* ($M = 3.658$), *task difficulty* ($M = 4.438$) or *luck* ($M = 3.750$). Moreover, individuals who attributed their performance to *effort* ($M = 3.658$) scored significantly lower than those who attributed their performance to *task difficulty* ($M = 4.438$).

4 Discussion

This study aimed to examine the relationship between computer-related self-concept and computer-related attributions in computer users. This section discusses the findings of the present study, its limitations, and offers suggestions for future research and practice.

4.1 Relationship Between Self-concept and Causal Attributions

The present study shows that users with more positive or negative computer-related self-concept differ in attributing their computer performance. Results yield significant differences especially with respect to the attributional dimension of locus. In situations of failure, similar results to other research domains were found. Users with a positive self-concept attributed failures to external reasons and tend to blame the system when something goes wrong. However, they also attribute successful outcomes to external factors. This is an interesting finding since this is contrary to the findings of a number of previous studies in other domains [34–38]. Other researchers even found the opposite in a study conducted with students in a classroom environment: Students with a positive self-concept attributed success and failure more to internal causes, while students with negative self-concept attributed success and failure more to external causes [49]. It is likely that the assigned tasks were perceived as too easy and the participants with a positive self-concept had ample experience with the respective applications or devices. On the other hand, participants with a negative self-concept may have less experience and were, therefore, proud to be able to solve the task at all. Therefore, they rather attributed the success to themselves.

This was also reflected in regard to the *attributional patterns* (ability, effort, task difficulty, and luck) but merely in situations of failure. Users that attribute causes of failures to their own *abilities* (internal/stable) showed a significantly more negative computer-related self-concept than users that attribute causes of failures to factors such as *effort*, *task difficulty*, or *luck*.

In summary, the present study found that users with positive and negative self-concept partly differ in respect to their attributions for computer-related outcomes. The findings also suggest that the relations regarding computer technology differ from the relations in other research domains.

4.2 Implications and Recommendations

The findings can be used in human-computer research and practice to understand better why users think, feel, or behave in a certain way. Thus, design principles could be developed to support different types of users in a specific way. To our knowledge, this is the first study that combines these two cognitive theories and directly examines the impact of computer-related causal attributions on users' computer-related self-concept, and vice versa.

Therefore, this study contributes to a more complete and detailed knowledge of users' computer-behavior. The results encourage further research on the relations between different cognitive theories in the field of HCI.

There are also implications for practitioners who develop and design computer systems. This study sheds light on different types of computer users who show characteristics of self-concept. In order to assist people to develop a more positive self-concept, several measures might be explored. For example, *attributional retraining* [50], which suggests that individuals' performance will increase when they learn to ascribe causes to more favorable attributions, could be a promising approach. Thus, our results are valuable for developing and improving existing computer learning training strategies and methods, as well as support and assistance mechanisms for users. Practitioners should attempt to adapt these findings and design specified systems by, for example, including attributional retraining strategies. This could be done, for example, by providing feedback that changes the beliefs of the users about the cause of computer-related outcomes (e.g., comments that contain the desired attributions). System developer and designers should bear this in mind and future research should take this into consideration.

4.3 Limitations and Future Research

The present study also faces some limitations. First, the relatively small sample size, especially regarding some subgroups, limits the generalizability of this study. This is particularly true for the attributional patterns (ability, effort, task difficulty, and luck). Future studies with larger samples would enable to delineate the relations more clearly, especially regarding group comparisons.

Despite the careful selection of the participants, the sample is relatively homogeneous. It consists predominantly of educated and experienced computer users who self-assessed their computer expertise as high. This limits the generalizability of our findings since there is some evidence that socio-demographic factors, like e.g. the educational level, have an impact on attribution processes [23]. Maybe this is also an explanation regarding the few significant differences in situations of success. To investigate the possible influence of the sample characteristics we re-ran our analyses with these factors (level of education, years of computer experience, self-assessed computer skills) as a covariate. However, the analysis showed no differences, as the variance of the values turned out to be too low. Therefore, future research will need to involve a more heterogeneous or representative sample.

The research design of this study also carried certain limitations. Standardized use situations were chosen in this study to create a similar experience for all participants. However, a drawback of this method is that the situations were somewhat artificial and unrelated to the participants' normal use habits, which might result in reduced intensity and significance of the experience (see [22] for a comparison of different data collection methods).

Furthermore, participants are from Germany only. In regard to possible intercultural differences, future studies should investigate cultural differences by expanding into a more international context.

The results presented here give first insights regarding the relation of computer-related self-concept and computer-related attributions. More research is needed to provide a rich understanding of how and to what extent these factors play a role in HCI research and practice. In this regard, it should be noted that this study has an explorative character. Therefore the results should be interpreted with caution. Nevertheless, this calls for more research to corroborate the findings.

Our next step is to investigate the relations in more detail as well as the effects of reattribution training methods on the computer-related self-concept.

References

1. Bracken, B.: Handbook of Self-concept: Developmental, Social and Clinical Considerations. Wiley, Canada (1996)
2. Burns, R.B.: Self-concept Development and Education. Holt, Rinehart and Winston, London (1982)
3. Weiner, B.: An attributional theory of achievement motivation and emotion. Psychol. Rev. **92**, 548–573 (1985). doi:10.1037/0033-295X.92.4.548
4. Heider, F.: The Psychology of Interpersonal Relations. Lawrence Erlbaum Associates, London (1958)
5. Leflot, G., Onghena, P., Colpin, H.: Teacher-child interactions: relations with children's self-concept in second grade. Infant Child Dev. **19**, 385–405 (2010). doi:10.1002/icd.672
6. Flook, L., Repetti, R.L., Ullman, J.B.: Classroom social experiences as predictors of academic performance. Dev. Psychol. **41**, 319–327 (2005). doi:10.1037/0012-1649.41.2.319
7. Shavelson, R.J., Bolus, R.: Self-concept: the interplay of theory and methods. J. Educ. Psychol. **74**, 3–17 (1982). doi:10.1037/0022-0663.74.1.3
8. Shavelson, R.J., Hubner, J.J., Stanton, G.C.: Self-concept: validation of construct interpretations. Rev. Educ. Res. **46**, 407–441 (1976). doi:10.3102/00346543046003407
9. James, W.: The principles of psychology (review). J. Hist. Philos. **21**, 270–272 (1890). doi:10.1353/hph.1983.0040
10. Cooley, C.H.: Human Nature and the Social Order. Transaction Publishers, New Brunswick (1992)
11. Coopersmith, S.: The Antecedents of Self-esteem. Consulting Psychologists Press, San Francisco (1967)
12. Marx, R.W., Winne, P.H.: Construct interpretation of three self-concept inventories. Am. Educ. Res. J. **15**, 99–109 (1978)
13. Marsh, H.W., Shavelson, R.: Self-concept: its multifaceted, hierarchical structure. Educ. Psychol. **20**, 107–123 (1985). doi:10.1207/s15326985ep2003_1
14. Hansford, B.C., Hattie, J.A.: The relationship between self and achievement/performance measures. Rev. Educ. Res. **52**, 123–142 (1982). doi:10.3102/00346543052001123
15. Byrne, B.M.: The general/academic self-concept nomological network: a review of construct validation research. Rev. Educ. Res. **54**, 427–456 (1984). doi:10.3102/00346543054003427
16. Janneck, M., Vincent-Höper, S., Ehrhardt, J.: The computer-related self concept: a gender-sensitive study. Int. J. Soc. Organ. Dyn. IT **3**, 1–16 (2013). doi:10.4018/ijsodit.2013070101
17. Rosenberg, M., Hovland, C.: Cognitive, affective, and behavioral components of attidudes. In: Attitude Organ Change an Analysis Consistency Among Attitude Components, pp. 1–14 (1960)
18. Weiner, B.: Achievement Motivation and Attribution Theory. General Learning Press, Morristown (1974)
19. Antaki, C., Brewin, C.R.: Attributions and Psychological Change: Applications of Attributional Theories to Clinical and Educational Practice. Academic Press, London (1982)
20. Weiner, B., Frieze, I., Kukla, A., Rosenbaum, R.M.: Perceiving the causes of success and failure. Dev. Psychol. **11**, 103 (1971). doi:10.1037/h0076132
21. Stiensmeier-Pelster, J., Schürmann, M., Eckert, C., Pelster, A.: Der Attributionsstil-Fragebogen für Kinder und Jugendliche (ASF-KJ): Untersuchungen zu seinen psychometrischen Eigenschaften. [Attribution style questionnaire for children and adolescents (ASF-KJ): Investigations on its psychometrical properties]. Diagnostica **40**, 329–343 (1994)

22. Niels, A., Janneck, M.: Computer-related attribution styles: typology and data collection methods. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9297, pp. 274–291. Springer, Cham (2015). doi:10.1007/978-3-319-22668-2_22
23. Niels, A., Guczka, S.R., Janneck, M.: Computer-related causal attributions: the role of sociodemographic factors. In: Procedia Manufacturing, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015, vol. 3, pp. 3448–3455 (2015)
24. Niels, A., Guczka S.R.: The impact of causal attribution s on system evaluation in usability tests. In: Proceedings of SIGCHI Conference Human Factors in Computing Systems, pp. 3115–3125 (2016). doi:10.1145/2858036.2858471
25. Hunter, M., Barker, G.: “If at first…”: attribution theory in the classroom. Educ. Leadersh. **8**, 50–53 (1987)
26. Solley, C.M., Stagner, R.: Effects of magnitude of temporal barriers, type of goal and perception of self. J. Exp. Psychol. **51**, 62–70 (1956). doi:10.1037/h0049139
27. Ames, C.: Children’s achievement attributions and self-reinforcement: effects of self-concept and competitive reward structure. J. Educ. Psychol. **70**, 345–355 (1978). doi:10.1037//0022-0663.70.3.345
28. Fitch, G.: Effects of self-esteem, perceived performance, and choice on causal attributions. J. Pers. Soc. Psychol. **16**, 311–315 (1970). doi:10.1037/h0029847
29. Ickes, W., Layden, A.: Attribution styles. In: Harvey, J.H., Ickes, W., Kidd, R.F. (eds.) New Directions in Attribution Research, p. 420. Lawrence Erlbaum Associates, Hillsdale (1978)
30. Weiner, B.: A theory of motivation for some classroom experiences. J. Educ. Psychol. **71**, 3–25 (1979). doi:10.1037/0022-0663.71.1.3
31. Burns, R.B.: The Self concept in Theory, Measurement, Development, and Behaviour. Longman Group Ltd., London (1979)
32. Hattie, J.: Self-concept. Lawrence Erlbaum Associates, Hillsdale (1992)
33. Marsh, H.W., Relich, J.D., Smith, I.D., Marsh, H.: Self-concept: the construct validity of interpretations based upon the SDQ. J. Pers. Soc. Psychol. **45**, 173–187 (1983). doi:10.1037/0022-3514.45.1.173
34. Meyer, W.-U.: Das Konzept von der eigenen Begabung [The concept of one’s own ability]. Huber, Bern (1984)
35. Schwarzer, R., Jerusalem, M.: Selbstwertdienliche Attributionen nach Leistungsrückmeldungen. [Self-serving attributions after performance feedback.]. Z Entwicklungspsychol Padagog Psychol **14**, 47–57 (1982)
36. Pekrun, R.: A social-cognitive, control-value theory of achievement emotions. Adv. Psychol. **131**, 143–163 (2000). doi:10.1016/S0166-4115(00)80010-2
37. Moller, J., Koller, O.: Spontaneous and reactive attributions following academic achievement. Soc. Psychol. Educ. **4**, 67–86 (2000). doi:10.1023/A:1009690303176
38. Faber, G.: Academic self-beliefs in the spelling domain: empirical research findings on elementary school students’ subject-specific self-concept, causal attributions, and test anxiety. In: Vargios, E.M. (ed.) Educational Psychology Research Focus, pp. 65–120. Nova Sciece Publishers, Inc., New York (2007)
39. Försterling, F., Harrow, J.T.: Attribution Theory in Clinical Psychology. Wiley, Oxford (1988)
40. Grünke, M., Castello, A.: Attributionstraining [Attribution training]. In: Lauth, G.W., Grünke, M., Brunstein, J.C. (eds.) Interv. bei Lernstörungen. Förderung, Train. und Ther. der Prax., 2nd edn., pp. 484–492. Hogrefe Verlag (2014)
41. Heller, K.A.: Attributional Retraining as an Attempt to Reduce Gender-Specific Problems in Mathematics and the Sciences. LIT Verlag, Berlin (2010)

42. Marsh, H.W., Byrne, B.M., Shavelson, R.J.: A multifaceted academic self-concept: its hierarchical structure and its relation to academic achievement. J. Educ. Psychol. **80**, 366–380 (1988). doi:10.1037/0022-0663.80.3.366
43. Kelley, H., Compeau, D., Higgins, C.A., Parent, M.: Advancing theory through the conceptualization and development of causal attributions for computer performance histories. ACM SIGMIS Database **44**, 8 (2013). doi:10.1145/2516955.2516957
44. Guczka, S.R., Janneck, M.: Erfassung von Attributionsstilen in der MCI – eine empirische Annäherung [Investigation of attribution styles in HCI - an empirical approach]. In: Reiterer, H., Deussen, O. (Hrsg.), Mensch Comput. 2012 Interakt. Inf. – allgegenwärtig und allumfassend!? pp. 223–232 (2012)
45. Hanrahan, S.J., Grove, J.R., Hattie, J.A.: Development of a questionnaire measure of sport-related attributional style. Int. J. Sport Psychol. **20**, 114–134 (1989)
46. Rogers, C.: Client Centred Therapy: Its Current Practice, Implications and Theory. Houghton Mifflin, Boston (1951)
47. Ames, C., Felker, D.W.: Effects of self-concept on children's causal attributions and self-reinforcement. J. Educ. Psychol. **71**, 613–619 (1979). doi:10.1037/0022-0663.71.5.613
48. Elrod, M.M., Crase, S.J.: Sex differences in self-esteem and parental behavior. Psychol. Rep. **46**(3), 719–727 (1980)
49. Madonna, S., Bailey, G.K., Wesley, A.L.: Classroom environment and locus of control in identifying. Psychol. Rep. **66**, 1152–1154 (1990)
50. Försterling, F.: Attributional retraining: a review. Psychol. Bull. **98**, 495–512 (1985). doi:10.1037/0033-2909.98.3.495

Greater Heart Rate Responses to Acute Stress is Correlated with Worse Performance of Visual Search in Special Police Cadets

Xiaofang Sun[1,2], Yi Yuan[3], Zhuxi Yao[1,2], Kan Zhang[1],
and Jianhui Wu[1,4](✉)

[1] Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
sunxf@psych.ac.cn, wujh8@szu.edu.cn
[2] University of the Chinese Academy of Sciences, Beijing, China
[3] Special Police Academy, Beijing, China
[4] College of Psychology and Sociology, Shenzhen University, Shenzhen, China

Abstract. Special police often need search targets under stress situation. In this study we investigated the relationship between the autonomic stress response and performance of visual search. Eighty-two male special police cadets were randomly assigned to walk on an aerial robe ladder to induce stress response or walk on a cushion on the ground as control condition. And then participants were asked to completed two visual search tasks including detecting targets (experiment 1) and detecting and identifying targets (expriment 2). We found a negative association between the heart rate of treatment stage and accuracy of the identification task for the participants who were under stress condition. This result suggested that greater autonomic stress response are related with worse performance of visual search in professional working crowed such as special police.

Keywords: Acute stress · Heart rate · Visual search · Special police cadets

1 Introduction

Visual search is a fundamental activity in our daily life. People might look for a car in a crowed parking lot, or search a name in a long list. In some special populations, people who work on public security, such as searching criminal suspects from surveillance video or picking up the target by radar, often do their works under high stress. These professional works require people to have both an appropriate physiological response and behavioral adjustment to meet the demands of emergencies.

1.1 Visual Search

Visual search is defined as the ability to find one item in a visual world filled with distracting items [1]. In the laboratory, the most common paradigm of visual search requires participants to detect whether the target is present or absent [2]. And some researchers argued that the more common visual search task in real world is to identify

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 200–208, 2017.
DOI: 10.1007/978-3-319-58472-0_16

the target with specific feature in crowed distractors such as looking for your car in crowded parking lot. In this situation, people are required to detect the target and then determine its identification [3]. Both the two paradigms are used pervasively in attention research. But researchers had little agreement with that whether there are different cognitive processes under the two tasks [5]. Horowitz and Wolfe found that when people were asked to detect a target among distractors, their visual system did not accumulate the information of the target identity [4]. Cameron et al. found that accuracy of identification was higher compare to detection [9]. However, Saarinen et al. found that the reaction time in identification task was slower than the reaction time in detection task [10].

1.2 The Relationship Between Acute Stress and Cognitive Functions

Exposure to acute stress leads to both a rapid activation of the sympathetic nervous system (SNS) and a slow activation of the hypothalamic-pituitary-adrenal (HPA) axis [12]. The activation of SNS leads to rapidly increasing heart rate, and the physiological responses are intimately correlated to behaviors and brain under stress [8, 13].

Studies found that many cognitive functions including working memory and executive control were influenced by acute stress, some results found the facilitated effect [6–8], for example, Yao et al. found that in acute stress condition, the stronger autonomic physiological responses was positively related with better post-error adjustment [8]. However other studies found stress attenuated cognitive function [11, 12], Plessow et al. found that acute psychosocial stress attenuated cognitive flexibility [11].

1.3 The Relationship Between Stress and Visual Search

Some previous studies focused on the effect of stress related emotion such as anxiety on visual search. Murray and Janelle found that reaction time of visual search task reduced in the low-anxious group but increased in the high-anxious group, but in this study, participants were grouped based on trait anxious [14]. Williams et al. found that anxiety which was induced by competition and prize money impaired performance especially the task imposed a heavy demand of working memory [15]. But few studies researched about the direct effect of acute stress on visual search processing and performance.

The aim of present study was to investigate the relationship between the acute stress and the performance of visual search including the detection paradigm and identification paradigm. Participants were male special police cadets. They were induced stress and then completed two visual search tasks. To induce acute stress, the participants of stress group were asked to walk on an aerial rope ladder bridge. And the participants of control group were required to walk on a cushion on the floor for the same distance as the stress group. We used a wireless chest HR transmitter and wrist monitor to recorder heart rate of participants.

2 Method

2.1 Participants and the Exclusion Criteria

Eighty-two healthy special police students participated this study, they were all male and aged 20 to 27 (mean = 22.47, SD = 1.25). They were randomly assigned to stress condition or control condition, 59 participants were in stress condition and the rest 23 were in control condition.

The criteria for exclusion were as follows: (1) smoking more than 5 cigarettes a day or alcohol abuse, (2) presence of cardiovascular, endocrine, neurological, or psychiatric disease, and other chronic major diseases. (3) Irregular sleeping patterns, (4) took medicine or had a cold two weeks before the study. And all participants were asked to not to do strenuous exercise and rest properly the day before the experiments.

2.2 General Procedure

The study run in the afternoon from 14:30 to 18:30 in order to control for the circadian fluctuation. Participants were given a brief introduction to the experiment in group sessions. Then they were required to fill up forms and questionnaires including informed consent, demographic information and personality questionnaires. After completing these forms, participants were asked to seat and be relaxed for 30 min, and then baseline heart rate data were recorded by a wireless chest rate transmitter and wrist monitor recorder (Polar RSC800CX, Polar Electro, Finland). Then, the participants were randomly assigned to the stress or control condition, at that time, participants had known that he would experience stress condition (see below) or control condition, recorded their heart rate of expect. And then, participants were required to do the stress treatment or control condition, during which heart rate data were continuously recorded. After completing the treatment, they were asked to complete the visual search task immediately in computers. At last recorded the posttreatment heart rate after the cognitive tasks.

2.3 Stress Induction

In the stress condition, the participants in the stress group were required to walk on an aerial rope ladder bridge to induce acute stress. All the participants were experienced such training for the first time. The aerial rope ladder which bridged between two buildings was 18 m above ground, and 12 m in length. Participants in stress condition were asked to wear harness and then walk on the aerial for 8 m (there was a mark on the two-thirds of the ladder), and stayed here and looked down for 5 s, and then turned back (16 m in total).

For the control condition, participants were asked to walk on a cushion on the floor for the same distance in a straight line and to turn around at the midpoint.

2.4 Visual Search Tasks

Based on previous study [3], two experiments were conducted. Both two experiments included 8 practice trails and 64 experimental trials. Figure 1 illustrates the sequence of events in a single trial of experiment 1. The trial began with a premask that stayed on for 600 ms, and was followed by the presentation of the search array that terminated with a response, or remained on the screen for maximum of 5,010 ms. After the presentation of the search stimuli, a postmask was presented for 600 ms, followed by feedback presented at the centre of the screen for 900 ms. The intertrial interval was 550 ms, during which time the screen was blank. On each trial, stimuli were black L and T on a white background, the target was a T, and the distractors were Ls. Both targets and distractors were in for orientations (0°, 90°, 180° or 270°). And the set size

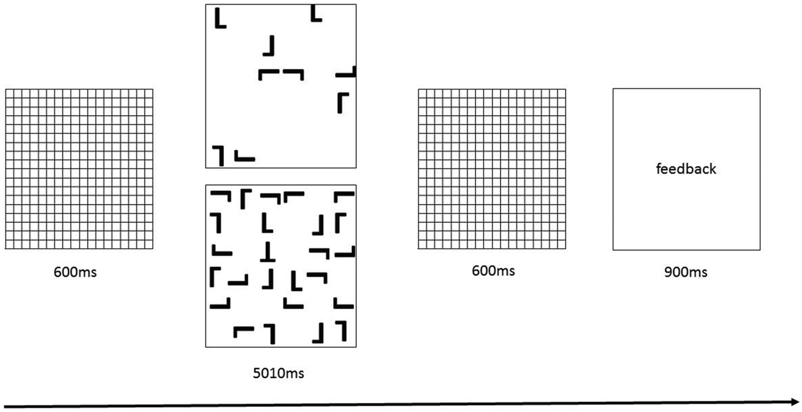


Fig. 1. An example of a trial of detection task. The stimulus on the top was a sample has 9 items without target, the other one was a sample has 25 items with target.

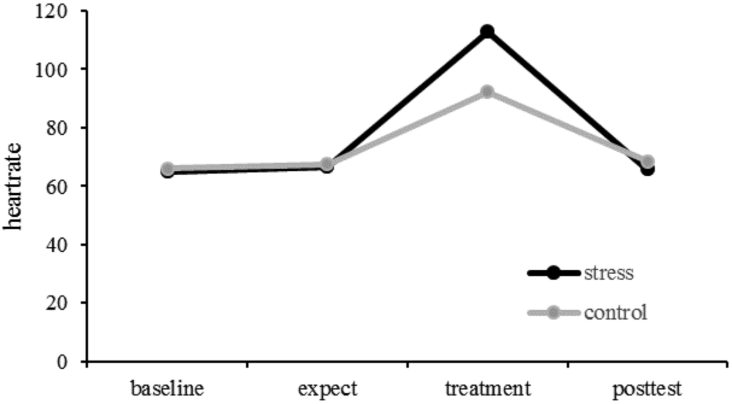


Fig. 2. Mean heart rate at prepare (baseline), expect (be introduced the stress task), during and after the acute stress and control condition.

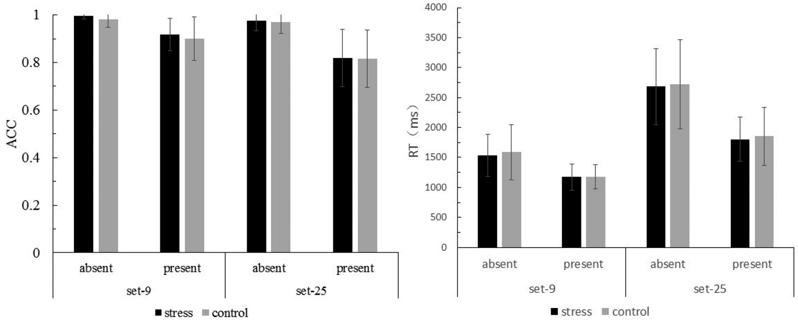


Fig. 3. The left figure showed accuracy (ACC) of experiment 1. And the right figure was mean reaction time (RT) of experiment 1. All the error bars were standard error.

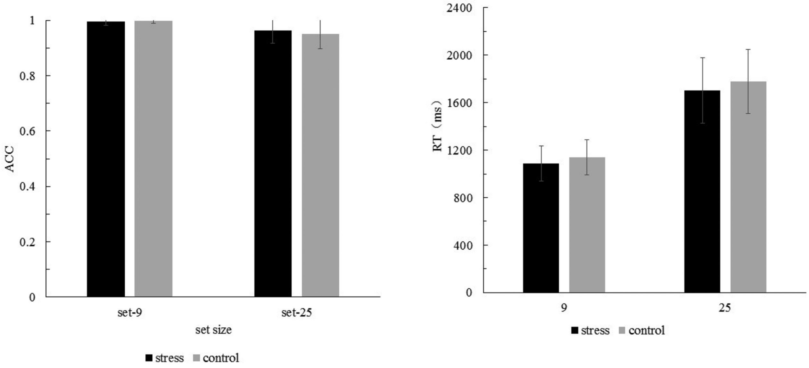


Fig. 4. The left figure showed accuracy(ACC) of experiment 2, and the right figure was mean reaction time(RT) of experiment 2. All the error bars were standard error.

of stimuli was 9 items or 25 items. Participants were asked to detect whether the target ‘T’ was present or not among distractors ‘L’. If the target was present, pressed left arrow, and if the target was absent, pressed right arrow.

In experiment 2, the procedure of single trial and stimuli were similar to experiment 1, but target T was present on all trails. Participants were asked to detect the target and determined the orientation of targets by pressing four corresponding arrow keys.

3 Results

3.1 Stress Effect

The repeated measures ANOVA, 2 (stress/control) × 4 (baseline/expect/treatment/posttest), revealed significant main effects in groups, F(3, 240) = 646.05, $p < 0.05$, $\eta^2 = 0.89$. And a main effect between groups was also significant, F(1, 80) = 5.79,

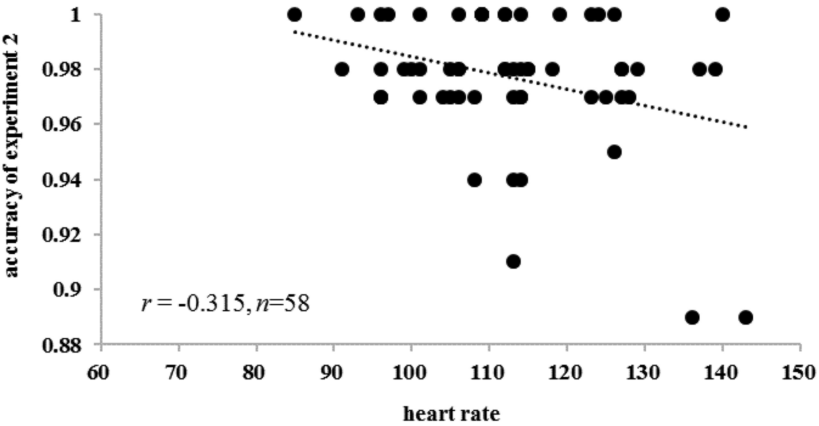


Fig. 5. The scatter plots of simple correlation analysis show the negative correlation between the heart rate and the accuracy of experiment 2 for participants of stress condition.

Table 1. Correlation between heart rate of treatment and accuracy and reaction time of participants in stress condition in both experiment 1 and 2.

	ACC		RT	
	r	*p*	*r*	*p*
Experiment 1	0.017	0.9	0.014	0.913
Experiment 2	−0.315*	<0.05	0.03	0.819

*p < 0.05; ns = not significant.

$p < 0.05$, $\eta^2 = 0.068$. The further simple effect analysis revealed that (1) the heart rate in treatment stage was higher than the other stages for both stress condition and control condition, (2) in treatment stage, heart rate of the participants under stress condition was higher than that of control condition ($p < 0.05$), (3) in the other stages included baseline, expect and posttest, no significant differences were found between stress and control condition (Fig. 2).

3.2 Data Analysis

In experiment 1, the repeated measures ANOVA of accuracy of 2 (stress/control) × 2 (set size:9/25) × 2 (absent/present) revealed significant main effects of set size and absent/present of target, $F_{set\ size}(1, 80) = 578.12$, $p < 0.05$, $\eta^2 = 0.88$; $F_{present}(1, 80) = 157.80$, $p < 0.05$, $\eta^2 = 0.66$. No significant difference was found between stress condition and control condition, $F(1, 80) = 1.16$, $p = 0.29$, $\eta^2 = 0.014$. The interaction effect of set size and present was significant, $F_{size \times present}(1, 80) = 17.77$, $p < 0.05$, $\eta^2 = 0.18$. Other interaction effects were not found.

The results of reaction time was similar. Both effects of set size and present were significant. $F_{set\ size}(1, 80) = 578.15$, $p < 0.05$, $\eta^2 = 0.88$; $F_{present}(1, 80) = 329.18$,

$p < 0.05$, $\eta^2 = 0.80$. The interaction effect of set size and present or absent was significant, $F_{size \times present}(1, 80) = 110.78$, $p < 0.05$, $\eta^2 = 0.58$. the difference between stress condition and control condition was not significant $F(1, 80) = 0.15$, $p = 0.70$, $\eta^2 = 0.02$ (Fig. 3).

In experiment 2, the repeated measures ANOVA of 2(stress/control) × 2(set size:9/25) revealed the main effect of set size were significant for both accuracy and reaction time, $F_{ACC}(1, 80) = 50.36$, $p < 0.05$, $\eta^2 = 0.39$, $F_{RT}(1, 80) = 616.98$, $p < 0.05$, $\eta^2 = 0.89$. But no differences between stress condition and control condition were significant for both accuracy and reaction time (Fig. 4).

Regarding the relationship between heart rate and the performance of visual search tasks, the correlation analysis between heart rate of treatment stage and both ACC (accuracy) and RT(reaction time) of two experiments revealed that for the participants of stress condition, a significantly negative correlation between HR and the ACC of experiment 2 ($r = -0.315$, $p < 0.05$) was found (Table 1 and Fig. 5).

The results of present study demonstrated that the physiologically autonomic response induced by acute stress was related with the identification stage of visual search, but not related with the detection stage. And lower heart rate under stress condition implicated response was related with better performance of visual search. Horowitz and Wolfe [4] argued that the visual system does not accumulate information about identity of targets. The identification task required participants to pay more mental resource which is limited under stress.

4 General Discussion

In this study we investigated the relationship between physiological responses to acute stress and the performance of visual search in special professional people such as special police. We asked participants to walk aloft as a stressor to induce an acute stress response. These participants were induced higher heart rate than participants under control condition. And higher heart rate was negatively correlated with accuracy of targets' identification. This result suggests that stronger autonomic stress responses are related to worse performance of identification stage of visual search.

This negative association between physiological response under acute stress and behavioral performance of visual search may suggest that the stronger responses for stress impair the cognitive processing of visual search. The current result was consistent with previous studies. Some studies found that acute stress attenuated cognitive functions such as executive function and working memory [11, 12]. And this result was also consistent with the study about the attenuated effect of anxiety on visual search [14, 15].

The mechanism underlying this result may be that higher heart rate reduces the cognitive resource which was used to detecting and recognizing targets. And stronger responses of acute stress may change the strategy of visual search or trade off. Another explanation of the underlying mechanism may be some common factor effects both heart rate and the performance of task. Previous study found that people trained would have smaller heart rate change under stress, this is to say, people who trained may have a smaller physiological responses under stress [16, 17].

Some limitations of this study have to be acknowledged. First, participants in this study were special police cadets. Further studies should test other special professional group such as security check inspectors or air traffic controllers. Second, the stress intensity of our study was relatively moderate, this may be the reason that only the accuracy of identification task was effected attenuated by stress.

The findings of the present study implicate that stress influenced on certain stage of visual search. This point is very helpful to making more appropriate direction to improve the interface between human and computer or other devices. And the results also provided suggestions about employee selection of high-stress job, such as special police, air traffic controller and so on. Physiological responses to stress can be a predictor of the job candidates' work efficiency in emergency.

References

1. Wolfe, J.M.: Guided search 2.0 a revised model of visual search. Psychon. Bull. Rev. **1**(2), 202–238 (1994)
2. Palmer, J.: Attention in visual search: distinguishing four causes of a set-size effect. Curr. Dir. Psychol. Sci. **4**(4), 118–123 (1995)
3. Dukewich, K.R., Klein, R.M.: Finding the target in search tasks using detection, localization, and identification responses. Can. J. Exp. Psychol./Revue canadienne de psychologie expérimentale **63**(1), 1 (2009)
4. Horowitz, T.S., Wolfe, J.M.: Visual search has no memory. Nature **394**(6693), 575–577 (1998)
5. Busey, T., Palmer, J.: Set-size effects for identification versus localization depend on the visual search task. J. Exp. Psychol. Hum. Percept. Perform. **34**(4), 790 (2008)
6. Hockey, G., Robert, J.: Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol. **45**(1), 73–93 (1997)
7. Henderson, R.K., Snyder, H.R., Gupta, T., Banich, M.T.: When does stress help or harm? The effects of stress controllability and subjective stress response on stroop performance. Front. Psychol. **3**, 1–15 (2012)
8. Yao, Z., et al.: Greater heart rate responses to acute stress are associated with better post-error adjustment in special police cadets. PloS One **11**(7), e0159322 (2016)
9. Cameron, E.L., et al.: Signal detection theory applied to three visual search tasks—identification, yes/no detection and localization. Spat. Vis. **17**(4), 295–325 (2004)
10. Saarinen, J., Vanni, S., Hari, R.: Human cortical-evoked fields during detection, localisation, and identification of 'pop-out' targets. Perception **27**(2), 215–224 (1998)
11. Plessow, F., et al.: Inflexibly focused under stress: acute psychosocial stress increases shielding of action goals at the expense of reduced cognitive flexibility with increasing time lag to the stressor. J. Cogn. Neurosci. **23**(11), 3218–3227 (2011)
12. Qin, S., et al.: Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. Biol. Psychiatry **66**(1), 25–32 (2009)
13. Ulrich-Lai, Y.M., Herman, J.P.: Neural regulation of endocrine and autonomic stress responses. Nat. Rev. Neurosci. **10**(6), 397–409 (2009)
14. Murray, N.P., Janelle, C.M.: Anxiety and performance: a visual search examination of the processing efficiency theory. J. Sport Exerc. Psychol. **25**(2), 171–187 (2003)

15. Mark Williams, A., Vickers, J., Rodrigues, S.: The effects of anxiety on visual search, movement kinematics, and performance in table tennis: a test of Eysenck and Calvo's processing efficiency theory. J. Sport Exerc. Psychol. **23**(4), 438–455 (2001)
16. Dishman, R.K., et al.: Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. Int. J. Psychophysiol. **37**(2), 121–133 (2000)
17. Rimmele, U., et al.: Trained men show lower cortisol, heart rate and psychological responses to psychosocial stress compared with untrained men. Psychoneuroendocrinology **32**(6), 627–635 (2007)

On-time Measurement of Subjective Anxiety of a Passenger in an Autonomous Vehicle: Gradually Changing Sounds Decreases Anxiety of Passenger

Akitoshi Tomita(✉), Etsuko T. Harada, Satoshi Ando,
Kozue Miyashiro, Maito Ohmori, and Hiroaki Yano

University of Tsukuba, Tsukuba, Ibaraki, Japan
tomita.akitoshi.ge@u.tsukuba.ac.jp

Abstract. The current study examined the possibility of measuring the subjective anxiety in real-time by means of a novel handle-shaped device, caused by riding an autonomous car. In our experiment, a participant was shown computer graphics (CG) animation, which gave the person a virtual experience of riding a new autonomous car. The CG animation stimuli were made with three variables: maximum speed (19 km/h, 160 km/h, or 320 km/h), acceleration/deceleration pattern (linear or exponential), and with and without ascending/descending sounds (sound, or no sound). The participants grasped the handle and moved it in the longitudinal direction, i.e., pulling when they experienced anxiety and pushing when they felt relaxed. Results of experiments by 16 participants showed that they moved the handle depending on the stimulus of speed at that instant, which indicated that our handle-shaped device was useful in assessing the participants' anxiety on time. In addition, results indicated that sounds, especially those which gradually ascending with acceleration, could diminish the subjective anxiety under some conditions.

Keywords: Subjective anxiety · Autonomous car · Psychological experiment with virtual reality

1 Background

Nowadays, numerous machineries are being altered to enable them to operate autonomously using the rapid developments in information technology and artificial intelligence. The same is true for automobiles. For example, TESLA Inc., USA, has manufactured a full self-driving system. However, from a user perspective, some people have worries or doubts regarding the safety and reliability of self-driving cars, which may lead to anxiety while using such cars. It can cause worries not only to drivers but also to passengers, when those autonomous cars are used in the public transportation system. Especially passengers usually do not have sufficient knowledge about such technologies and the specific risks associated with them. As Peyre et al. [1] indicated, the question regarding whether riding automatically driven vehicles as a passenger can cause anxiety or not is an important issue [1].

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 209–219, 2017.
DOI: 10.1007/978-3-319-58472-0_17

Anxiety is a difficult concept to assess and it is often measured on a five-to seven-point scale in response to questions on subjective feelings of anxiety, e.g., the State-Trait Anxiety Inventory (STAI) [2]. However, such scales cannot capture anxiety as a real time index because of limitations related to the non-continuous style of data sampling and its retrospective reporting. Subjective anxiety has also been assessed by researches using biometric indexes, especially cardiac wave analysis, i.e., indexes of the parasympathetic nerves (high frequency cardiac waves: HF) and sympathetic nerves (low frequency cardiac waves divided by HF: LF/HF) [3, 4]. These methods can measure some aspects of anxiety continuously. However, it is hard to capture these indexes as direct values of the anxiety level because such biomedical indexes reflect different physiological states and not just anxiety. Additionally, there are latencies between subjective feelings of anxiety and the measurement of these indexes.

In this study, we have attempted to measure the subjective feeling of anxiety continuously from the time before vehicular movement starts to the time when the destination is reached using a new haptic device, which can slide forward and backward. It is possible to assess real-time subjective feelings, anxiety in this case, directly while a participant experiences the stimulus of riding an autonomous car. In order to test its effectiveness, we executed an experiment using computer graphics (CG) delivered through a head-mounted display to measure the anxiety, which a participant experienced while virtually riding in a new automated vehicle. To determine their sensitivity, the maximum speed of the vehicle was varied from 19 km/h to 320 km/h and patterns of acceleration/deceleration were introduced.

Another simulation variable was introduced during the experiment, which was aimed at decreasing the passenger's anxiety. We used sound that varied according to the change in speed of the vehicle as a possible way to decrease the passenger anxiety in autonomous cars. The sound was in ascending/descending musical scale, in which the tone changed gradually in accordance with the speed of the train (as an example: Keihin electric Express Railway Co. Ltd, Japan). In our experiment, gradually changing sounds were presented for 6 s at the beginning of acceleration and another 6 s before stopping. The introduction of these sounds may be able to eliminate or diminish passenger anxiety because they can help the passenger to predict speed changes.

If this new method, which uses a handle-like device to ask and record subjective feelings of anxiety continuously, is effective for measuring anxiety as well as other feelings like fun and excitement, it can be useful in evaluating the *design of motion control* in autonomous cars in the future. We executed a psychological experiment to determine whether this new method would be able to measure passenger anxiety and to see the effectiveness of sound in decreasing passenger anxiety. This experiment is the first of its kind to evaluate the subjective anxiety of a passenger continuously.

2 Methods

2.1 Experimental Design

In order to assess the effectiveness of our new method in capturing the participant's anxiety, we set three conditions of maximum speed (19 km/h, 160 km/h and 320 km/h) and two accelerate/decelerate patterns (linear and exponential). Those speeds were chosen from data obtained from pilot experiments, which were executed to determine the velocity needed for a participant to feel some anxieties with the CG simulation set used in our study.

In addition, experiments were performed with and without sound. Totally, there were 3 (levels of speed) × 2 (speed change patterns) × 2 (with and without sound) conditions, all of which were within-subject factors. This resulted in twelve experimental conditions. Each one in those twelve conditions were presented three times, thus participants had thirty-six trials, which were presented in random orders.

2.2 Stimuli and Devices

We created the handle device for measuring the participants' subjective feelings of anxiety using a 3D printer. The handle device consisted of a T-shaped handle, which could move in the longitudinal direction and had a rectangular-shaped base (Fig. 1). The device included elastic that held the handle at the center. The more the handle was pushed (or pulled), the more was the reaction force to return to the center. This device was connected to a computer that tracked handle position at all times during the experiment. This data was used to capture the conditions under which the participants experienced anxiety.

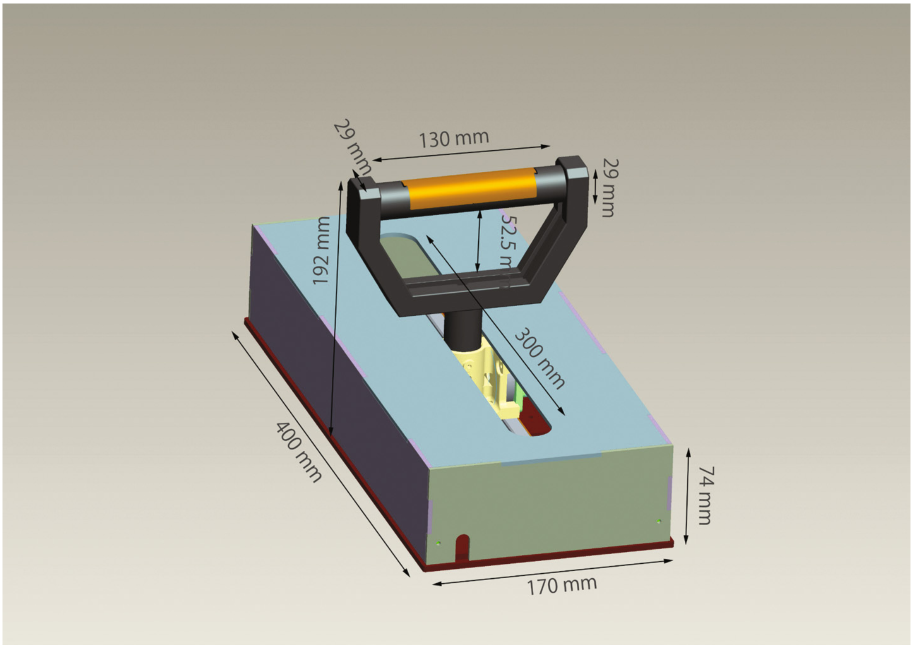


Fig. 1. Image of the handle device. The base can be made to slide forward and backward by pulling or pushing the handle. The handle can move a length of 300 mm. Participants grasped the handle and were instructed to pull it when they experienced anxiety or push it when they felt relaxed.

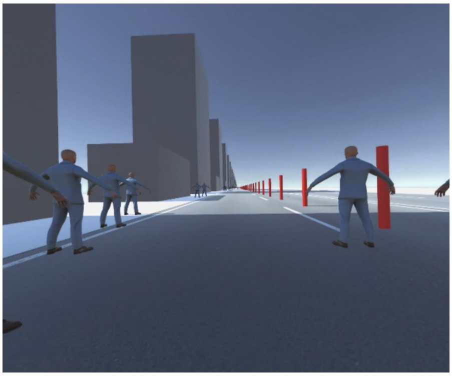


Fig. 2. An example of the stimulus presented in the experiment. The stimulus consisted of computer graphics created by Unity and contained some human models, buildings, and poles. The stimulus was created for giving the participants the experience of virtually riding in an autonomous car, where the car automatically moved in the longitudinal direction. Participants watched the stimulus through a head-mounted display.

We used 3D scenery as an experimental stimulus that gave the participant the illusion of moving forward. This stimulus was developed using a computer graphics platform, Unity R (Unity technologies). The scenes consisted of a road with two lanes, a sidewalk, some human models, buildings, and poles (Fig. 2). The stimuli were automatically changed to make the participants experience a feeling of moving ahead on the straight road. The first ten seconds from the start consisted of gradually moving ahead with acceleration. After reaching the maximum speed, the participants experienced ten seconds of moving at a constant speed followed by ten seconds of moving with gradually decreasing speed until coming to a halt. The scenes moved along a longitudinal direction and were presented through a head-mounted display (Oculus Rift Development Kit 2, Oculus VR, Inc.).

We used two types of acceleration and deceleration patterns, namely linear pattern and exponential (non-linear) pattern; in the former, the speed during acceleration and deceleration was changed at a constant ratio (Fig. 3.a), which resulted in a linear pattern, while in the latter, the speed was changed at a non-linear or exponential ratio, which resulted in a non-linear pattern (Fig. 3.b). Further, the maximum speed was set to 19 km/h, 160 km/h, and 320 km/h in different test runs.

Two sound patterns were created for acceleration (ascending musical scale: sound A) (Fig. 4.a) and deceleration (descending musical scale: sound B) (Fig. 4.b) in the with-sound condition. There was no sound stimulus in the without-sound condition. Sound stimuli were presented through a headphone system (Sennheiser, HD 380 pro).

Participants were also asked to wear a heart rate sensor, myBeat (UNION TOOL CO., WHS-2), to record their heart beat. This data was used for investigating the relationship between the continuous evaluation of subjective anxieties and biometrical indexes, which will be reported in another study.

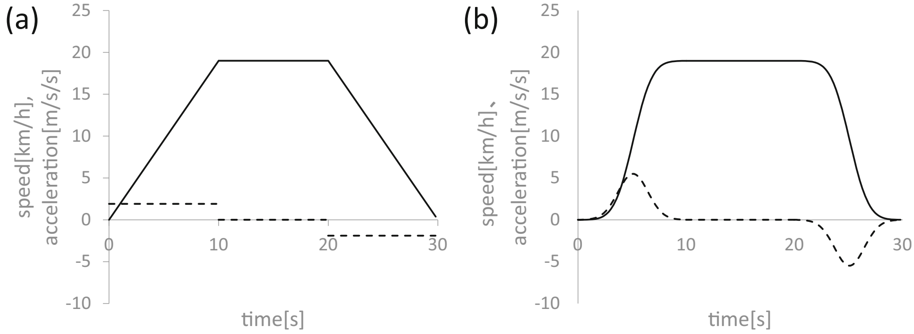


Fig. 3. Graphs indicating the speed patterns of the autonomous car with acceleration and deceleration. The X-axis indicates time while the Y-axis indicates the speed of the car. The solid lines show the speed of the car and dotted lines indicate the acceleration/deceleration ratio. Graph (a): linear pattern; Graph (b): exponential (non-linear) pattern.

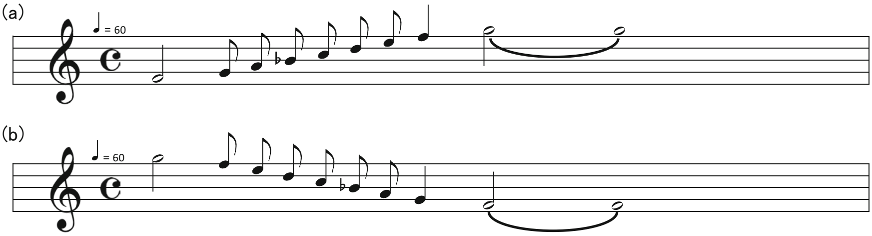


Fig. 4. Sounds presented during acceleration and deceleration. Upper (a): musical score indicating the sound presented during acceleration. Lower (b): musical score indicating the sound presented during deceleration. Both sound patterns lasted 6 s.

2.3 Participants

Sixteen undergraduate or graduate students (of which eight were female) of the University of Tsukuba participated in this experiment voluntarily. All the participants had normal or corrected-to-normal vision and normal hearing.

2.4 Procedure

The experiment was executed separately for each participant. When a participant came the laboratory, s/he was explained the purpose of this experiment and s/he gave their consent to participate in the experiment in writing. Next, the participant was made to wear a heart rate sensor and asked to sit on a seat fitted with the handle device. S/he wore the head-mounted display and a headphone and was asked to grasp the device handle while watching the simulated scenes.

The participants could watch the scenes in all directions using the head-mounted display. They were instructed to grasp the handle while watching the scene and pull the handle when they felt anxious; the more anxiety they felt, the more they pulled the handle toward themselves. On the other hand, they were asked to push the handle when they felt relaxed; the more they felt relaxed, the more they pushed the handle forward. At the end of the experiment, the participants were questioned on their acceptability of the vehicle: “Would you wish to ride this vehicle if it was a kind of public transport?” They were asked to respond on a seven-point scale: “Absolutely not, I won’t” (1) to “Yes, I would” (7). The next test was started after s/he had answered the question. After two practice trials, the experiment was executed with a 5 min resting time between every nine tests. The experiment lasted 90 min approximately.

3 Results

The report in this study covers the anxiety levels as recorded by the handle on the handle device. The remaining data consisting of acceptability rating and biometric indexes will be reported in another paper. We analyzed the data relating to handle position and calculated the average handle position per second. We normalized the data for each participant and separated the data into three sections: the acceleration part, the part with constant speed, and the deceleration part. With this process, we obtained ten data corresponding to the three sections under each of the twelve conditions.

The acceleration and deceleration sections alone were analyzed in this study because the objective of this study was to assess the change in anxiety level with change in speed. The data corresponding to these 2 sections were analyzed individually by 3 × 2 × 2 × 10 analysis of variances (ANOVAs) with repeated measures corresponding to three maximum speeds (19 m/h, 160 km/h, or 320 km/h), two patterns for changing speed (linear vs. non-linear), two sound conditions (with sound vs. no sound), and ten divided sections (intervals of every 1 s from 1 to 10 s for acceleration and deceleration). The average of the handle position standardized for each participant is depicted in Fig. 5. The results of the ANOVA are summarized in Tables 1 and 2.

During acceleration, the effect of the maximum speed was significant ($F(2, 30) = 29.605$, $p < .001$, $\eta_p^2 = .664$; see Table 1). Participants pulled the handle more during the maximum speed of 320 km/h than during 160 km/h ($p < .001$). Similarly, the pulling movement on the handle was more at 160 km/h than at 19 km/h ($p = .007$) and more at 320 km/h than at 19 km/h ($p < .001$). These results indicate that our method could measure the subjective anxieties, which were dependent on the maximum moving speed.

There were significant interactions between the pattern of speed change and sound ($F(1, 15) = 7.251$, $p = .017$, $\eta_p^2 = .326$), speed changing pattern and divided sections ($F(9, 135) = 24.818$, $p < .001$, $\eta_p^2 = .623$), and maximum speed and divided sections ($F(18, 270) = 31.089$, $p < .001$, $\eta_p^2 = .675$). These results indicated that when the simulation was accompanied by sound, the participants pulled the handle more during linear acceleration than during non-linear acceleration ($F(1, 15) = 6.091$, $p = .026$, $\eta_p^2 = .298$). During non-linear acceleration, participants pulled the handle more when the simulation was not accompanied by sound than when sound was present ($F(1, 15) = 5.443$, $p = .034$, $\eta_p^2 = .266$). In addition, in the interval from 3 s to 5 s, the participants pulled the handle more during linear acceleration ($ps < .05$). On the other

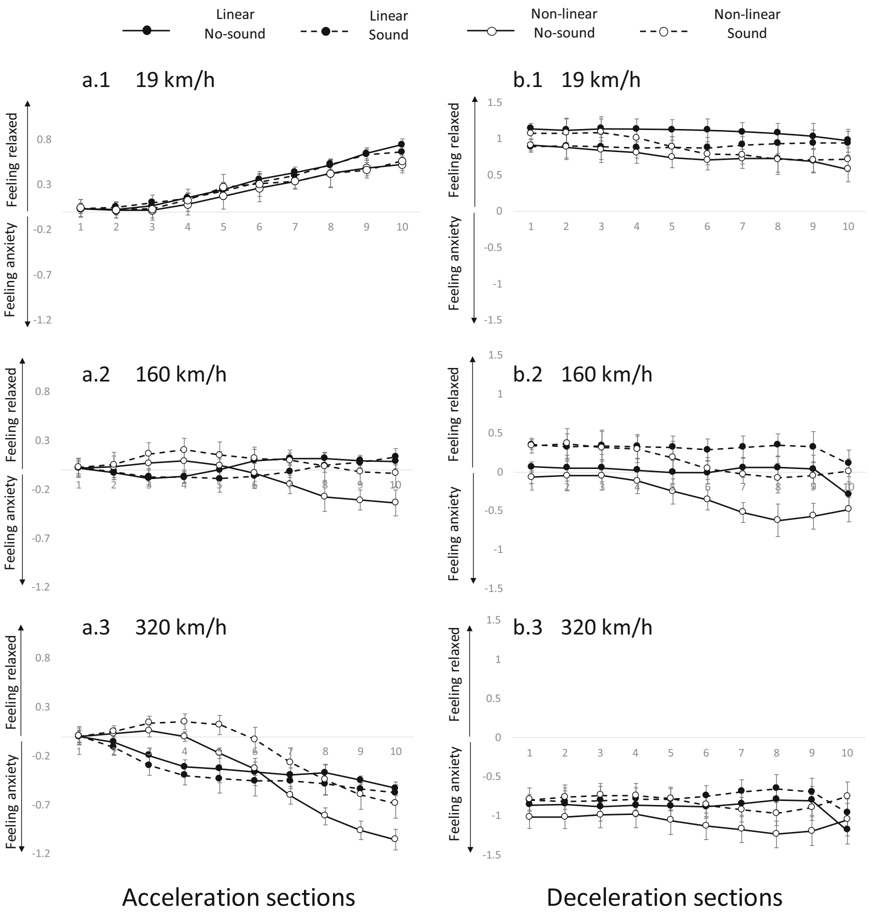


Fig. 5. The mean of the handle positon maintained by participants while they watched the simulation. The graphs on the left (a) indicate the handle positions during acceleration and the graphs on the right (b) indicate the handle positions during deceleration. The solid line with black circles indicates linear pattern of acceleration with no sound. The dotted line with black circles indicates linear pattern of acceleration accompanied by sound. The solid line with white circles indicates non-linear pattern of acceleration with no sound. The dotted line with white circles indicates non-linear pattern of acceleration with sound. The error bar shows the standard error.

hand, in the interval from 8 s to 10 s, they pulled it more during non-linear acceleration than during linear acceleration (*ps* < .05). Furthermore, there were differences between the three conditions of maximum speed for nearly every divided section except at 2 s (*ps* < .05).

In addition, there were three-way interactions between the speed changing pattern, maximum speed, and the divided 10 s ($F(18, 270) = 5.455$, $p < .001$, $\eta_p^2 = .267$), and between speed changing pattern, sound, and the divided 10 s ($F(9, 135) = 3.504$, $p = .001$, $\eta_p^2 = .189$). There were no four-way interactions ($F(18, 270) = 0.867$, *n.s.*,

Table 1. The results of ANOVA of handle position during acceleration. This table lists the main effects, interactions, and significant post hoc multiple comparisons. Lower values in mean indicates higher anxiety, by pulling handles more. A single asterisk indicates a p-value lower than 5%, two asterisks indicate a p-value lower than 1%, and three asterisks indicate a p-value lower than 0.1%. A plus sign indicates a p-value lower than 10%. This statistical analysis was performed using IBM SPSS Statistics 24.

Source	Type III Sum of Squares	df	Mean Squear	F	Sig.	η_p2	Post hoc comparisons
Speed Changing Pattern	0.022	1	0.022	0.021	0.887	0.001	
Maximum Speed	111.157	2	55.578	29.605	0.000	0.664 ***	320 < 160 < 19
Sound	1.132	1	1.132	1.068	0.318	0.066	
Divided Section	0.935	9	0.104	0.801	0.616	0.051	
Speed Chaning Pattern * Maximum Speed	1.526	2	0.763	1.249	0.301	0.077	
Speed Chanigng Pattern * Sound	3.930	1	3.930	7.251	0.017	0.326 *	Under condition with sound non-linear acceleration < linear acceleration * Under non-linear acceleration, without sound < with sound *
Maximum Speed * Sound	0.441	2	0.220	0.520	0.600	0.034	
Sound * Divided Section	0.499	9	0.055	0.742	0.669	0.047	
Speed Changing Pattern * Divided Section	12.005	9	1.334	24.818	0.000	0.623 ***	Under divided section 3 to 5, linear acceleration < non-linear acceleration ** Under divided section 8 to 10, non-linear acceleration < linear acceleration *
Maximum Speed * Divided Section	66.589	18	3.699	31.089	0.000	0.675 ***	Except divided section 2, 320 < 160 < 19 * Under 19 and 320 condition, 1 < 2 < …< 10 +
Speed Changing Pattern * Maximum Speed * Sound	1.467	2	0.734	1.869	0.172	0.111	
Speed Changing Pattern * Maximum Speed * Divided Section	4.105	18	0.228	5.455	0.000	0.267 ***	Under linear acceleration, except section 2, 1 < 3 < … < 10 * Under non-linear acceleration, except section 2 to 4, 1 < 5 < … < 10 *
Speed Changing Pattern * Sound * Divided Section	1.300	9	0.144	3.504	0.001	0.189 ***	Under non-linear acceleration, without sound < with sound *
Maximum Speed * Sound * Divided Section	0.769	18	0.043	1.437	0.114	0.087	
Speed Changing Pattern * Maximum Speed * Sound * Divided Section	0.581	18	0.032	0.867	0.620	0.055	

$\eta_p^2 = .055$). For example, at 320 km/h with non-linear acceleration (lower left image in Fig. 5a.3), the handle position clearly varied across the divided 10 s. In the divided sections from 2 to 10 s, the handle position was significantly or marginally lower in the no-sound condition than with sound (*ps* < .10). In addition, the ascending sound, i.e., the gradual change in upper musical scale in accordance with the accelerating speed, showed decreasing feelings of anxiety in the non-linear acceleration condition alone with maximum speeds of 160 km/h and 320 km/h.

During deceleration (Table 2), there were significant and marginally significant effects of the speed changing pattern ($F(1, 15) = 8.573, p = .010, \eta_p^2 = .364$), maximum speed ($F(2, 30) = 54.993$, $p < .001$, $\eta_p^2 = .786$), sound ($F(1, 15) = 3.775$, $p = .071$, $\eta_p^2 = .201$), and divided 10 s ($F(9, 135) = 6.890$, $p < .001$, $\eta_p^2 = .315$) (Fig. 5(b)). Additionally, there were significant or marginally significant interactions between speed

Table 2. The results of ANOVA of handle position during deceleration. This table lists the main effects, interactions, and significant post hoc multiple comparisons. Lower values in mean indicates higher anxiety, by pulling handles more. A single asterisk indicates a p-value lower than 5%, two asterisks indicate a p-value lower than 1%, and three asterisks indicate a p-value lower than 0.1%. A plus symbol indicates a p-value lower than 10%. This statistical analysis was performed using IBM SPSS Statistics 24.

Source	Type III Sum of Squares	df	Mean Squear	F	Sig.	η_p2	Post hoc comparisons
Speed Changing Pattern	15.329	1	15.329	8.573	0.010	0.364 **	Non-linear deceleration < linear deceleration
Maxximum Speed	1038.690	2	519.345	54.993	0.000	0.786 ***	320 < 160 < 19
Sound	14.730	1	14.730	3.775	0.071	0.201 +	
Divided Section	10.141	9	1.127	6.890	0.000	0.315 ***	Section 10 < 2 = 3 = 4
Speed Changing Pattern * Maxximum Speed	1.053	2	0.526	0.267	0.767	0.017	
Speed Changing Pattern * Sound	4.985	1	4.985	4.410	0.053	0.227 +	Under conditions without sound, linear deceleration < non-linear deceleration *** Under non-linear deceleration, without sound < with sound **
Maxximum Speed * Sound	13.358	2	6.679	3.321	0.050	0.181 *	Under 160km/h condition, without sound < with sound *
Speed Changing Pattern * Divided Section	8.804	9	0.978	10.676	0.000	0.416 ***	Under section 5 to 9, non-linear deceleration< linear deceleration *
Maxximum Speed * Divided Section	2.310	18	0.128	1.427	0.118	0.087	
Sound * Divided Section	0.374	9	0.042	0.635	0.766	0.041	
Speed Changing Pattern * Maxximum Speed * Sound	0.716	2	0.358	0.185	0.832	0.012	
Speed Changing Pattern * Maxximum Speed * Divided Section	2.545	18	0.141	4.119	0.000	0.215 ***	Under 19km/h condition, and in section 5 to 10, deceleration without sound < deceleration with sound * Under 160km/h and 320 km/h condition, and section 6 to 9, deceleration without sound < deceleration with sound *
Speed Changing Pattern * Sound * Divided Section	0.221	9	0.025	0.690	0.717	0.044	
Maxximum Speed * Sound * Divided Section	0.241	18	0.013	0.296	0.998	0.019	
Speed Changing Pattern * Maxximum Speed * Sound * Divided Section	0.732	18	0.041	1.592	0.062	0.096	

changing pattern and sound ($F(1, 15) = 4.410$, $p = .053$, $\eta_p^2 = .227$), maximum speed and sound ($F(2, 30) = 3.321$, $p = .050$, $\eta_p^2 = .181$), and speed changing pattern and divided 10 s ($F(9, 135) = 10.676$, $p < .001$, $\eta_p^2 = .416$). There was three-way interaction between speed changing pattern, maximum speed, and divided 10 s ($F(18, 270) = 4.119$, $p < .001$, $\eta_p^2 = .215$). The results relating to the deceleration section indicated that differences in simulation conditions with sound and without sound were significant only for the maximum speed of 160 km/h ($F(18, 270) = 4.119$, $p < .001$, $\eta_p^2 = .215$). However, the subjective anxiety was higher for non-linear deceleration than for linear deceleration without sound ($p < .001$). In the case of non-linear deceleration, the subjective anxiety was higher without sound ($p = .015$). Moreover, we found that there was a significant difference between the first five seconds and last five seconds during non-linear deceleration and maximum speed of 19 km/h (*ps* < .05).

4 General Discussion

The present study investigated whether our new haptic device was effective in measuring real-time subjective anxiety caused by riding in an autonomous car. In the experiment, a participant who was wearing a head-mounted display watched a CG animation that simulated the feeling of riding in an autonomous car. The CG animation stimuli were operated using three variables: maximum speed (19 km/h, 160 km/h, or 320 km/h), acceleration/deceleration pattern (linear or non-linear), and with or without ascending/descending sounds (sound and no sound). While the participant was watching the simulation, s/he grasped the handle device and pulled when he/she experienced anxiety and pushed when they felt relaxed.

The results of the experiment demonstrated that the higher the stimulus of speed was, the higher was the anxiety the participants felt. Additionally, subjective anxiety was higher during non-linear changes in speed than during linear speed changes. The results indicated that our new method could measure the anxiety experienced by the participants in accordance to the speed at the time, while riding an automated vehicle. Moreover, we found that sounds can decrease feelings of anxiety during automated driving. The result showed the possibility that the sound might eliminate passenger anxiety related to riding an autonomous car if an appropriate sound was presented especially when the car started to accelerate.

Even though we got some interesting results here, it should be noted that this experiment is not “real” in some meaning, because there are much differences between the circumstances of an actual passenger riding an autonomous car and one who watches a CG simulation of the same. Although in some previous papers reported that people could perceive self-movement using CG simulation, by the sensation of self-motion or vector [5, 6], it is not clear that those sensation are same with those in the real riding. Those differences might be related to the fact that the maximum speed in the current study was 320 km/h. This was because most participants did not experience anxiety when they were shown a simulation where the speed was less than 320 km/h in our CG animation. The speed of 320 km/h cannot be a practical speed for an autonomous car, which implies that there was really a gap between the CG animation and the real-life situation. Those differences should be passengers’ feeling of the gravitational acceleration, vibrations due to the movement of the car, or airflows from the window, in a real-life situation. If we would like to use this method with simulations to evaluate the design of motion control actually, methods to fill those gaps might be important.

The experiment performed in current study showed that the handle device could measure the continuous subjective anxiety, which varied with factors in maximum speed, acceleration or deceleration pattern, and presence of sound in the simulation presented using a head-mounted display. The result indicated the possibility that using computer simulation with the handle device enabled the measurement of not only anxiety but also of other emotions such as “fun associated with riding.” Actually, even in current results, the data upper than the “anxiety 0” point, pushing the handle might mean a participant’s relaxing or enjoying the riding. Using the new device to evaluate more various *KANSEI* subjective feeling might be useful, to create more acceptable designs of motion control. Possibilities and limits for capturing real-time or on-time subjective feeling should be investigated in future.

References

1. Peyre, W., Cestac, J., Delhomm, P.: Intention to use a fully automated car: attitudes and a priori acceptability. Transp. Res. Part F **27**, 242–263 (2014)
2. Spielberger, C.D.: State-Trait Anxiety Inventory: a Comprehensive Bibliography. Consulting Psychologists Press, Palo Alto (1989)
3. Promeranz, B., Macaulay, R.J., Caudill, M.A., Kutz, I., Adam, D., Gordon, D., Kilborm, K.M., Barger, A.C., Shannon, D.C., Cohen, R.J., Benson, H.: Assessment of autonomic function in humans by heart rate spectral analysis. Am. J. Physiol.-Heart Circulatory Physiol. **248**, 151–153 (1985)
4. Akselrod, S., Gordon, D., Ubel, F.A., Shannon, D.C., Barger, A.C., Cohen, R.J.: Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. Sci. New Series **213**, 220–222 (1981)
5. Seno, T., Fukuda, H.: Stimulus meanings alter illusory self-motion (vection) – experimental examination of the train illusion. Seeing Perceiving **25**, 631–645 (2012)
6. Palmisano, S., Allison, R.S., Schira, M.M., Barry, R.J.: Future challenges for vection research: definitions, functional significance, measures, and neural bases. Frontiers Psychol. **6**(193), 1–15 (2015). doi:10.3389/fpsyg.2015.00193

Investigating the Influence of Emotion in Air Traffic Controller Tasks: Pretest Evaluation

Martina Truschzinski[1(✉)], Georg Valtin[2], and Nicholas H. Müller[2]

[1] Department of Automation Technology, Technische Universität Chemnitz, Chemnitz, Germany
`mtru@hrz.tu-chemnitz.de`
[2] Department of Media Psychology, Technische Universität Chemnitz, Chemnitz, Germany

Abstract. Air traffic controllers (ATC) have to perform under highly demanding circumstances, and they are often responsible for the lives of more than 5.000 people in mid-air at any given moment. Therefore, besides regular scheduled breaks to recover, it is of paramount importance to be in a level-minded state in order to maximize cognitive capabilities. Within this context and based on our previous research with active air traffic controllers, we designed a comparable scenario in which two airplanes from different directions converge and where subjects have to prevent collision. They watched an emotion inducing video shortly before participating in the experiment. The induced positive or negative valence lead to significant group-differences during their performance of the ATC task.

1 Introduction

According to current statistics, the amount of airline passengers will continue its positive development over the next years, with expected annual growth rates of up to five percent [3,23]. To maintain the resulting needs and ensure smooth and safe traveling, the duty of air traffic controllers (ATCs) is of high importance. However, tasks of air traffic controllers (ATC) are complex and demanding [31]. Especially, factors which influence the general workload like traffic complexity or frequency congestion also increase the mental workload and therefore might influence the failure rate of an air traffic controller [6,20].

In general, the performance of high demanding tasks or decision-making processes are depending on several influences [22,29,32,33,40]. Individual differences can be assessed by personality traits [14]. Especially, the traits neuroticism and conscientiousness seem to be important when attempting to simulate the rate of cognitive processes. Higher amounts of neuroticism usually lead to a lower performance of participants, because the cognitive load imposed by neuroticism (e.g. worry) reduced the processing capacity of the working memory system. People with a higher amount of conscientiousness, on the other hand, results in better overall training performances [40].

Another influence on task-performance depends on the difficulty of the tasks and the time pressure, which influences the workload and therefore affects the performance of problem solving [13,44].

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 220–231, 2017.
DOI: 10.1007/978-3-319-58472-0_18

However, time pressure does not simply increase the cognitive workload, it also influences the emotional valence, resulting in a measurable increase of the arousal and less control options over tasks, compared to scenarios with less challenging time constraints [9,42]. This experience of task-induced emotions could cause a limitation of cognitive resources needed to solve high demanding tasks [12]. This might in turn lead to an increase of the mental workload which once again impairs the task-performance [32]. Emotional influences, however, do not always result in negative effects. By investigating how emotions influences different executive functions, [29] found that positive mood decreases the performance of working memory tasks and simple problem solving tasks (Tower of London) whereas it increases the task performance which assessed fluency and creativity. In the investigation of influences of negative mood to performance, there exists contradictory results as well: Mitchell and Phillips [29] found no significant influence of negative mood to their investigated tasks performances. Jeon et al. [25] found in a driving scenario that negative mood impairs the performance of the task negatively, Berggren et al. [2] and Jeon and Coshere [24] found evidences that emotional stimuli could slow down executive functions and van Dillen and Koole [11] found that contrary to positive mood which lasts longer, negative mood might be distracted by solving complex and demanding tasks.

2 Emotion and Workload Within Air Controller Tasks

We present research based on a previously developed and validated emotion model [30]. Within our new project "StayCentered – Methodenbasis eines Assistenzsystems für Centerlotsen" (methodical base for an air traffic controller assistance system - MaCeLot) we improve and validate the aforementioned model to assess emotional and cognitive influences on the performance of air traffic controller.

In general, air traffic controllers working within designated air space areas, where they have to maintain the flow of air traffic by contacting aircraft pilots and providing them with advice, instructions and information about weather conditions and safe flight, ascent and descent paths. Under normal conditions, any given airspace sector is monitored by a dyad of air traffic controllers. While they are located at a shared workspace, they assume different roles: The executive is responsible for the coordination of the flights as well as the communication with the pilots, while the planner is managing the acceptance or handover of flights from or to other sectors (for details: Pfeiffer et al. [34]). The work of air traffic controllers is considered to be very demanding, since it is characterized by high responsibility and complex decision making under time pressure which can result in a negative stress response [21,35]. Short term consequences of negative stress such as anxiety, despondence, anger, and cognitive impairments (e.g. Khansari et al. [27]) can directly effect the emotional and cognitive state of the air controllers and therefore hamper their performance [39]. While the existing literature covers several factors that influence the performance of aviation-related workers like fatigue [8], situation awareness [45] and mental imagery [38] as well as assessment techniques to measure workload (e.g. Vewey and Veltman [43]), it is rather unclear how emotional conditions, for instance as a result of stress, influence the air traffic controllers performance.

Therefore, we examined the influences of emotion and workload on the performance in an air traffic scenario.

3 Pretest Experiment

We collected data from 7 volunteers located at the University of Technology Chemnitz (57.1% male, 42.9% female, $M_{Age} = 27.29$; SD = 2.75). The majority of 85.7% participants had no prior experience in air traffic control tasks (including, but not limited to, video games).

3.1 Experimental Design

We conducted this pretest as a between-subjects design, in which each participant completed a simulated air traffic controller task. The simulation was divided into a practice session and two conditions of 4 min, a neutral condition (NC) and an emotional condition (EC). Before the emotional session each participant watched either a positive or a negative film clip to induce the respective mood.

The aim of this pre-test was to revise if the used methods are able to recognize the influences of workload and mood on the performance of the air controller task.

3.2 Methods

Since mixed results exist in literature about the influence of emotions to the performance assessed with different measurements, we decided to use both subjective and objective measurements. We used these measurements to verify as the subjective perception of the emotional and cognitive state of each participant is coherent with the objective measurements.

Measurements of Mood. Changes in the mood were assessed with a questionnaire "Aktuelle Stimmungsskala" (ASTS). This questionnaire [10] is a shortened version of the "profile of mood state"- scale (POMS) from McNair et al. [28]. The ASTS consists of 19 German adjectives calculating five different scales, representing anger, sadness, hopelessness, positive mood and tiredness. Subjects have to estimate their current feeling by rating on a scale from 7 (very strong) to 1 (not at all) how well this adjective represents their feeling. The questionnaire is validated and considered a good to very good reliability.

Furthermore, we recorded the skin conductance during the whole experiment to measure arousal. This data is used as a additionally indicator for the mood induction to check how strong participants are affected by the videos and also to identify how long such an induced mood lasts and influences subjects during the task.

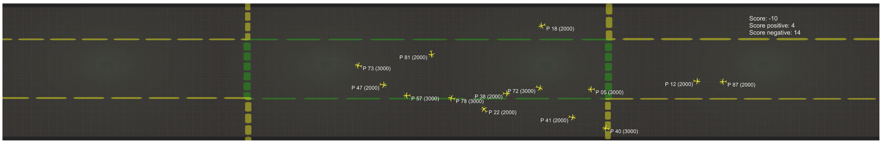


Fig. 1. The presentation of the stimulus represents the visual style of the radar screen. (Color figure online)

Measurements of personality. The personality traits neuroticism and conscientiousness of each participants were assessed with the NEO-FFI questionnaire [5].

Measurements of Workload. Changes in the workload were assessed with the NASA-TLX based on the weighted mean of 6 sub-scales: mental demand, physical demand, temporal demand, effort, performance and frustration level [17]. This questionnaire is used and revised in air traffic control since more than two decades [16].

As objective measurements, we recorded the pupillary response as [1] referred that different task difficulties and workload led to different pupil dilation. With this measurement, we prove if the reported workload is coherent increased or decreased to the workload assessed by recorded pupil dilation.

Measurements of Performance. During the experiment, the simulation counted and logged if airplanes left an airspace unharmed or if a collision happened. Each airplane left the airspace unharmed counted +5 points, each airplane which collided in the responsible area gives −3 points and every airplane collided in the outer airspace area counted −2 points. The overall score was calculated and displayed on time during neutral and emotional conditions.

3.3 Material

Simulation of the air controller task. We designed a radar screen with different airspaces using Unity3D (see Fig. 1). These airspaces were displayed on three screens in front of the participants. During the whole experiment, every 4 s a pair of airplanes appeared from top, down, left or right heading towards the opposite side of the screen. At randomly chosen collision points, a framed green rectangle in the middle screen, reflected the so called responsible airspace. If a collision happened, the program played a collision sound and counts negative points to the score. If an airplane left the screen successfully, the software counted these events as positive points. The simulation stopped automatically, if the scheduled maximum experiment time was reached.

Mood Induction. Participants were randomly assigned to a happy or sad mood condition. We used two video clips with a duration of 4 min to induce the assigned mood (happy or sad). This induction methodology is described by [36] as well as used and revised in several studies [4,15,18]. Participants in the sad condition watched a film clip from "The Lion King" and participants in the happy condition watched a film clip from "When Harry mets Sally". The videos were played on one of the three screen, on which the simulation of the air controller task is going to be displayed.

We also conducted a small pretest with video material were participants had to watch two videos in each condition (sad and happy) at there home. Afterwards they filled out a questionnaire by rating each video on a scale from 5 (strong) and 1 (not at all) how well this video represented the intended mood. The videos we used for the negative mood were "The Lion King" and "The Champ"(English) while the videos for a positive mood were "When Harry met Sally" and "Zoomania" (Trailer). After evaluating the questionnaires, we decided to use "The Lion King" and "When Harry mets Sally" for our mood induction as they were rated the highest and in the native language (German) of most of the expected participants.

Empatica-bracelet. The Empatica E4 wristband with a relatively small weight of only 40 grams combines sensors for the galvanic skin response, heart-rate, skin temperature and an accelerometer. Therefore, it is a valuable tool to unobtrusively measure biophysiological responses without wires. The skin-conductance is measured by two electrodes on the inside of the wristband. It measures in μs at 4 Hz while the resolution is at about 900 picosiemens with a range of 0.01 and 100 μs. The heart-rate is measured by four photo-diods via photoplethysmographie which is based on volume changes of the arterial blood-flow in the outer wrist (BVP). The rate of measurement is 64 Hz and besides the BVP it also records the heart-rate-variability.

Eye tracker. For the recordings of gaze-patterns and pupillary responses, we used a SMI ETG2 mobile eye-tracker which records eye-movements and the changes of the pupil-dilation at 60 Hz. The whole eye-tracking-glasses setup weighs about 47 grams and is about the size of everyday protective glasses. They have a tracking accuracy of 0.5° and tracks the gaze of a user at 80° horizontally and 60° vertically. The coax camera is installed inside the frame of the glasses and records the field of view at 960×720 p with 30 frames per second.

3.4 Procedure

The study was conducted in a prepared experiment room at the University of Technology Chemnitz. After the participants completed the consent forms, they completed the NEO-FFI, the first ASTS and a demographic questionnaire (see Fig. 2). Afterwards, the preparation and calibration phase started including the setup of the Empatica-bracelet and the eye tracker glasses.

In the following practice session, each participant got the instruction and an easy task to learn how to manage the airplanes on the screens. This section was finished if the subject was able to manage the scenario (10 consecutive correct answers) and the participant completed another ASTS and the first NASA-TLX questionnaire. After the practice section the participant was exposed to a so

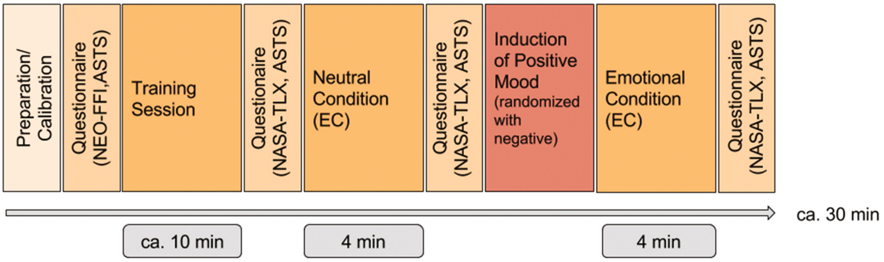


Fig. 2. Overview of procedure for the pretest.

called neutral section in which the participant has to maintain all airplanes on the screen for 7 min.

In this section, subjects controlled airplanes on the radar with their speech (see Fig. 3) as a wizard-of-oz-experiment. Each participant was the responsible air traffic controller of the middle airspace (green rectangle) and had to maintain all airplanes which randomly appeared in the outer air spaces. The airplanes appeared as a couple at a given frequency with a given number and a random height from both sides or from top and down heading to the same randomized point in the responsible airspace. During the whole experiment, participants have to keep in mind, that since we displayed a 3D scene on a 2D screen, airplanes which do not collide in 3D could appear at the same location of the screen.

The used control commands to maintain the air traffic flow and to avoid collisions were similar to real air controller commands (see Fig. 3). Therefore, subjects had to include in their commands the number of the chosen airplane plus the information about what they want to change, for example the heading or height of the airplane. The experiment leader played the role of the pilot controlling the airplanes in the background.

The neutral section was closed by completing the third ASTS and the second NASA-TLX questionnaire. Afterwards, each participant had to watch a video

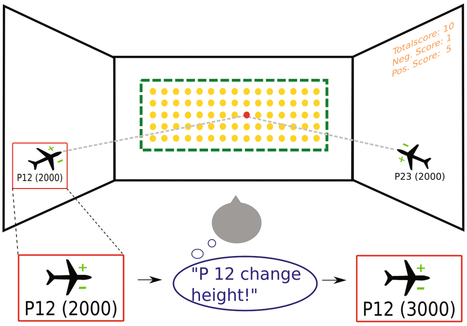


Fig. 3. How participants changes the height of an airplane. (Color figure online)

either a positive or a negative one and, subsequently performed last trial section. Since we induced a mood within this section, it is called emotion section. Within this trial, subjects has to maintain the airplanes 7 min as they did in the neutral session. Finally,the experiment was closed by filling out the last ASTS and NASA-TLX.

3.5 Data Preparation

After conducting the experiment and preparing the pupil diameter data recorded by the eye tracker (see material section), data had to be cleaned and artifacts to be removed, blinks and other undesired patterns in the data stream [1] were filtered out. Therefore, we used MATLAB-functions to implement standard methods for cleaning and analyzing pupil diameter data. First, we deleted all blinks in the signal, which are characterized by zero values in the data stream. Then, we interpolated the missing values and used a MATLAB function that detects and deletes outliers (values outside the 25th and 75th percentile of the range of pupil diameter in the whole experiment) and a median filter in order to smooth the signal. Participants with more than 18% blinks or zeros in the data stream were excluded from the statistical analysis as the filtering functions and the evaluation could be falsified by very noisy signals.

The EDA-data of the Empatica-bracelet was also cleaned by removing artifacts and outliers with the MATLAB-functions. Furthermore, we exclude participants with small values (mean of all conditions < 0.3) and low variance in their data, because it seems that their measure of the skin conductance is not valuable for gathering information about their arousal levels.

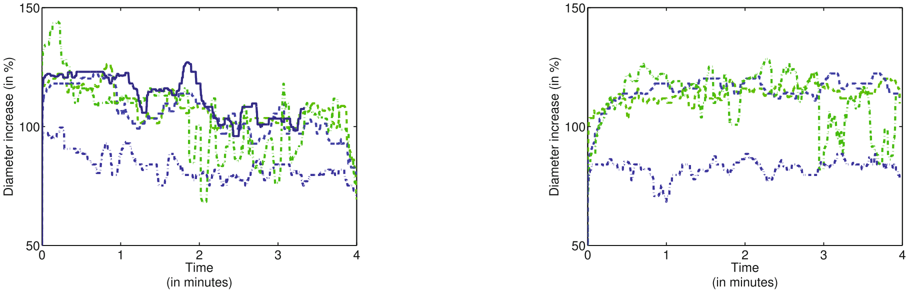
4 Preliminary Results and Discussion of the Used Methods

Experiment Design. For a maximum of external validity, the measurement of the cognitive and emotional states should ideally be performed under realistic conditions. However, due to legal constraints and security issues it is impossible to conduct experiments with German air traffic controllers while at work. Therefore, we needed to find an adequate scenario to gather data of participants while fulfilling typical air traffic controllers tasks. For this purpose, we decided to create a subtask regarding the prevention of aircraft collisions, since the whole bandwidth of air controller tasks is much too complex for laypersons to handle, and it is also influenced by too many variables to be subject of an controlled experiment (for an overview over the cognitive complexity in air traffic control: [19]). The main sources of information about the ongoing traffic in the sector are the flight strips and the radar screen. Since understanding and properly using the flight strips requires special knowledge, we used a simulated radar screen as device for stimulus presentation. Prior research suggests a possible confounding influence of the visual presentation on the mental workload in comparable tasks [26] due to different levels of perceptual load associated with the visual stimuli [7]. To keep our task comparable to the air traffic controllers work, we re-created a radar-screen using the same color scheme as well as the highly similar icons and information texts (see Fig. 1). Although realistic scenarios involve multiple airplanes at the same time with differences in heading, speed and flight level, this would be overly complex for participants without expertise. Therefore, we use only two airplanes which move faster than the ones the real radar-screen to compensate for the lower demands due to the reduced number of airplanes. The characteristics of the experimental tasks remains the same as for the air traffic controllers: Participants have to percept visual stimuli, retrieve relevant information and perform an adequate input, if necessary.

ASTS. The analysis of the ASTS questionnaires showed no significant correlation with the score, yet. But we saw also huge differences in the small sample size. Participants reported no problems with this questionnaire. Thus, we will use it for the upcoming experiment.

NASA-TLX. The used NASA-TLX questionnaire causes several questions and confusion about deciding how they should fulfill this questionnaire. In general, subjects said that the questionnaire is unintuitive and the description difficult to understand. Thus we decide to replace the NASA-TLX with the Instantaneousness self-assessment of workload (ISA) in a questionnaire form [41]. This technique is also used in air traffic control to reporting the current workload to their supervisors as well as in simulations.

Pupil diameter. We analyzed the change in the pupil diameter during the video session and in the emotional session. In Fig. 4 we show the recorded curve progressions of all participants. In the analysis of the pupil diameter or workload in the emotional session we see that the diameter within the positive session are more fluctuating than the pupil diameter during the negative session. However, we see that the pupil diameter is increased during the air traffic task (see



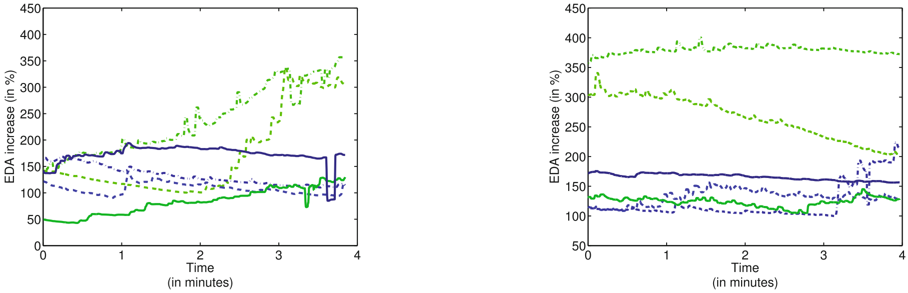
(a) Diameter change during Video session.

(b) Diameter change during Emotion session.

Fig. 4. Recorded diameter during video and emotional session. Green lines represent positive mood induced participants and blue lines negative mood induced participants. (Color figure online)

Fig. 4(b)) in contrast to the video session, we see that pupil diameter during the pupil diameter respectively the workload is decreased (see Fig. 4(a)). Thus, we assume that the recorded pupil diameter is able to represent the workload.

EDA. In contrast to the pupil diameter, the EDA-values are increased during the video session in most of the cases (see Fig. 5(a)). This was expected and shows that the mood induction with the videos was successful. The comparison of the EDA curve progression between positive (green lines) and negative (blue lines) session shows that participants watching the positive video are more aroused as the participants watched the negative video. This might be the case as the induced emotion in the negative video is sadness and this might be not an emotion that is represent by a high arousal [37]. During the emotional session participants doing the air controller tasks this arousal influence lasts for three participants the whole time (see Fig. 5(b)). Two of the participants watched the



(a) EDA change during Video session.

(b) EDA change during Emotion session.

Fig. 5. Recorded EDA during video and emotional session. Green lines represent positive mood induced participants and blue lines negative mood induced participants. (Color figure online)

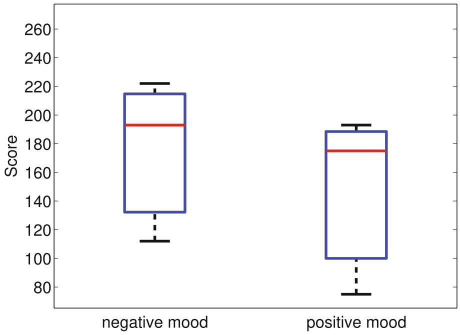


Fig. 6. Comparison of the reached scores between positive and negative mood induced participants.

negative video were more aroused by the tasks and one positive induced mood participant shows an decrease in the curve progression.

Score. If we compared the score in the experiment of negative and positive mood induction (see Fig. 6) we see that participants with negative induced mood reached a higher score than participants with the positive induced mood. Thus, we see that the experiment setting is able to investigate the influence of positive and negative mood to the performance of an air controller task.

5 Conclusion

We described how we investigate the influence of emotion and workload on the performance of air controller tasks. We discussed the possible influences and how we are able to investigate this influences within an experiment with students. We present the structure and methods of our pretest in order to conduct a real experiment with a higher sample size. We presented preliminary results and discussed all used methods in respect to their expected out-coming and their usability for the participants. Therefore, we decided to use all described techniques except the NASA-TLX, which we will replace with the ISA- questionnaire. Based on this results we are able to conduct our experiment investigating the influences of workload and emotion on air controller tasks.

References

1. Beatty, J., Lucero-Wagoner, B.: The pupillary system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology, 2nd edn., pp. 142–162. Cambridge University Press, Cambridge (2000)
2. Berggren, N., Koster, E.H.W., Derakshan, N.: The effect of cognitive load in emotional attention and trait anxiety: an eye movement study. J. Cogn. Psychol. **24**(1), 79–91 (2012)
3. Boeing: Estimated annual growth rates for passenger and cargo air traffic from 2016 to 2035, by region. In statista - the statistics portal. Technical report (nd). https://www.statista.com/statistics/269919/growth-rates-for-passenger-and-cargo-air-traffic
4. Bohn-Gettler, C.M., Rapp, D.N.: Depending on my mood: mood-driven influences on text comprehension. J. Educ. Psychol. **103**(3), 562–577 (2011)
5. Borkenau, P., Ostendorf, F.: NEO-FFI, 2nd edn. Hogrefe, Göttingen (1992)
6. Brookings, J.B., Wilson, G.F., Swain, C.R.: Psychophysiological responses to changes in workload during simulated air traffic control. Biol. Psychol. **42**, 361–377 (1996)
7. Chen, S., Epps, J.: Using task-induced pupil diameter and blink rate to infer cognitive load. Hum.-Comput. Interact. **29**(4), 390–413 (2014)
8. Chu, T.C., Wang, T.Y., Ke, G.C.: Fatigue minimization work shift scheduling for air traffic controllers. Int. J. Autom. Smart Technol. **3**(2), 91–99 (2013)
9. Cœgnet, S., Miller, H., Anceaux, F., Naveteur, J.: How do time pressured drivers estimate speed and time? Accid. Anal. Prev. **55**, 211–218 (2013)

10. Dalbert, C.: Subjektives wohlbefinden junger erwachsener: theoretische und empirische analysen der struktur und stabilität. Zeitschrift für Differentielle und Diagnostische Psychologie **13**, 207–220 (1992)
11. van Dillen, L.F., Koole, S.F.: Clearing the mind: a working memory model of distraction from negative mood. Emotion **7**(4), 715–723 (2007)
12. Freese, M., Jakobi, A., Jipp, M.: Emotions in decision-making processes of air traffic management stakeholders. In: Proceedings of the Human Factors and Ergonomics Society Europe, Groningen, Niederlande (2015)
13. Galy, E., Cariou, M., Mélan, C.: What is the relationship between mental workload factors and cognitive load types? Int. J. Psychophysiol. **83**(3), 269–275 (2012)
14. Goldberg, L.R.: The Big Five trait taxonomy: history, measurement, and theoretical perspectives. J. Pers. Soc. Psychol. **59**(6), 1216 (1990)
15. Gross, J.J., Levenson, R.W.: Emotion elicitation using films. Cogn. Emot. **9**(1), 87–108 (1995)
16. Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **50**(9), 904–908 (2006)
17. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research (1988)
18. Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., Bartussek, D.: A revised film set for the induction of basic emotions. Cogn. Emot. **7**(19), 1095–1109 (2005)
19. Hilburn, B.: Cognitive complexity in air traffic control: a literature review. EEC note 4(04) (2004)
20. Hilburn, B.: Cognitive task analysis of future air traffic control concepts: the TCAS downlink scenario. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 51, pp. 98–101. Sage Publications (2007)
21. Hopkin, V.D.: Human Factors in Air Traffic Control. Taylor & Francis, London (1995)
22. Hu, Y., Wang, D., Pang, K., Xu, G., Guo, J.: The effect of emotion and time pressure on risk decision-making. J. Risk Res. **18**(5), 637–650 (2015)
23. IATA: Annual growth in global air traffic passenger demand from 2005 to 2017. In statista - the statistics portal. Technical report (nd). https://www.statista.com/statistics/193533/growth-of-global-air-traffic-passenger-demand
24. Jeon, M., Croschere, J.: Sorry, i'm late; i'm not in the mood: negative emotions lengthen driving time. In: Harris, D. (ed.) EPCE 2015. LNCS (LNAI), vol. 9174, pp. 237–244. Springer, Cham (2015). doi:10.1007/978-3-319-20373-7_22
25. Jeon, M., Walker, B.N., Yim, J.: Effects of specific emotions on subjective judgment, driving performance, and perceived workload. Transp. Res. Part F: Traffic Psychol. Behav. **24**, 197–209 (2014)
26. Jessee, M.S.: Ocular activity as a measure of mental and visual workload. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **54**(18), 1350–1354 (2010)
27. Khansari, D.N., Murgo, A.J., Faith, R.E.: Effects of stress on the immune system. Immunol. Today **11**(5), 170–175 (1990)
28. McNair, D.M., Lorr, M., Droppleman, L.F.: Profile of Mood States. Educational and Industrial Testing Service, San Diego (1971)
29. Mitchell, R.L.C., Phillips, L.H.: The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. Neuropsychologia **45**, 617–629 (2007)
30. Müller, N.H., Truschzinski, M.: An emotional framework for a real-life worker simulation. In: Kurosu, M. (ed.) HCI 2014. LNCS, vol. 8511, pp. 675–686. Springer, Cham (2014). doi:10.1007/978-3-319-07230-2_64

31. Mogford, R.H., Guttman, J.A., Morrow, S.L., Kopardekar, P.: The complexity construct in air traffic control: a review and synthesis of the literature. Technical report, DTIC Document, Springfield, Virginia (1995)
32. Navon, D., Gopher, D.: On the economy of the human-processing system. Psychol. Rev. **86**(3), 214–255 (1979)
33. Pessoa, L.: How do emotion and motivation direct executive control? Trends Cogn. Sci. **4**(13), 160–166 (2009)
34. Pfeiffer, L., Valtin, G., Müller, N.H., Rosenthal, P.: The mental organization of air traffic and its implications to an emotion sensitive assistance system. Int. J. Adv. Life Sci. **8**, 164–174 (2015)
35. Roske-Hofstrand, R.J., Murphy, E.D.: Human information processing in air traffic control. In: Smolensky, M.W., Stein, E.S. (eds.) Human Factors in Air Traffic Control, pp. 65–114. Academic Press (1998)
36. Rottenberg, J., Ray, R.D., Gross, J.J.: Emotion elicitation using films. In: Handbook of Emotion Elicitation and Assessment. Series in Affective Science, pp. 9–28. Oxford University Press, New York (2007)
37. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
38. Shorrock, S.T., Isaac, A.: Mental imagery in air traffic control. Int. J. Aviat. Psychol. **20**(4), 309–324 (2010)
39. Staal, M.A.: Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework. BiblioGov, Moffett Field (2013)
40. Studer-Luethi, B., Jaeggi, S.M., Buschkuehl, M., Perrig, W.J.: Influence of neuroticism and conscientiousness on working memory training outcome. Pers. Individ. Differ. **53**(1), 44–49 (2012)
41. Tattersall, A.J., Foord, P.S.: An experimental evaluation of instantaneous self-assessment as a measure of workload. Ergonomics **39**(5), 740–748 (1996)
42. Van Kleef, G.A., De Dreu, C.K.W., Manstead, A.S.R.: The interpersonal effects of emotions in negotiations: a motivated information processing approach. J. Pers. Soc. Psychol. **87**(4), 510–528 (2004)
43. Verwey, W.B., Veltman, H.A.: Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. J. Exp. Psychol. Appl. **2**(3), 270 (1996)
44. Wickens, C.D.: Multiple resources and performance prediction. Theoret. Issues Ergon. Sci. **3**(2), 159–177 (2002)
45. Wickens, C.D.: Situation awareness and workload in aviation. Curr. Dir. Psychol. Sci. **11**(4), 128–133 (2002)

Stressor Load and Stress Resilience: A New Perspective for Occupational Stress

Lijing Wang[1](✉), Yanlong Wang[1], Yingchun Chen[2], Dayong Dong[2],
and Wenjun Dong[2]

[1] School of Aeronautic Science and Engineering, Beihang University,
37th Xuanyuan Road, Haidian District, Beijing 100191, China
`wanglijing505@126.com`, `wyllaf@sina.com`
[2] Commercial Aircraft Corporation of China Ltd,
5188th Jinke Road, Pudong New Area, Shanghai 201210, China
`{chenyingchun, dongdayong, dongwenjun}@comac.cc`

Abstract. In modern society with fierce competition, occupational stress has become a critical factor for people's health and well-being. Through several outstanding models proposed for occupational stress have been widely used and verified in different industries under various cultural backgrounds, they still have some limitations and/or difficulties in theory and application. Therefore, this paper defines the concept of stressor and stressor load, and puts forward the Stressor Load-Stress Resilience Model (SLSRM) from a new perspective. Based on this model, the perceived stress depends on severity of the stressor in the job context and individual ability to withstand stress (i.e. stress resilience). Stressors are classified into seven categories in this model, including the interpersonal relationship, organizational management, career development, job requirements, working environment, role ambiguity and work-family conflict. The individual stress resilience is determined by the positive personal characteristics, including psychological capital, knowledge/skill, working experiences and personality. Moreover, importance (weight) of stressor for different persons is also taken into account. An overview of a validation research is provided. This model provides a new approach for understanding and measuring the occupational stress and has the potential to be integrated into/developed to a complete stress management framework.

Keywords: Occupational stress model · Stressor · Stressor load · Personal characteristics · Stress resilience

1 Introduction

With the high-speed development and increasing intense competition in the modern society, occupational stress has made a more and more important effect on human health and well-being. Studies show that the occupational stress has significant effects on the staff performance, mental health, job satisfaction, and job burnout, etc. [1, 2]. Occupational stress may cause health hazards and lead to physiological diseases, which have also been certified by many studies [3–5]. Moreover, for some special occupations such as the pilot and doctor, whose occupational stress may not only influences the

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 232–243, 2017.
DOI: 10.1007/978-3-319-58472-0_19

individual, but also may threaten the public health and safety. It is considered that the air crash event of the airline Germanwings happened in 2015 was caused due to mental health of the pilot. However, one of the causes leading to his mental health is the over-high occupational stress possibly. Therefore, the occupational stress is taken more seriously by the enterprise and academia.

Currently, there are many kinds of stress theories and models, among which, the most important and empirically most successful ones include the Transactional Model [6, 7], Person-Environment Fit (PEF) [8, 9], Job Demand-Control (-Support) model (JDC(S)) [10–12] and model of Effort-Reward Imbalance at work (ERI) [13]. Although these models have been applied in a wide range of industries, but there are still some problems need to be solved. Transactional Model focuses too much on the individual appraisal process. It is not conducive to identify the stressors in the job context and provides little useful information for effective intervention. PEF model does not specify the particular content dimensions on which person and environment should be examined, and not determine the relationships between content dimensions, as well as their impact on stress. The requirement of commensurate dimensions also makes PEF model more difficult for application. JDC model confines stressors only on the aspect of work characteristic and only treats skill and decision power as control resource without consideration of the effect of personality and psychology capital. While the reward in ERI model is more of a variable as a result, which overlook the process nature of stress. The consequent balance between effort and reward does not mean there is no stress during the work process.

In recent years, the research on occupational stress mainly focuses on the application, reliability and validity of existing stress models in different industries and cultural background [14–16], and some certain extent of improvement, richness and development for the existing models [17–21]. These efforts did not offer us a new perspective to understand and study stress.

The above reasons motivate the effort in this paper. Based on the concept of stressors, this paper puts forward a novel model illustrating the formation mechanism of occupational stress, which was intended to provide a different approach for the understanding and measurement of occupational stress. In this model, a basic assumption is that the individual abilities to withstand stress, determined by skill/knowledge, psychology capital and other personal characteristics, are different. And the relationship between the total stressor load in job context and the individual capacity to resist stress determines the perceived stress level. We do not claim to answer these questions mentioned above, but we want to demonstrate how they can be approached in a somewhat different conceptual framework.

The following contents of this paper are arranged as follows. In Part II, from three aspects of stressor and stressor load, stress resilience, and stressful situations, the assumption, concept and framework of the model proposed in this paper are introduced. In part III, an overview of the development of a special occupational stress scale for Chinese pilots is presented to illustrate the rationality and validity of the SLSRM model. In Part IV, the problems still need to be resolved and the advantages of the proposed model. Part V gives the conclusion and points out the ongoing and future efforts.

2 Stressor Loads-Stress Resilience Model

As per the model put forward in this paper, the perceived occupational stress is co-determined by the stressor in working context and stress resilience (personal characteristic) of the employee. One's coping mode can regulate the psychological, physiological and behavior influences exerted by the perceived stress. The organizational intervention impacts all aspects of the stress indirectly, as shown in Fig. 1. Because the models proposed in this paper focus on the stress production mechanism, the influences from coping and organizational intervention as shown in Fig. 1 are not discussed in details herein. The stress production mechanism is based on the stressor load perspective. Stressor are all factors that may make employees perceive the stress in the occupational context, including the job, organizational management, career development, interpersonal relationship, work-family conflict, etc. Moreover, the stressor has the potential capability for stress production, which is defined as stressor load. The model is based on such a hypothesis that one has a definite stress-withstanding capability that is decided by personal characteristics such as knowledge/skill, psychological capital, personality traits, etc. The stress level perceived by the employee is co-determined by the total stressor load in the working context and individual stress-withstanding capability. The higher the total load is and the lower the individual stress-withstanding capability is, the more stress the employee will experience. The detailed introduction to the stress production mechanism (stressors and personal characteristics) will be provided in the following section.

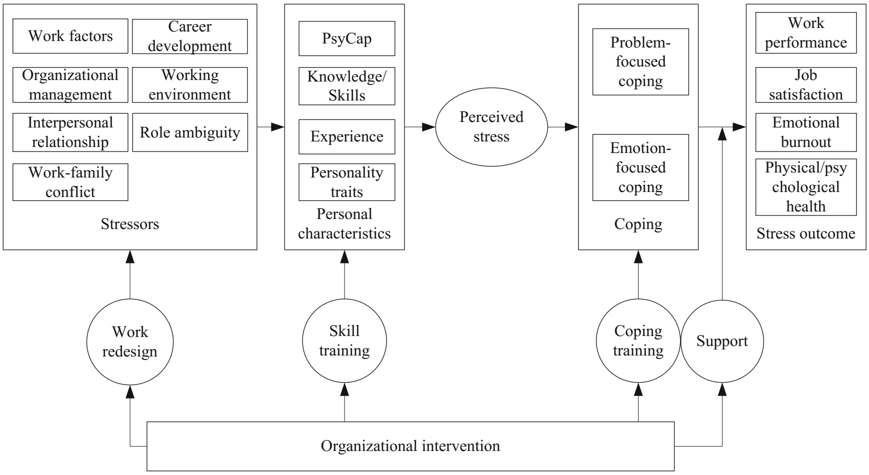


Fig. 1. A novel perspective for occupational stress formation and management

2.1 Stressor and Stressor Load

Stressor and stressor load definition. Stress is formed due to the complex mutual effect between the human and his environment (including physical and social environments). Existing stress models emphasize different aspects of transaction between the person and the environment, but they all involve the stressor concept in direct or indirect form. According to Lazarus's cognitive theory of emotion, cognitive appraisal or evaluation of an experienced stressor precedes any form of emotional response [22]. Karasek regarded the stressor in the job context such as workload requirement as the job demand [10]. In the JDCS model, support is believed could buffer the effect of stressors on strain [12]. Cohen & Edwards regarded the environment as stressors and researched the moderating effect of personality on relationship between the environment stressors and the health. In the PEF model, environmental requirements, internal and external sources and rewards are similar to the concept of the stressor [9].

However, these models give emphasis on "fit", "imbalance" or "congruence", they do not provide a concrete definition for stressor. Some researches give the definition of stressor but do not reach a consensus. Caplan thought that job stressors were certain job environment characteristics which can threaten the individual, such as workload, job complexity, and role ambiguity, etc. [23]. Jamal thought that job stressors were threats existing in the job context which require certain response to be taken. Tache & Selye held that all life events cause some stress, and the stressor is the stimulus eliciting a need for adaptation [24]. McEwen and Mendelson [25, 26] believed that a stressor was an event that challenged homeostasis, with disease the consequence of failure of the normal adaptive system.

The stressor is the basis of the model proposed in this paper. Therefore, the definition of the stressor is provided here: A stressor means a threat in the occupational context, including any factor that may cause the employee to perceive the stress. In order to present the potential ability for stress production of these stressors, this paper introduces a definition of stressor load, which is elicited from the workload concept in ergonomics. The higher the stressor potential is, the higher the corresponding stressor load is. We suppose the loads of independent stressors in the environment can be added herein, so a sum of all stressor loads is the total stressor load (TSL) in this occupation. The total stressor load mentioned in this model only reflects the capability for generating stress in one occupation and does not indicate the stress level perceived by the employee actually, which is different from methods used by Morano, Riolli, etc. [27, 28]. They gained grades for all stressors by means of self-rating stressor questionnaires and the grade sum served as the employees' stress level directly. Giving the definition of the stressor and the stressor load provides the conceptual foundation for the model proposed in this paper, as well as helps to understand the similarities and differences between existing stress models and theories.

Stressor identification and category. The identification and classification of stressors are correlated to the specific research and occupation closely. Houghton et al. [29] and Roddenberry & Renk [30] made studies on main stressors of undergraduates, ranging from the demands of their academic coursework to challenges in managing interpersonal relationships. Carmen Morano [27] identified 14 problematic behaviors as stressors of disease caregivers. Hatton et al. developed the Staff Stressor Questionnaire (SSQ) and distinguished seven types of stressor in Services for Adults with Intellectual Disabilities, including user challenging behavior, poor user skills, lack of staff support, lack of resources, low-status job, bureaucracy and work-family conflict [31]. Cooper's four-way model [32] and Robbins's "stressor-perceived stress-stress result" model [33] are general occupational stress models. The four-way model divides the occupational stressor into factors intrinsic to the job, the managerial role, relationships with other people, career and achievement, organizational structure and climate, and home/work interface. Robbins distributed the stressor into environmental factor, organizational factor and personal factor. On the basis of literature review and stressor definition provided herein, the stressor in the model of this paper falls into interpersonal relationship, organizational management, career development, job requirements, working environment, role ambiguity and work-family conflict, with personal factors eliminated. To be mentioned here, most researches or models do not take working environment as a stressor. However, we consider working environment such as bad weather is a key factor in affecting employees' workload for air transportation, offshore oil platform, etc., therefore, the working environment should be taken as a stressor.

Total stressor load. Robins points out that stress factors are addable, and that each new and continuous stress factor will enhance the individual stress level. In fact, one stressor may be insignificant. However, if it is added to a very high stress level, it may become "the last straw". In our model, the total stressor load (TSL) in a specific occupation context is a summation of all independent stressor loads.

However, just as pointed out by Edwards when commentating the PEF model, demands-ability and needs-supply match are considerably more relevant to people when the stimuli are important to them. Edwards has referred to this as dimension importance [12]. We also reckon the stressor means different importance for different persons, for example, someone thinks highly of career development and promotion while others may value family. Thereby the same stressor load from career development and work-family conflict may cause different stress perception for different persons. We advise to take the stressor weight into consideration, which is attained from personal assessment as a representative for individual difference. The total stressor load should be a weighted sum of individual stressor loads.

2.2 Stress Resilience

The stress resilience is defined as a higher-order construct composed of different positive personal characteristics, which indicates one's ability for maintaining high efficiency, health and well-being under stress. In this paper, the term stress resilience and the term ability to withstand stress can be used alternately. The model hypothesizes a definite ability to withstand stress of one person and the stress perceived by one person depends on the suffered total stressor load and comparison of ability to withstand stress. When TSL exceed one's ability to withstand stress, one will be confronted with high stress levels. We think everyone has a certain resilience to stress. The amount of stress a person can withstand depends very much on the individual. Stress-resilient personal characteristics cover psychological capital, knowledge/skill, experience and personality. In our opinion, the person with excellent knowledge/skill, rich experience, high-scored psychological capital and positive personality can resist more stress, thus withstanding higher stressor load. The model in this paper differs from many stress theories and researches taking personal characteristic as a medium or mediation variable between the stress and result in such aspect that it uses the personal characteristic as a direct factor for determining the perceived stress that results in an adverse stress outcomes.

Psychological capital. Initial research in the Industrial-organizational psychology field confirms a positive relationship between Psychological capital (PsyCap) and well-being [34] as well as other important work attitudes, behaviors, and performance [35]. PsyCap, is a meta-concept that incorporates various traits that have been found to foster psychological resilience. PsyCap is defined as:

> "*an individual's positive psychological state of development and is characterized by: (1) having confidence (self-efficacy) to take on and put in the necessary effort to succeed at challenging tasks; (2) making a positive attribution (optimism) about succeeding now and in the future; (3) persevering toward goals and, when necessary, redirecting paths to goals (hope) in order to succeed; and (4) when beset by problems and adversity, sustaining and bouncing back and even beyond (resilience) to attain success.*" [36]

According to this definition, psychological capital is conceptualized as a combination of efficacy, optimism, resilience, and hope. Each individual construct of optimism, hope, efficacy, and ego resiliency is imperfect in representing general resilience to stress, and thus their common factor (PsyCap) should provide a more complete index of the domain.

In most researches, PsyCap is applied to the world of work has been hypothesized to aid employees cope with stressors, or play a role of mediator or moderator in the workplace. In accordance with Riolli et al. [28], among undergraduate students from a university, PsyCap mediated between stress and indices of psychological and physical well-being. In the case of Psychological Symptoms and Health Problems, PsyCap buffered the impact of stress so that the relationship between stress and negative outcomes was reduced. In a word, psychological capital buffered stressors with the negative stress outcomes, and augmented the positive outcome. We reckon that the mechanism for these mediating effects is that PsyCap is a key factor in deciding the resilience to stress and only the actually-perceived stress exerts a positive or negative influence on the employee, which is consistent with Riolli's opinion [28], saying that persons high in PsyCap will more readily withstand stress and maintain physical and psychological well-being and happiness in the face of stress.

Knowledge/skill and experience. In PEF model, knowledge/skills are regarded as a part of one's ability for satisfying environmental requirements [9]. When mismatched demands-abilities affect the reception of supplies, the stress is generated. According to Edwords, excess abilities may decrease strain by providing supplies for needs, as when being able to complete one's work more quickly than required creates time for reading, socializing, or other pleasurable activities. Alternately, excess abilities may decrease strain by allowing the person to conserve personal resources (e.g., time, energy) to apply toward future demands. In the model put forward herein, we think that sufficient knowledge/skill is necessary for dealing with difficulties and challenges in the job and assisting employees coping pertinent stressor load such as job requirements, etc., with ease. Consequently, it should be as a favorable factor for stress resilience.

Working experience is an important factor in workload and human reliability analysis (HRA) research field, but it is rarely applied in occupational stress research. However, in Robins's stress model, working experience is served as a part of individual difference and served to regulate the stress perceived by one. Based on Robins, the new and uncertain job context brings people stress, but the stress will disappear or decrease greatly when the working experience is attained. The working experience, as a good stress reducer, is in inverse proportion to working stress. For this phenomenon, Robbins considered there are two causes: one for selective withdrawal, indicating that a person with strong stress perception is prone to job-hopping automatically while those persons working in the organization for a long term possess higher stress-withstanding capabilities or could bear more of the kind of stress in the organization where they work; the other is that an anti-stress mechanism will be generated with time, which takes a long period of time. Therefore, seniors in the organization have a stronger adapting capability and perceive little stress [33]. We believe experienced employees are familiar with the work process and ready for challenges or changes possibly occurred in the job and can endure higher stressor load under the same condition.

Personality traits. Personality traits are regarded as regulation variables for the stress in many stress theories, whereas the Type A personality in the personality traits and Locus of Control applied as the moderating variable in OSI scale [37]. The individual difference in Robins' stress model also involves the Locus of Control [33]. The personality, which is a power system for physical and psychological system inside a person, decides the special regulating mode of one person for environmental characteristics. Personality traits are stable personality characteristics represented by one person in different contexts, which cover gumption, ambition, etc.

Locus of Control shows one's attitude towards life; for example, some regard themselves as the master of destiny with control right, which are called internals; while others think they are controlled by external forces and all changes in life rely on fortunate and opportunity, which are called externals. It is researched that the internals search for information positively before making decision and have strong motivation for achieving success, thus prone to control of the environment. Persons with this personality allow individuals to view events more positively, less negatively, so they do well in the job full of challenges and can withstand more stress.

If some persons are always willing to participate in high-strength competition activities and perceive time stress for a long term, they have Type A personality, who push themselves to do more work in the shortest time and attack other persons or events hindering their efforts, thereby resisting higher stressor load. Based on the opinion from the model put forward in this paper, personnel with Type A personality should be able to endure more stress theoretically.

2.3 Stressful Situations

As per this model, the perceived stress level (PSL) is determined by the total stressor load (TSL) and individual stress resilience (ISR). Their relationship is as shown in formulation (1):

$$\mathrm{PSL} = \mathrm{nTSL}/\mathrm{nISR} \tag{1}$$

In which, nTSL and nISR refer to normalized total stressor load and individual stress resilience respectively. The bigger the PSL value is, the higher the stress level experienced by the employee is. When PSL = 1, the employees are under full load. Based on PEF or ERI models, only when “misfit” or “imbalance” appears, there will be stressful experience. However, this model is different from the preceding ones and the PSL in this model is always greater than 0, indicating stress is existent all the time. Only when PSL is small enough, one will not experience the stress obviously. Only when PSL reach a certain value, the stress can be perceived and this point is referred as perceptible stress point (PSP). The specific value of the PSP should be determined via further researches, but we think it is a value less than 1, that is, employees should have experienced the stress before they are under full load. Before PSL reaches 1, the stress is good to the job performance. However, it will exert an adverse influence on the job performance after exceeding 1, that is, the inverted U relationship between the stress and performance appears and “PSL = 1” is the critical point. However, long-term full load or near to full load will exert adverse effect on psychological and physical health and wellbeing.

3 Occupational Stress Survey Using SLSRM Based Scale

To illustrate the rationality and validity of the SLSRM model, a special occupational stress scale for Chinese pilots is developed based on this mode and a 385 pilots joined stress investigation is conducted then. The special scale, named “Chinese civil aviation pilot occupational stress scale”, included two sub-scales, as the stressor subscale and the individual characteristics subscale. The original list of stressors were derived from open-ended questionnaire survey and review of relevant literatures. Then they were tailored by factor analysis after the large-scale investigation. The questionnaire used a six-point Likert-type Scale, 6-1 represented for strongly agree, agree, agree a bit, disagree a bit, disagree and strongly disagree, respectively.

Modified by item analysis, the final occupational stress subscale contains 64 items classified into 7 dimensions, including interpersonal relationship, organizational management, intrinsic to the job, work environment, occupational role, family factor, and career development were extracted. The personal characteristic subscale of this questionnaire was formed by the combination of the psychological capital scale compiled by Luthans and the self-compiled personal ability scale, which was consisted of 40 items.

The procedure of questionnaire score calculation includes three steps: (a) calculate the sum scores of each subject of the subscale respectively; (b) score normalization process; (c) calculate the ratio of normalization score obtained by the scale of pressure source and the scale of the personal characteristics. By fitting Minnesota Satisfaction short scale, the question score was adjusted.

KMO test was used for validity analysis. The KMO test results of stressors and personal characteristics subscales are 0.757 and 0.943 respectively and the significance of Bartlett test results of the two scales are less than 0.01, which means the score met the requirements of factor analysis. For reliability analysis, two estimates of the reliability of the occupational scale for Chinese Civil Pilots were determined: split-half and internal consistency. The results show that the questionnaire has a good consistency reliability and indicate a satisfactory level of consistency among items. The result of the stress survey is compared and contrasted with that of the job satisfaction scale to check its validity and to demonstrate it can serve as alternative of other successful stress scales. Pearson correlation coefficient, chosen as the criterion validity, is –0.609, the correlation of external validity is significant.

The overview of the development of the Chinese civil aviation pilot occupational stress scale and the large-scale stress survey aims to illustrate the rationality and validity of the SLSRM model. The detail description of these efforts will be presented in another paper.

4 Discussion

The SL-SR model brought forward in this paper is based on the stressor, which is mentioned in most kinds of stress theories and models. Some researches even use the appraisal of stressors as the perceived stress, such as Staff Stressor Questionnaire, but do not define the stressor definitely due to emphasis on "appraisal", "fit", "imbalance", etc. Based on the consensus implied in those theories, models and researches, this paper provides a definition for the stressor and puts forward the concept of stressor loads for representing the severity of stressor. The proposed stressor and stressor load concepts will assist further understanding for commonness in current stress theories. The stress model put forward in this paper supplies a new route for knowing the production mechanism of occupational stress and has remarkable advantages in stress coping and intervention aspects. However, some questions should be taken into consideration during actual application of the model.

Firstly, it is the determination of stressor load. Self-report is a commonly used method in the stress research field, with which a reasonable outcome can be procured through approval by the application research. Consequently, the self-report will serve as a useable stress load measuring method. However, self-report measures of stressor loads always have already contained some degree of the actually-experienced stress by the employee and individual difference influence and do not reflect the stressor load objectively. It might be predicted that persons with high-score and low-score personal characteristics perceive identical stressor quite differently, with the former giving lower ratings for stressor simply because their ability to withstand stress render those events less troublesome. These problems do not invalidate the use of self-report measures, but it is important to include objective measures to guard against them。 Except bringing about the objective measures of stressor load, another method is to reduce the influence on individual characteristics by using elaborate stressor questionnaires. Taking the work complexity as an example, the question "Compared with other jobs, my job possesses more complexity" can be asked to the subject instead of "I think my job is very complex".

Secondly, it is the determination of stress resilience. This model present that the individual resilience to stress depends on the knowledge/skills, working experience, psychological capital and personality traits, but appraisals on those traits are not direct measurement for resilience to stress. Therefore, when PSL = 1, it does not necessarily mean balance between the total stressor load and individual resilience to stress. The real balance point may be determined via further researches.

Both of the above problems also exist in a great deal of stress theoretical models. For example, there is much less research assessing the congruence of self- and objective ratings for job demands than that for job decision latitude. A self-report of a "demanding" job on the indicator probably will also express an element of subjective perception of stress [38]. Though above-mentioned problems needs resolutions, this model offers a new route for effectively understanding the stress production process, and compared with other models, can be easily integrated with occupational stress coping and intervention to constitute a complete occupational stress management framework, as shown in Fig. 1.

Stress models such as Transactional Model, PEF model, JDC(S) model and ERI model pay more attention to the stress production process. They are not complete stress management models because they do not provide more detailed operational suggestions for stress management method. At present, the researches on relationship between the occupational stress and result variable, and functions of the mediating variables and moderating variables are always independent from stress model. The relationships between different variables are researched individually [2, 27, 39–42] and a systematic complete stress production and management models are not formed. In order to construct a complete stress management model, management practices and operational connotations of stress model should be enriched from individual and organizational dimensionalities besides researches on stress production and influence mechanism.

5 Conclusion

This paper is the initial effort of developing a novel occupational stress model which is motivated by the problems emerged in existing stress models and theories. The fundamental concepts of the SLSRM model, stressor and stress resilience, are extracted from the common elements of existing occupational stress models and theories. The proposed model may not able to answer all these questions mentioned above, but we try to provide a different approach for understanding and control occupational stress. Furthermore, the SLSRM model can give practical recommendations for stress management and has the convenience to be developed to a comprehensive stress management framework.

Stress management aims at effective integration and dispelling of different occupational stress factors, regulation of mental strain, creation of excellent occupational stress environment and turning of stress into power, thus allowing employees to work efficiently in a healthy occupational stress status under continuous stimulus. The established structural cognition and analysis frameworks for the occupational stress will help find out possible causes for the stress and resolutions to facilitate systematic stress management in the organization. In future work, we will improve the stress scale based on the survey result, identify useable coping measures and organizational intervention plans, research on methods for assessing their influences and effects, and constitute a complete occupational stress assessment and management framework eventually.

References

1. Pan, B., Shen, X., Li, L., Yang, Y., Wang, L.: Factors associated with job satisfaction among university teachers in Northeastern region of China: a cross-sectional study. Int. J. Environ. Res. Public Health **12**, 12761–12775 (2014)
2. Li, X.R., et al.: The mediating role of psychological capital on the association between occupational stress and job burnout among bank employees in China. Int. J. Environ. Res. Public Health **12**, 2984–3001 (2015)
3. Lundberg, U., Cooper, C.L., Marmot, M.G.: The science of occupational health: stress, psychobiology and the new world of work. J. Ment. Health **20**, 210–214 (2011)
4. Wirtz, P.H., Siegrist, J., Rimmele, U., Ehlert, U.: Higher overcommitment to work is associated with lower norepinephrine secretion before and after acute psychosocial stress in men. Psychoneuroendocrino **33**, 92–99 (2008)
5. Li, J., Yang, W., Cho, S.I.: Gender differences in job strain, effort-reward imbalance, and health functioning among Chinese physicians. Soc. Sci. Med. **62**, 1066–1077 (2006)
6. Lazarus, R.S.: Stress and Emotion: A New Synthesis. Free Association, London (1999)
7. Lazarus, R.S.: Toward better research on stress and coping. Am. Psychol. **55**, 665–673 (2000)
8. French, J.R.P., Caplan, R.D., Harrison, R.V.: The mechanisms of job stress and strain. Wiley, London (1982)
9. Edwards, J.R., Caplan, R.D., Harrison, R.V.: Person-environment fit theory: conceptual foundations, empirical evidence, and directions for future research. In: Cooper, C.L. (ed.) Theories of Organizational Stress, pp. 28–67. Oxford University Press, Oxford (1998)
10. Karasek, R.A.: Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign. Adm. Sci. Q. **24**, 285–308 (1979)
11. Karasek, R.A., Theorell, T.: Healthy Work: Stress, Productivity and the Reconstruction of Working Life. Basic Books, New York (1990)
12. Johnson, J.V., Hall, E.M.: Job strain, work place social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. Am. J. Public Health **78**, 1336–1342 (1988)
13. Siegrist, J.: Adverse health effects of high-effort/low-reward conditions. J. Occup. Health Psychol. **1**, 27–41 (1996)
14. Lien, L.L., Steggell, C.D., Iwarsson, S.: Adaptive strategies and person-environment fit among functionally limited older adults aging in place: a mixed methods approach. Int. J. Environ. Res. Public Health **12**, 11954–11974 (2015)
15. Li, J., Yang, W.J., Cheng, Y.W., Siegrist, J., Cho, S.I.: Effort-reward imbalance at work and job dissatisfaction in Chinese healthcare workers: a validation study. Int. Arch. Occup. Environ. Health **78**, 198–204 (2005)
16. Cheng, Y., Luh, W.M., Guo, Y.L.: Reliability and validity of the Chinese version of the job content questionnaire (C-JCQ) in Taiwanese workers. Int. J. Behav. Med. **65**, 1247–1251 (2010)
17. Demerouti, E., Bakker, A.B., Nachreiner, F., Schaufeli, W.B.: The job demands-resources model of burnout. J. Appl. Psychol. **86**, 499–512 (2001)
18. Crawford, E.R., Lepine, J.A., Rich, B.L.: Linking job demands and resources to employee engagement and burnout: a theoretical extension and meta-analytic test. J. Appl. Psychol. **95**, 834–848 (2010)
19. Van den Broeck, A., De Cuyper, N., De Witte, H., Vansteenkiste, M.: Not all job demands are equal: differentiating job hindrances and job challenges in the job demands-resources model. Eur. J. Work Organ. Psychol. **19**, 735–759 (2010)

20. Kinnunen, U., Feldt, T., Makikangas, A.: Testing the effort-reward imbalance model among finnish managers: the role of perceived organizational support. J. Occup. Health Psychol. **13**, 114–127 (2008)
21. Laschinger, H.S., Finegan, J.: Situational and dispositional predictors of nurse manager burnout: a time-lagged analysis. J. Nurs. Manag. **16**, 601–607 (2008)
22. Lazarus, R.S.: Emotion and Adaptation. Oxford University Press, New York (1991)
23. Caplan, R.D., Cobb, S., French, J.R.: Relationships of cessation of smoking with job stress, personality, and social support. J. Appl. Psychol. **60**, 211–219 (1975)
24. Taché, J., Selye, H.: On stress and coping mechanisms. Issues Ment. Health Nurs. **7**, 3–24 (1985)
25. Mcewen, B.S.: Protective and damaging effects of stress mediators. N. Engl. J. Med. **338**, 171–179 (1998)
26. Mcewen, B.S.: Allostasis and allostatic load: implications for neuropsychopharmacology. Neuropsychopharmacology **22**, 108–124 (2000)
27. Morano, C.L.: Appraisal and coping: moderators or mediators of stress in Alzheimers disease caregivers? Soc. Work Res. **27**, 116–128 (2003)
28. Riolli, L., Savicki, V., Richards, J.: Psychological capital as a buffer to student stress. Psychology **03**, 1202–1207 (2012)
29. Houghton, J.D., Wu, J.P., Jeffrey, L.G., Christopher, P.N., Charles, C.M.: Effective stress management. J. Manag. Educ. **36**, 220–238 (2012)
30. Roddenberry, A., Kimberly, R.: Locus of control and selfefficacy: potential mediators of stress, illness, and utilization of health services in college students. Child Psychiatry Hum. Dev. **41**, 353–370 (2010)
31. Hatton, C., et al.: Staff stressors and staff outcomes in services for adults with intellectual disabilities: the staff stressor questionnaire. Res. Dev. Disabil. **20**, 269–285 (1999)
32. Williams, S., Cooper, L.: Managing Workplace Stress: A Best Practice Blueprint. Wiley, West Sussex (2002)
33. Robbins, S.P., Judge, T.A.: Organizational Behavior, 15th edn. Prentice Hall, Englewood Cliffs (2012)
34. Culbertson, S.S., Fullagar, C.J., Mills, M.J.: Feeling good and doing great: the relationship between psychological capital and well-being. J. Occup. Health Psychol. **15**, 421–433 (2010)
35. Avey, J.B., Reichard, R.J., Luthans, F., Mhatre, K.H.: Meta-analysis of the impact of positive psychological capital on employee attitudes, behaviors, and performance. Hum. Resour. Dev. Q. **22**, 127–152 (2011)
36. Luthans, F., Youssef, C.M., Avolio, B.J.: Psychological Capital. Oxford University Press, New York (2007)
37. Lyne, K.D., Barrett, P.T., Williams, C., Coaley, K.: A psychometric evaluation of the occupational stress indicator. J. Occup. Organ. Psychol. **73**, 195–220 (2000)
38. Lazarus, R.S.: Psychological Stress and the Coping Press. McGraw-Hill, New York (1966)
39. Devereux, J.M., Hastings, R.P., Noone, S.J., Firth, A., Totsika, V.: Social support and coping as mediators or moderators of the impact of work stressors on burnout in intellectual disability support staff. Res. Dev. Disabil. Multi. J. **30**, 367–377 (2009)
40. Fried, Y., Shirom, A., Gilboa, S., Cooper, C.L.: The mediating effects of job satisfaction and propensity to leave on role stress-job performance relationships: combining meta-analysis and structural equation modeling. Int. J. Stress Manag. **15**, 305–328 (2008)
41. Lu, L., Liu, L., Sui, G.Y., Wang, L.: The associations of job stress and organizational identification with job satisfaction among Chinese police officers: the mediating role of psychological capital. Int. J. Environ. Res. Public Health **12**, 15088–15099 (2015)
42. Luthans, F., Avolio, B.J., Avey, J.B., Norman, S.M.: Psychological capital: measurement and relationship with performance and satisfaction. Pers. Psychol. **60**, 541–572 (2007)

Situation Awareness and Control

An Integrated Approach of Human Oriented Interactions with Complexity

Cedric Bach(✉), Viviane Perret, and Guillaume Calvet

Human Factors and Ergonomics Department,
Bertin Technologies, Toulouse, France
{cedric.bach, viviane.perret,
guillaume.calvet}@bertin.fr

Abstract. This paper presents a theoretical study in the field of Human Oriented approach of interactions with complexity. This study aims to contribute to one of the current challenges in Human Factors research consisting to provide methods and structured approaches to face the increasing complexity of ATM concepts. This theoretical study consists to articulate two models of complexity, previously published, in a single matrix able to delimit the scope of Human Oriented Approach of Complexity (HOAC). This matrix named the HOAC matrix aims to provide properties to classify methods and to structure HF activities in order to support a systemic approach of the design of interactions improving a better integration of humans in complex systems. The HOAC matrix is illustrated by four examples of methods supporting the design of *simplex artefacts.*

Keywords: Human factors · Methods · Complexity · Systemic · User centered design · Framework

1 Introduction

One of the current challenges in the aeronautics field is to design systems and Human-machine interactions supporting human performance dealing with complexity characterized as a system of systems. Obviously, the automated and interconnected systems contribute to the overall safety of aeronautics [1]. But the evolution of these systems also increases, for instance, the number of parameters to monitor and the mediation of pilot control [14]. Pilot's role is increasingly focused on his expertise and adaptation capacity to manage high context variability including multidimensional complexity variations. Indeed novelties arisen in different fields contribute to increase the overall systems of systems network, but also open new investigation fields for the Human Factors (HF) research. Indeed the past decades produced a technological (r)evolution from simple artefact to interconnected and multimodal systems and this evolution has multiple side effects on the HF specialist job, tools and methods.

Thus, the current HF challenge is to design systems both: (a) structurally complex from a technological and organizational standpoint; and (b) conceptually simple from an operator standpoint. In other words, the aeronautics has to create interaction dealing with complexity or in short: simplex systems [17].

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 247–265, 2017.
DOI: 10.1007/978-3-319-58472-0_20

Today to ensure the overall performance of the Human-System Interactions, in aeronautical field the flight deck design is usually based on user-centered methods. These methods require a systemic approach of the dynamic interactions composing the whole activity of the pilots. This approach aims at diagnosing threats and vulnerabilities parameters involved in the activity. Nevertheless, an activity, as the pilot's role, is composed of large number of elements, exchanged and challenged during operation. Therefore one of the major difficulties for a Human Oriented designer is to compose a realistic model of an activity under analysis. The application of user-centered design in this field is usually efficient to produce materials describing the current and future operations (e.g. by use of *task models*, *interviews*, *design exposure*, *authority allocation…*). Of course, these different outcomes will reflect the complexity of the activity under study. Therefore the role of the Human Oriented designer will consist in organizing a variability of points of view in a coherent and realistic manner as a whole. Usually, at this point, the designer faces major issues: How to transform outcomes from user-centered methods as part of the project the designer is working for? Actually, the way to adapt user centered knowledge to a particular project will also strongly depend on the purpose of this project. In short, all this variability creates difficulties to generalize knowledge related to design and assessment of simplex systems. At this point the HF research field aims to provide some integrated approaches of HF methods [11]. Nevertheless the HF field is far to propose a standardized methodological framework to harness the complexity linked to the wide evolution of systems of systems.

Therefore the purpose of this paper is to expose a structured view of a Human Oriented approach of interactions with complexity. Most precisely, this paper exposes the result of a theoretical study providing a matrix supporting a Human Oriented systemic approach of the design and evaluation of interactions dealing with complexity. Actually this matrix, named Human Oriented Approach of Complexity matrix (HOAC matrix), can reflect the organization of methodologies currently developed or used to harness the design and evaluation of simplex systems. So, this paper presents both the details of the theoretical study conducting the HOAC matrix then illustrated with a panel of methods applied or under (i.e. *simplexity criteria* [16], *authority design* [12], *eye-tracking evaluation*, *complexity simulator*).

2 Theoretical Background

This section provides the main theoretical backgrounds supporting the ideas presented in this paper. This background involves both the *paradigm of complexity* and the emerging theoretical concept of *simplexity*. Of course the concept of *simplexity* is itself partially based on the concept of usability especially in the field of Human Machine Interactions. But we prefer the concept of *simplexity* to highlight the involvement of complexity characteristics in the design of simplex artefacts. Of course the design of these simplex artefacts shall include the main, and now well known, standardized principles of usability [13].

2.1 Paradigm of Complexity and Complexity Characteristics

Our postulate is to consider that complexity requires an adapted approach and should involve a set of methods, tools and theoretical backgrounds in line with and supporting such complexity. Indeed we consider that complexity is not only a dimension or a continuum between "simple" and "complex". Actually we consider complexity as a part of complexity sciences in such way complexity should be considered with its own characteristics or criteria. Therefore complexity could be viewed as a paradigm, the paradigm of complexity [15]. Through this paradigm, the notion of agents can be used to characterize that an activity is conducted by a defined entity either human or technical within an organizational setting [8]. The *paradigm of complexity* can help to generalize the notion of activity independently of the type of agent involved. To achieve expected performance, agents gather information from their external or internal environments, analyze and take decisions and implement and monitor the outcome of their decisions through a variety of cognitive or physical actions. Achieving performance depends also on having the adequate information on other agent's activity and knowing how to collaborate. Finally, the situational, societal and organizational characteristics of agents' environments determine the way how actions are performed in addition to the specificities of each entity, such as for example previous experience or individual motivations.

Cilliers [7] presents the following characteristics of complex systems. These characteristics match with the ATM ones:

- The number of elements is sufficiently large that conventional descriptions (e.g. a system of differential equations) are not only impractical, but cease to assist in understanding the system. Moreover, the elements interact dynamically, and the interactions can be physical or involve the exchange of information.
- Such interactions are rich, i.e. any element or sub-system in the system is affected by and affects several other elements or sub-systems. In this way ATM is usually described as a system of sytems.
- The interactions are non-linear: small changes in inputs, physical interactions or stimuli can cause large effects or very significant changes in outputs.
- Interactions are primarily but not exclusively with immediate neighbours and the nature of the influence is modulated.
- Any interaction can feed back onto itself directly or after a number of intervening stages. Such feedback can vary in quality. This is known as recurrence.
- Such systems may be open and it may be difficult or impossible to define system boundaries.
- Complex systems operate under far from equilibrium conditions. There has to be a constant flow of energy to maintain the organization of the system.
- Complex systems have a history. They evolve and their past is co-responsible for their present behaviour.
- Elements in the system may be ignorant of the behaviour of the system as a whole, responding only to the information or physical stimuli available to them locally.

2.2 Human Oriented Approach of Complexity: A Step Toward Simplex Artefacts

Human oriented approach as User Centred Design (UCD) places humans at the centre. But beyond an orientation of a design process, a Human Oriented approach of complex systems should be viewed as a prerequisite for the overall system optimization because Humans are already a great Embodied Complex Adaptive System. In short, we need humans in large complex systems because they are able to increase the overall efficiency of a system on the long term. In this way a Human agent should be viewed as an agent with a specific characteristic which is its embodiment [19].

The aim of a Human oriented design process, dealing with interactions with *complexity*, is to provide to Humans simplex interactions absorbing a part of environment, organizations, human or tasks complexity. One can name such products *Simplex artefacts*. These simplex artefacts are mandatory to ensure an efficient involvement of Humans in complex systems such as future ATM concepts.

3 Toward an Integrated Framework to Harness Human Oriented Approach of Complexity

To face such complexity the HF specialist needs an appropriate referential and adapted guidance material. The aim of this section is to present a global framework able to harness the main dimensions of the Human Oriented approach of complexity in aeronautics. This framework is mainly composed by two components, first a model describing the dimensions of Human Oriented complexity and secondly a set of layers on which the Human Oriented approach of complexity can apply. The following sections describe these two components.

3.1 A Pyramidal Model to Understand the Human Oriented Approach of Aeronautics Complexity

No consensus exists on a universal approach of complexity of pilot's activity, the different perspectives used (e.g., dynamic, sociotechnical system, cognitive complexity) take into account common parameters which are weighted according to the scope of the studies. These common parameters are grounded in the definitions of these notions closely related together. Complexity is, in this regard, associated with the intricate inter-twining or inter-connectivity of elements within a system and between a system and its environment [9]. In order to approach the operational complexity of pilots, this section presents (a) the methodology deployed to define our conceptual model; (b) the main components of our model to describe the operational context; and (c) the main dimensions of our model which may be applied on the components and their interactions to describe the operational situation.

Methodology. This model is based on a qualitative data analysis and has been previously presented in [16]. The data come first from the scientific literature related to the design of complex systems (e.g., on the complexity definition, on the Human-Automatism cooperation) and secondly from aeronautical operational reports. So, on the one hand, the model presented in this section refers to three models: the AUTOS pyramid [4], the STS model [18] and the task complexity model [14]. On the other hand, in order to structure the knowledge, a set of 810 recommendations have been extracted from gathered data and classified by complexity dimensions. All along this work, the content and formalism of the recommendations as well as of the classification and construction of it have been reviewed by Human Factors Experts.

Main Components Interacting in Operations. The first part of the presentation of the model is based on the description of the main components in interaction during pilots' activity (see Fig. 1). From a systemic point of view, it is generally admitted that user's performance will depend on the interactions between his/her own characteristics, the characteristics of the artefact(s), of the task(s) and of the environment(s) involved. Being Human-Oriented, the description of the main components of our model is based on two perspectives: environmentalism and interactionism.

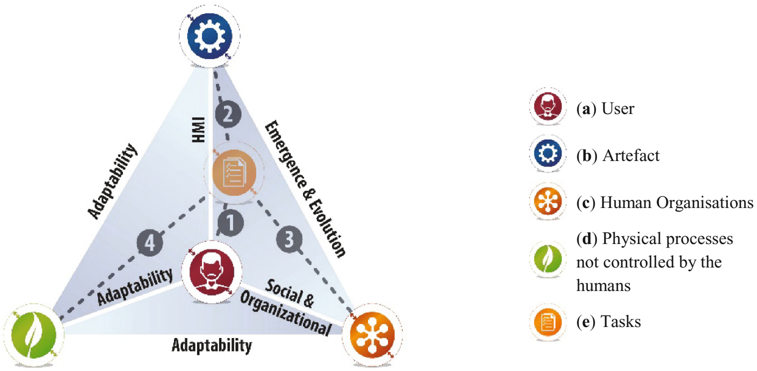


Fig. 1. Integrated view on the main concepts of the framework to understand a Human-Oriented complexity

On the one hand, environments may be defined in terms of internal and external standpoints with regard to user's activity. Usually, the internal environments refer to the resources available while the external environments refer to the elements which may have a direct or indirect incidence (positive or negative) on the user's activity. The difficulty in aeronautical domain is related to, for instance, the weather which is at the same time a resource for pilots' activity (e.g. fuel consumption strategy) and may disturb their activity (e.g. weather avoidance). So, to be adapted to pilots' activity, the viewpoints will be more dependent here on his/her direct (internal) or indirect (external) physical interactions with his/her environments regardless the interaction modality used (e.g., visual, auditory). More precisely, according to the components' references illustrated in Fig. 1, first, the internal environments of the activity refer here both to physical and operational properties of (a) the user or human operator and (b) the artefact such as computer, command and control system. Then, the external environments of the activity refer both to the characteristics of the social and infrastructural (c) organisations controlled by the human such as governmental organizations, airline, Air traffic controller and the atmospheric and topographical characteristics of (d) the physical processes not controlled by the human such as weather, birds, volcanic ash, landscapes.

On the other hand, from the perspective of the interactions involved, our model takes into account the direct interactions between the components previously presented [4]. For instance, the icing conditions may freeze the probes of the artefact. With respect to the interactions directly dependant from the tasks, the teleological, procedural, informational, spatial and temporal characteristics of (e) the tasks component are a mean to access to a set of parameters involved in pilots' activity because this component is the core of complex interactions. The characteristics of the tasks component are also a mean to analyse the interactions between the components. According to the references illustrated in Fig. 1, the analyses can refer to: (1) the task and activity requirements like the goals, action level, user's experience level [4, 18]; (2) the information and technological requirements like the automation and/or computerisation level [4, 18]; (3) the human organisation requirements like the responsibility, authority, role, operational procedure, training [4, 18]; (4) the adaptability requirements like the level of adjustment and temporal constraints.

The instantiation of these main components (i.e., User, Artefact, Human Organisations, Physical processes not controlled by the human and Tasks) in a space-time relationship allows the description of the operational context.

The complexity overview of the operational situation requires to apply some dimensions of complexity on each of the main components as well as on each of their interactions. The section hereafter explains these dimensions.

Main Dimensions Impacting the Complexity of Operations. The second part of the presentation of the model is therefore based on seven abstract parameters that may impact the complexity state of the operational situation. These dimensions extracted from the qualitative data analysis are: (a) *Variability* referring to the unstable characteristics of the elements [14] and to the nature of their influence on each other. For instance, small changes in inputs can cause large effects in outputs [7]; (b) *Ambiguity* referring to the lack of a perfect knowledge about all the elements [7, 14] like the cause-effect chain or else the behaviour of the system as a whole; (c) *Occurrence* referring to non-routine events and operations caused, for instance, by interruption [14]; (d) *Discrepancy* referring to the deviation of a component from the prescription and/or norm; (e) *Self-regulation* referring to the process permanent evolution like the own dynamic of a process or else the self-organisation of the elements [7, 17]; (f) *Dependency* referring to the relationship between the elements depending on their degree of direct dependence [7]. The relationship may be conflictual/negative [14] and/or positive when the elements have common characteristics such as references, history [7]; and (g) *Quantity* referring to the number of elements and/or interactions in their individual characteristics and in their relation to each other like the number of distinguishable properties, number of elements per unit area or unit volume for a given time. Quantity refers also to the quantity of resources required [14] like cognitive and/or physical resources, energy resources.

The interactions between the main components and their dimensions situated in a space-time relationship allow the description of the operational situation. In other words the situation complexity emerges from the interactions between the different components described above. Nevertheless, this model refers to the complexity of local execution, but the HF specialist designing a new operational concept or function have also to deal with other dimensions that are important to consider in an overall Human Oriented design process. We describe these different dimensions as a set of different layers in the next section. Actually these layers express different instantiations of the pyramidal model through layers with different proprieties.

3.2 Layers to Design Human Oriented Interactions with Complexity

Since agents in complex systems execute actions in collaboration and cooperation with other entities, the organization of these entities needs to be described more in detail to understand the wide-ranging impact of actions. Actions can be conducted by human or technical agents that belong to the same or different spatial, temporal or social units. These units can be characterized along layers ranging from a detailed Nano-Layer to a high-level Macro Layer. This differentiation is quite common in different domains such as economy, sociology as well as ecology. However additional levels were required for complex interactive systems and have been previously described in the field of incident report [20] and adapted to the context of authority sharing in ATM [12] introducing a NANO-layer. In line with the concept introduced by Bonfenbrenner [12], the Micro-layer covers a pattern of activities and roles in a specific context and can be understood as linked to an individual unit; the Meso-layer addresses a setting in which an individual engages for a certain amount of time, and the Macro-layer entails social and cultural values that exert strong influence. An agent may be connected to different layers at the same time. Depending on the layer where an agent is acting, its impact is different and in consequence the HF specialist may describe the interaction with complexity through the characteristic and scale of the corresponding layer. For example if the HF specialist is designing an authority philosophy (s)he could consider the interaction with an HMI at the Micro level and the definition of roles and responsibilities by authorities at a Macro level. On the same line the space and time of actions and design are impacted by the different layers. For example at the Micro level space and time may be the cockpit and tactical actions whereas at a Macro level the HF specialist should consider the impact of a new concept directly at a traffic scale (and none on a single aircraft) with long term perspective. The following sections provide instantiation of the layers properties by use of an example about authority and responsibility design, of course this example can be replicate following the same logic on other dimensions such as time, space or social units. The sections hereafter propose an adaptation of the layers used by different disciplines to the aeronautical domain.

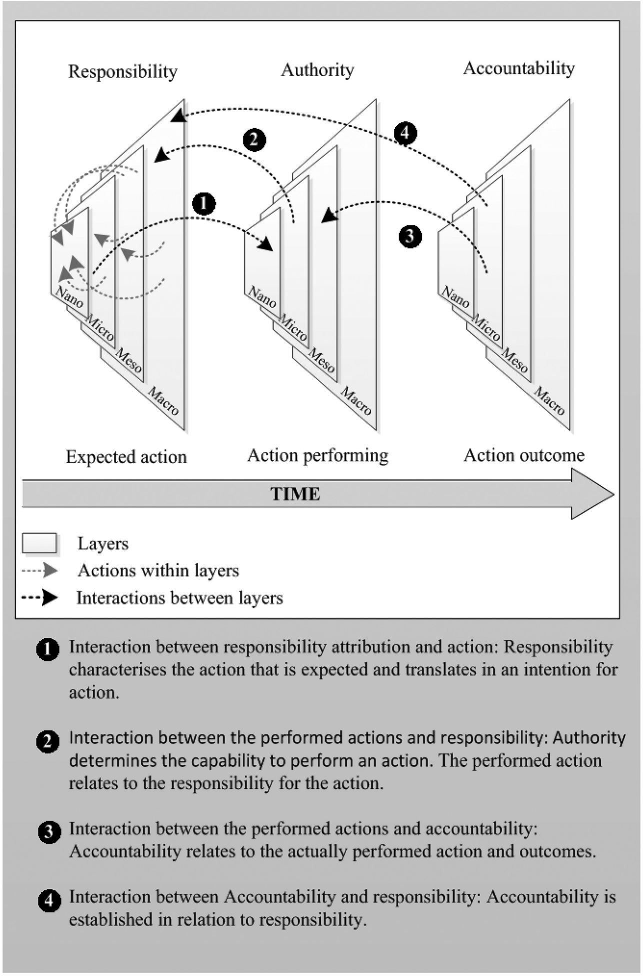
The **Macro-layer** usually characterizes large social units such as nations, globally acting institutions, cultures or societies (for example Eastern versus Western Societies) of groups of agents. The macro-level typically prescribes the responsibilities of the socio-technical components of the overall aviation system. Authorities (e.g. Military authorities, Civil authorities, FAA, EASA) usually design the regulations and global players such as major aircraft manufacturers, accident investigation offices, standardization groups, or the International Civil Aviation Organization (ICAO) may define the high-level principles of operations such as distance-based aircraft separations, air space organization, or free route. The rules defined at this level may impact the actions of agents either in a direct way (by applying the regulations) or indirectly (by applying the rules defined at a Meso-level). But also, at the Macro-layer actions and authority issues can be found, since the units of this layer are proper entities. The time frame for such activities is long and involves many contributors and can directly impact human activities, systems, organizations and the environment (e.g. CO_2 management, traffic capacity).

The **Meso-layer** refers to social units of an intermediate size, which can be articulated in form of specific organizations of groups of agents executing the high-level rules of operations defined at the Macro-level or sharing a common set of activities. Such organizations (e.g. Airlines, Industrial collaborations, Pilot/ATCO associations) usually define global missions for activities. The Meso level refers mainly to the management of both planning and execution phases of a mission (such as flight leg). Responsibilities at this level are related to an overall area, mission, flight, or work amplitude. The missions defined at this level directly impact the actions of agents. Equally, action and authority issues can also be found at this level. The time frame follows the same logic and is at a mission scale, the environment consideration is also at a mission level and is mainly supported by the forecasts (e.g. traffic density at arrival, weather).

The **Micro-layer** refers to the social unit of a single agent which can be either a human or a technical system. The events produced at this level usually refer to the execution phase. An example for authority at this level can be seen in an alerting system, which has the power to provide information or guidance. This level also deals with the allocation of authorities between human and machine (e.g. TCAS, Flight Mode Annunciator). The Micro-level should consider the properties of an agent such as its competences, strengths and weaknesses, and of course its responsibilities and authorities to perform actions. The time frame is tactical (e.g. less than 10 min of anticipation) and the environment considerations are based on detection by sensors.

Finally, the **Nano-level** refers to the specific properties (such as concrete knowledge, inputs, etc.) or components (such as system functions) that produce the actions. At Nano-level, also the definition of roles can be introduced. Roles are related to an aggregated set of connected actions, hence an agent may integrate different roles. An agent may play different roles depending on the level of the societal context he is involved with. The time frame and environment are immediate. Figure 2 characterizes how Layers can be linked to actions around responsibility/authority/accountability attribution and may be instantiated on a case by case basis through the different layers.

The attribution of responsibility/authority/accountability may also change over time. The example of the 2002 Ueberlingen accident reflects the change of the responsibility organization in relation to TCAS. This technical agent triggers instructions to pilots who have authority to execute a tactical avoidance manoeuvre. Before



① Interaction between responsibility attribution and action: Responsibility characterises the action that is expected and translates in an intention for action.

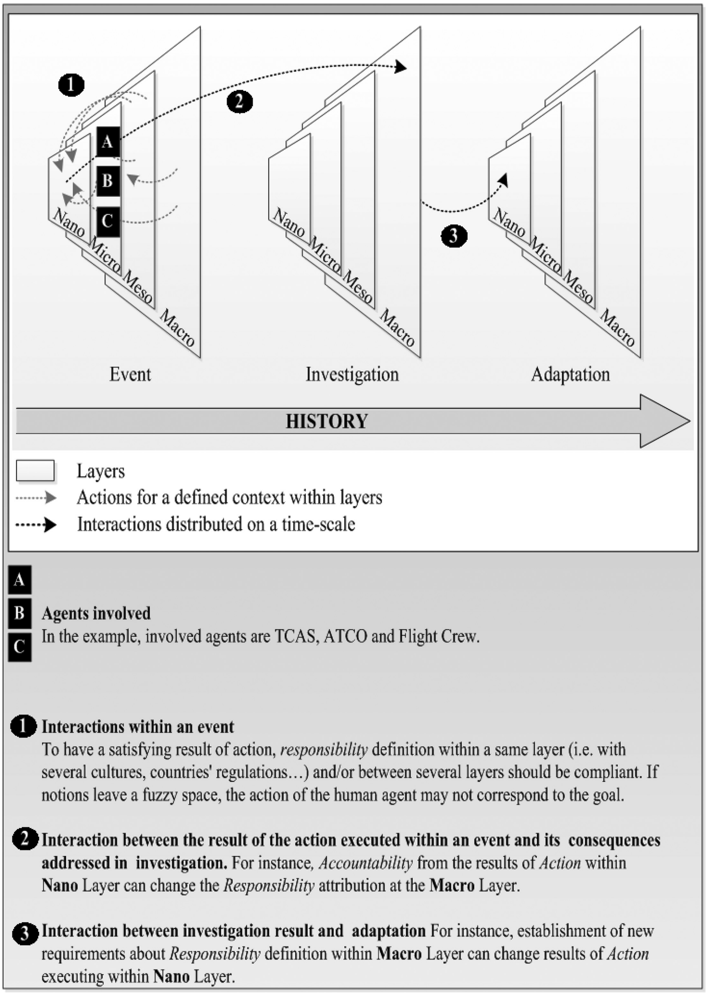
② Interaction between the performed actions and responsibility: Authority determines the capability to perform an action. The performed action relates to the responsibility for the action.

③ Interaction between the performed actions and accountability: Accountability relates to the actually performed action and outcomes.

④ Interaction between Accountability and responsibility: Accountability is established in relation to responsibility.

Fig. 2. Relationship between Authority, Responsibility, and Accountability based on the action concept

this accident, there was an absence of harmonization between German and Russian procedures at the Macro-level, one giving the authority to make decision to the TCAS agent and the other one to the ATCO. This lack of harmonization led pilots and ATCO, at the Nano-level, to follow different uncoordinated instructions during the execution of action causing a mid-air collision. After this event, the procedures were harmonized giving the authority to TCAS whatever the context. It means that an event that occurred at the Nano-level led to modifications at the Macro-level, which was in turn again resulting in modifications at the Nano level. Figure 3 illustrates the relationship between concepts.



Agents involved
In the example, involved agents are TCAS, ATCO and Flight Crew.

① **Interactions within an event**
To have a satisfying result of action, *responsibility* definition within a same layer (i.e. with several cultures, countries' regulations…) and/or between several layers should be compliant. If notions leave a fuzzy space, the action of the human agent may not correspond to the goal.

② **Interaction between the result of the action executed within an event and its consequences addressed in investigation.** For instance, *Accountability* from the results of *Action* within **Nano** Layer can change the *Responsibility* attribution at the **Macro** Layer.

③ **Interaction between investigation result and adaptation** For instance, establishment of new requirements about *Responsibility* definition within **Macro** Layer can change results of *Action* executing within **Nano** Layer.

Fig. 3. Dynamic relationships between Authority, Responsibility, and Accountability notions after an event across the different layers

4 A Framework for Human Oriented Interactions with Complexity

Obviously the components presented in the two previous sections could be mixed in order to draw a big picture of Human Oriented approach of complexity in aeronautics. In fact this is the main aim of this paper. Nevertheless the entire integration of these two components cannot be express in the frame of this presentation. Therefore we focus only on a part of the pyramid model presented in the Sect. 3.1. So we describe only the interaction between the Tasks and the vertex of the pyramid: HMI, Environment, Human Organization and Users. These four relations are numbered from 1 to 4 on the

Fig. 3. The following sections illustrate more in detail this mix between the interaction proper to the pyramid with the four layers Nano, Micro, Meso and Macro.

4.1 The Complexity Framework Mixing Pyramid Models and Layers

The main purpose of the complexity framework is to improve the awareness of the HF specialist regarding is own activity and also in the choice of the appropriate methodology in line with the level of his/her intervention. As we mentioned in the previous section we only consider the relation of the task with the vertex, of the pyramid model, with the different layers. Therefore we could consider the mix between these two components as a 4 * 4 matrix presenting in the lines the different relations to the task and in the column the four different layers. In the cells of this matrix might emerge the different HF activities addressing interactions with complexity and by extension the different methods and tools supporting these different activities. The Table 1 shows the result of this mix between the two components and presents a codification that we will use in the rest of the paper to guide the description of different methodologies able to cover the different cells of this matrix.

Table 1. Mix of pyramidal model and layers also called the Human Oriented Approach of Complexity matrix (HOAC matrix)

	Nano layer	Micro layer	Meso layer	Macro layer
1 task → User	1 (na)	1 (mi)	1 (me)	1 (ma)
2 task → HMI	2 (na)	2 (mi)	2 (me)	2 (ma)
3 task → Orga.	3 (na)	3 (mi)	3 (me)	3 (ma)
4 task → Envi.	4 (na)	4 (mi)	4 (me)	4 (ma)

The Table 2 provides some examples of instantiations through the HOAC matrix. These examples expect to guide the HF specialist in aeronautics to understand more concretely the link between his/her own (daily) activity with the different complexity dimensions of the HOAC framework. Indeed this instantiation reflect some activities in which the HF specialist, or the HF arguments, may be involved.

We can focus on some of these cells to highlight their content in terms of activities in which the HF specialist can be involved and their relations and independencies. As a case study we propose to follow the main needs involved in a new concept of Trajectory Based Operation in cruise phase. The TBO could be viewed only through a board system optimisation such as the improvement of vertical profile according to wind and temperature forecast in cruise for a single aircraft. But in this case the entire complexity of the concept is not considered because dimension as 4 (ma) the long term benefit on the traffic capacity is not considered as well as the (3 me) sharing of this TBO with the ground that may be have a strong positive impact on traffic predictability for ATC. Furthermore with a Human Oriented approach in design of simplex system the HF *transition factors* involving the elaboration of new (2 mi and 2 me) mental

Table 2. Examples of instantiations of the combination in the HOAC matrix

	Nano layer	Micro layer	Meso layer	Macro layer
1 task → User	concrete user actions/inputs for the task execution	properties of an agent such as its competences, strengths and weaknesses	users acting directly on planning and execution phases of a mission (a flight leg)	definition by authorities of users responsibilities in a task execution or long term impact of user on task and *vice versa*
2 task → HMI	concrete execution of components as system function	functional model supporting the immediate task execution at a tactical level	HMI supporting the mission execution at a leg level (e.g. FMS, ND, Display of ATC sector)	design of HMI interaction philosophy at a manufacturer level
3 task → Orga.	agent may embodied different roles and a set of aggregated set of connected actions	transfer of responsibility to an agent to other one according to defined procedure	co-execution, sharing a common set of operations at a mission level (ATC/pilot cooperation, pilots task sharing)	definition of rules or law structuring the overall inner and inter organisations
4 task → Envi.	concrete and direct action of the environment such as gust or flock of birds	adaptation of the task to the local sensed environment (constraints due to detected weather, terrain)	adaptation of the mission to the forecasted environment (weather forecast, escape routes)	study of the impact of task on the overall environment such as fuel consumption and traffic impact

representations of board and ground actors using this new TBO in cruise have to be considered, studied and rationalised. Otherwise this may negatively impact the acceptance of this concept both by 1 (ma) actors in execution phases but also by 3 (ma) organisation that can buy and maintain this new concept in operations.

In the case of these different dimensions are not considered in the design of that TBO concept, this may viewed as a way of simplification (on the contrary of simplex) of the design process and a way to win time in development on a short term. Nevertheless as this already mentioned in [3] if the whole needs of a new ATM concept are not considered or ignored in the early steps of a design project this may have a strong negative impact on the long term success of this concept because some important needs will be not identified whereas they still exist.

4.2 Examples of Methods Supporting the Human Oriented Approach of Complexity

This section expects to continue the illustration of the HOAC matrix but this time in terms of methods to support a Human oriented approach of complexity. On this line we introduce and describe four different methods candidate to be use or already supporting different cells of the complexity matrix. Obviously this set of methods is not exhaustive and is used only to illustrate the capacities of the HOAC matrix to delimit the scope of a Human oriented approach of complexity, to classify methods and HF activities, to support the understanding of the relations between the HOAC matrix cells and finally to predict the needs to new methods or new interactions between methods. The combination of these four HOAC matrix capacities is its main added value in the overall design process of *simplex artefacts*. In this way the following sections describe a first proposal of *Simplexity criteria*, a systemic approach for *Authority design*, two different views of *eye-tracking studies* and the last method deals with *complexity simulator*. To ensure the link between the description of methods and the HOAC matrix we use the codification able to localize the cells of the matrix impacted by the different methods.

Simplexity Criteria [1 (mi); 2 (mi); 3 (mi); 4 (mi)]. The diagnosis task is a key task within a design process. Indeed, the diagnosis both, emerges from the need analysis and/or results from assessments carried out, and allows to define requirements and recommendations for the conception or else key points needed to be assessed through user-test. The approach presented hereafter focuses on the method implemented to define simplexity criteria which may be implemented at the Micro layer in the HOAC matrix.

These simplexity criteria are based on the principles presented in previous sections and especially the Sect. 3.1 describing the pyramidal model. More precisely, the aim of these simplexity criteria is to assist the identification and the classification of ergonomic problems linked to the simplex artefact (e.g. usability, assistance flaws).

The criteria formalism as evaluative referents [6] has the advantage to be enough conceptual to be adaptive to various HF specialists backgrounds, study scopes, but also to be understandable by specialists from different disciplines. In short, criteria formalism allows exposing widespread knowledge required for a systemic approach and for convergence between several points of views.

The elaboration and assessment of these simplexity criteria follow a process in line with the process used for the ergonomic criteria focused on Human interactions with Graphical User Interfaces [5], or with Virtual Environments [2]. More precisely, the design of the criteria is first based on a qualitative data analysis in line with the *Grounded Theory* principles [10], and then follow a standard user-centered design. The users' feedbacks and performances are collected from two kind of empirical studies. Indeed, the validity level of the proposed criteria depends of their (a) intrinsic validity, that is their usability level and agreement level to support a problems classification; (b) extrinsic validity, that is their reliability level to identify problems in interactive systems.

Table 3. List of the main simplexity criteria

1st version	2nd version
1. Suitability with user requirements	1. Compatibility
2. Detection of information	2. Guidance
3. Recognition of information	3. Adaptability
4. Variability management	4. Actions and information costs
5. Actions implementation	5. Homogeneity/Consistency
6. Workload	6. Threat and error management

So, a first criteria list has been extracted and defined from the codification and classification of a set of recommendations, itself deciphered and grounded on an analysis of scientific and operational literature (see Sect. 3.1). This first list structured in two levels was composed of 6 main criteria (Table 3 – 1st version) and 24 elementary criteria. The definition of each criterion, as well as its rationale and examples of recommendations, are exposed in a technical paper report.

10 Human Factors specialists from the aeronautical filed participated to the assessment of the intrinsic validity of the criteria. This study was based on an assignment task. The results from this first study highlighted that the proposed criteria covered the scope of Human-Simplex System Interactions as well the relevance of these criteria for the HF specialists' diagnosis. The participants have appreciated the possibility to have quickly access to the main properties to consider and the integration within the criteria of current questions like the situated relationships between Augmented Reality and Real Environment. Nevertheless, the criteria should be improved, and so the intrinsic validity of the criteria should be reassessed. More precisely, the organization of this first version was too many distributed to carry out an ergonomic inspection, half of the criteria definitions required to be improved, and some of the criteria needed to be subdivided. An update on the criteria has been performed. This new version structured in three levels is composed of 6 main criteria (Table 3 – 2nd version) and 32 elementary criteria. The assessment of the extrinsic validity of the criteria has been performed based on this second version. 30 participants with Human Factors or engineer backgrounds have performed an inspection of two simplex artefacts. The results are currently ongoing analysis. The update on the criteria will thus be performed.

As the already implemented ergonomic criteria, the main difficulty related to the establishment of the Simplexity criteria lies in the fact to find the right level of abstraction which allows both to have a criteria number cognitively accessible and to ensure that the criteria covered the scope of the intended purpose, to diagnose ergonomic problems. Nevertheless this first tentative is promising because the participants to the two experiments highlighted the good level of guidance to define the scope of interactions with complexity. So, in spite of the difficulties to find the appropriate granularity detailing each criterion, the first step of the appropriate delimitation of the scope of Simplexity criteria, is a positive lesson learned.

Authority Design. The design of authority and responsibilities in new ATM concept is a key point to ensure a good cooperation between agents and by extension absorb a part of the operational complexity. We provide hereafter an abstract of an approach for *authority design* that has been previously adapted to a systemic approach of ATM and detailed in [12]. So, this approach is already adapted to a declension of authority/responsibilities and accountabilities across agents at a Nano, Micro and Meso level. In this way this approach is able to cover many cells of these layers in the HOAC matrix in terms of execution of authority/responsibility. But according to the system impacted by the design of authority some responsibilities have to be designed at a Macro level especially in the cell 3 (ma) in which standard and organization regulation are designed by authorities (e.g. ICAO, FAA, EASA) but on the basis of argumentations elaborated in the previous cells at Nano, Micro and Meso levels. We briefly illustrate the use of this method through a case study based on the management of *Wake turbulences during visual separation in approach* already presented in [12] and updated in order to match with the HOAC matrix. In this perspective, the following case study is tagged by use of the HOAC codification in order to identify which part of the case study match with this matrix.

The case study describes, by use of the proposed Authority design framework, the management of wake vortices during visual separation operations from a pilot perspective. In the US, visual separation in approach is a daily activity to increase traffic capacity. Operations at many airports are based on maintaining visual separation from preceding aircraft using 4 (na) direct visual information on this aircraft. This kind of operation has a strong impact on the 1 (mi) responsibilities of each agent. Indeed, when the 1 (na) flight crew accepts to maintain visual separation, the responsibility of separation 3 (ma) is delegated to flight crew and does not fall under the responsibility of ATCO. Visual Separation allows for reduced spacing between pairs of aircraft which 4 (ma) must be spaced greater than separation minima (i.e. ICAO Wake Vortices Separations) in standard operations. This operation is, for now, not supported by any 2 (mi) on-board function but solely based on 1 (mi) the pilot skills.

In other words, the task of the 1 (na) pilot consists in analyzing the distance from 3 (na) preceding aircraft taking into account 4 (na) contextual information in order to ensure that the own aircraft does not encounter Wake Vortices and, of course, 4 (mi) prevent traffic collision.

Figure 4 illustrates the overall *Assisted Visual Separation* concept by use of the framework of Responsibility and Authority. For each layer and action Responsibility [R] and Authority [A] are identified. The figure shows the big picture of the concept as it could be designed without reflecting the specific execution of the concept during a specific mission.

Around Nano-layer, the detailed actions are analyzed during the execution phase, as one or several agents may be required to conduct actions towards a same goal. Hence, the relevant item at this level is not the responsibility, as responsibility is allocated to a defined agent, but authority. The authority to decide on the adequate distance to maintain from preceding aircraft is given to the pilot.

A human factors issue in relation to this design is the difficulty of this task, since today no tool support is available. The question can be asked if 1(na) human and/or 2 (na) machine shall 3 (mi) and/or 4 (mi) analyze and decide on the distance and speed adjustment to the preceding aircraft. Such a situation can be considered being an authority delegation. Figure 4 shows the 2 (na) ADS-B as a potential mean for such tool support.

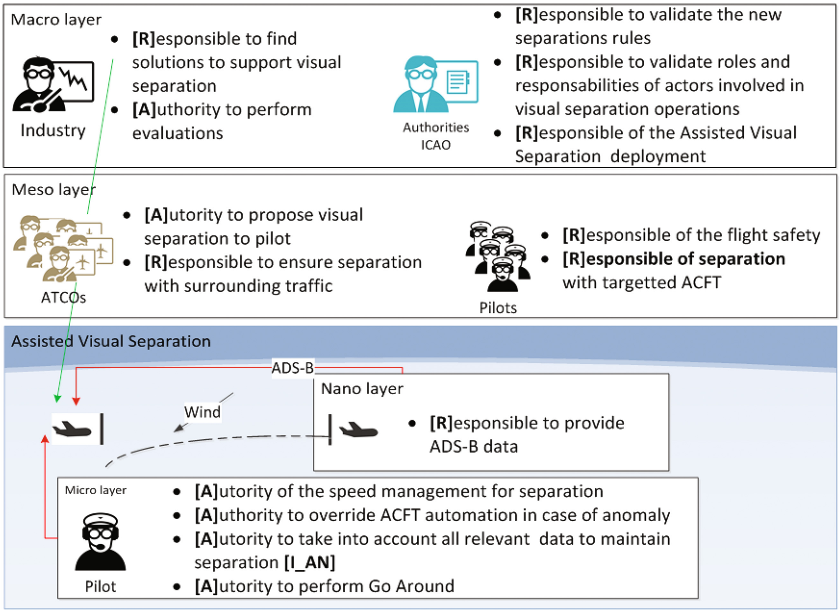


Fig. 4. The framework of responsibility and authority applied to the assisted visual separation concept

Around Micro-layer, in controlled areas the separation responsibility is allocated to the ATCO. Once visual separation accepted, it is the pilot becoming responsible.

Pilot has authority to accept the task of visual separation. Separation is not under her or his responsibility because s/he 3 (mi) is not obliged to accept according to operational descriptions. Only once he has accepted visual separation, he becomes responsible for separation and has to ensure it. Hence, the actions at the Nano-Layer with the pilot analyzing the distance to the preceding aircraft are to be seen in relation to this authority at Micro-Layer.

At the Meso-level, the responsibility between the actors and the social unit could for example be linked to the ways 3 (me) the airlines define directives for 1 (me) pilots depending on their overall mission. Such a directive could be linked to the conduct of the mission, and an example could be seen in requirements of maintaining additional distance to what is judged necessary by the pilot in order to ensure safety and/or the airlines would provide recommendations to their pilots regarding the visual separations.

At the Macro-layer institutions define laws, for example with regard to keeping a certain separation to the preceding aircraft depending on the aircraft category. They impact consequently the accountability of related actors. International institutions, such as the ICAO, have authority to introduce new regulation, but it is up to the definition of the responsibility of more regional authorities by the local states how they are introduced into concrete actions. On the same line the Industrial sector, in the case of *Assisted Visual Separations*, is responsible to provide technical solutions and have also authority to perform the evaluation of the overall efficiency including Human performance of such solutions.

Eye Tracking Evaluations. The increased complexity of ATM concept such as visual separations described, in the previous section, require adapted tools able to provide an adapted level of metrology. In this way eye-tracking evaluations and more broadly neuro-ergonomics provide more and more reliable direct measurements and in some cases indicators able to study more and more precisely fast human interactions in operations. Such measurements provide interesting quantitative and objective data able to assess in a new way complex concepts as for example the situation awareness, visual scan patterns [21]. Itself these eye-tracking studies are serious candidates to cover the cells 1 (na) and 2 (na) of the HOAC matrix. Nevertheless the impact of the eye-tracking use may be broader. Indeed eye-tracking could be used as a tool supporting the progress of student pilots through training programs at a meso level.

Complexity Simulator. In the frame of a Human Oriented approach of interactions with complexity, the HF factor specialist or at least its argumentation may need methods or study at the macro level. This is especially the case when the HF specialist has to deal with the design of new ATM concept with apparently few evolution compare to existing operations. In this type of projects the acceptance of the different actors are difficult to reach at a meso level and usually this is because the perception of benefits of the concept should be viewed on the long term and at a traffic level. Therefore the perception of benefits is almost impossible for designer because these benefits are the results of a long and complex overall system (e.g. positive impact on CO_2; traffic predictability. In such cases the HF specialist needs to access to study using specific tools able to simulate the result of numerous small operational micro modifications that can have an important macro benefit due to the complexity of local interaction on long term period. These kinds of simulations (usually named fast time simulation) can be performed, so far, only by specific ATM agencies such as Eurocontrol, Mitre or NATS. But their results are really important to ensure and deals with of transition factors such as the acceptability of the new ATM concept. In this way the HOAC matrix can locate this kind of specific activities 2 (ma)/4 (ma) and guide the HF specialist to understand the importance of such complexity simulation for his/her own HF activity.

5 Discussion and Limits

We describe in this paper the HOAC matrix able to support a big picture of HF specialist dealing with aeronautics complexity. The aim of this matrix is to objectively classify the HF methods, tools and specialist actions. We illustrate with few examples some cells of this matrix. Obviously this matrix can be viewed as a predictive structure able to develop activities and tools supporting the entire Human Oriented Approach of interactions with complexity. This contribution is a first attempt to structure and integrate the HF specialist activity in this area, that is an actual great challenge. Nevertheless our approach is mainly based on a theoretical study. In this sense our proposal includes different flaws and should be completed and structured with a set of studies assessing the overall utility of this matrix and its efficiency to structure the HF specialists. Furthermore our study focuses solely on the interactions between the tasks and the vertex of the pyramidal model. Obviously this theoretical study should be extended to the entire pyramidal model and the HOAC matrix should be completed with other methods and activities in future works.

References

1. Amalberti, R.: The paradoxes of almost totally safe transportation systems. Saf. Sci. **37**, 109–1206 (2001)
2. Bach, C., Scapin, D.L.: Comparing inspections and user testing for the evaluation of virtual environments. Int. J. Hum.-Comput. Interact. **26**(8), 786–824 (2010)
3. Bach, C., Biede, S.: Assessing needs from a human factors perspective when designing complex systems. In: Schwarz, M., Harfmann, J. (eds.) Proceedings of the 32nd Conference of the European Association in Aeronautics Psychology, Groningen, NL, pp. 224–242 (2017)
4. Barbé, J., Boy, G.: On board system design to optimise energy management. Paper Presented at the European Annual Conference on Human Decision-Making and Manual Control (EAM 2006), Valenciennes, France, September 2006
5. Bastien, J.M.C., Scapin, D.L.: Ergonomic criteria for the evaluation of human-computer interfaces: Institut National de Recherche en Informatique et en Automatique (1993)
6. Bonnardel, N.: Le processus de l'évaluation dans les activités de conception. Doctorat, Université de Provence/INRIA, Aix-en-Provence (1992)
7. Cilliers, P.: Complexity and Postmodernism: Understanding Complex Systems Complexity. Routledge, UK (1998)
8. Ferber, J., Gutknecht, O., Michel, F.: From agents to organizations: an organizational view of multi-agent systems. In: Giorgini, P., Müller, Jörg P., Odell, J. (eds.) AOSE 2003. LNCS, vol. 2935, pp. 214–230. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24620-6_15
9. Gell-Mann, M.: What is complexity? Complexity **1**, 16–19 (1995)
10. Glaser, B.G., Strauss, A.A.: La découverte de la théorie ancrée. Armand Colin, Paris (2010)
11. Harris, D.: Human Performance on the Flight Deck. Ashgate Publishing Ltd., Burlington (2011)
12. Henry, J., Bach, C., Biede, S.: A framework for designing authority and responsibility into complex sociotechnical systems. In: Proceedings of the 5th International Conference on Application and Theory of Automation in Command and Control Systems, Toulouse, France, September 2015, pp. 1–10. ACM (2015)
13. ISO 9241: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability (1998). International Organisation for Standardization
14. Liu, P., Li, Z.: Task complexity: a review and conceptualization framework. Int. J. Ind. Ergon. **42**, 553–568 (2012)
15. Morin, E.: From the concept of system to the paradigm of complexity (1990). Translator Sean Kelly. https://manoftheword.files.wordpress.com/2013/07/morin-paradigm-of-complexity.pdf
16. Perret, V., Bach, C., Calvet, G., Chevalier, A.: A user-centred approach of complexity: toward development of a method to support simplex systems design. In: Proceedings of HCI-Aero 2016, Paris, France, September 2016

17. e Cunha, M.P., Rego, A.: Complexity, simplicity, simplexity. Eur. Manage. **28**, 85–94 (2010)
18. Ragosta, M., Martinie, C., Palanque, P., Navarre, D., Sujan, M.A.: Concept maps for integrating modeling techniques for the analysis and re-design of partly-autonomous interactive systems. In: Proceedings of the 5th International Conference on Application and Theory of Automation in Command and Control Systems, Toulouse, France, September 2015
19. Varela, F., Thompson, E., Rosch, E.: The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridge (1991)
20. Winckler, M., Bach, C., Bernhaupt, R.: Identifying user experience dimensions for mobile incident reporting in urban contexts. IEEE Trans. Prof. Commun. **56**(2), 97–119 (2013)
21. Yu, C., Wang, E.M., Li, W.-C., Braithwaite, G.: Pilots' visual scan patterns and situation awareness in flight operations. Aviat. Space Environ. Med. **85**(7), 708–714 (2014)

Human-Swarm Interaction as Shared Control: Achieving Flexible Fault-Tolerant Systems

Jacob W. Crandall(✉), Nathan Anderson, Chace Ashcraft, John Grosh,
Jonah Henderson, Joshua McClellan, Aadesh Neupane,
and Michael A. Goodrich

Brigham Young University, Provo, UT, USA
`crandall@cs.byu.edu`

Abstract. Robot swarms modeled after hub-based colonies, such as ants and bees, potentially offer fault-tolerant capabilities at very favorable cost margins. However, relatively little is known about how to harness the potential of these swarms through command-and-control systems. In this paper, we study how to merge operator input with the underlying swarm behavior to maintain the fault-tolerant attributes of robot swarms while providing the operator with enough control to ensure that mission objectives are accomplished. We advocate that an effective mechanism for achieving this is *shared control*, wherein decision-making is shared between the human operator and the underlying swarm dynamics. We lay out characteristics of human-swarm systems that provide an effective balance between fault-tolerance and control, and we discuss preliminary designs of human-swarm systems for hub-based colonies based on these principles.

Keywords: Human-Swarm Interaction · Shared control · Hub-based colonies

1 Introduction

Robotic swarms potentially offer fault-tolerant and coordinated defense, surveillance, and delivery capabilities at very favorable cost margins. As such, they can be an important complement to existing precision-based systems [19]. These robot collectives are often modeled after biological swarms, such as hub-based colonies (e.g. ants and bees) and spatial swarms (e.g. birds, fish, and locusts), where individual members of the population act using simple sensor systems and behavioral strategies. Local interactions between these simple systems produce complex and intelligent behaviors that are robust to many kinds of attacks and environmental factors. Robot swarms modeled after these biological swarms have already been successfully developed for a number of applications [10,16,18,26], and thus have great potential.

Human-swarm interactions (HSI) [12], wherein one or more human operators manage the robot swarm through a command-and-control interface, are necessary to ensure that the robot swarm's behavior aligns with mission objectives.

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 266–284, 2017.
DOI: 10.1007/978-3-319-58472-0_21

To date, the majority of work in HSI has focused on robot collectives modeled after spatial swarms (e.g., [1,7,11,21,25]). Less is understood concerning how to harness the potential of hub-based colonies through HSI. Thus, this paper addresses the topic of human interaction with robot swarms modeled after these hub-based colonies.

For human interactions with robot swarms of this kind, it is tempting to view the human operator as a centralized, authoritative controller of the swarm. However, this philosophy negates the strength of swarm technologies (decentralized, fault-tolerant systems), and instead turns the operator into a potential single point of failure. Centralized operator control is particularly problematic when the operator has limited or incorrect information about the environment in which the swarm is operating, or in complex scenarios in which the operator cannot possibly attend to all aspects of the mission at once. Thus, an alternative control paradigm is required to preserve the fault tolerance of the swarm while giving the operator sufficient influence to align the swarm's behavior with mission objectives.

We advocate that an effective method for balancing human control and fault tolerance is *shared control*, wherein the human operator and the underlying situated dynamics of the swarm share the burden of decision-making. The impact of this design choice is that there is a potential trade-off between the control given to the human operator and the resulting fault tolerance of the system. The nature of this trade-off defines in part the success of the human-swarm system.

In this paper, we study various mechanisms for sharing control between a human operator and a robot swarm modeled after honey bees. In Sect. 2, we describe the underlying dynamics of the robot swarm. We then discuss, in Sect. 3, how human interactions with this system result in the sharing of control between the human and the robot swarm, which impacts the trade-off between operator control and the fault tolerance of the system. Finally, in Sect. 4, we describe a preliminary design of our human-swarm system in which the human operator interacts with the robot swarm described in Sect. 2.

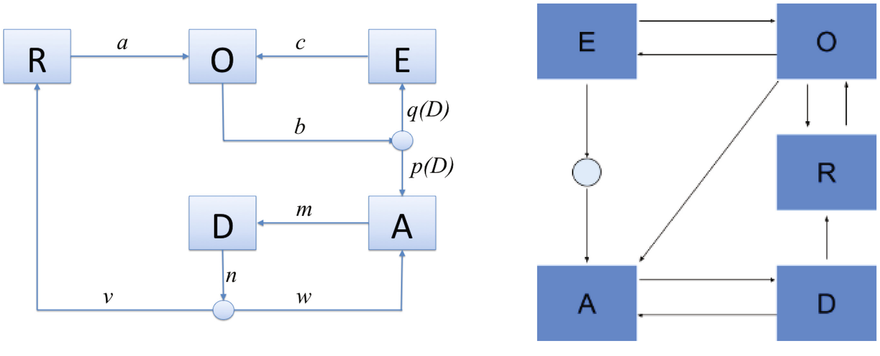
2 A Robot Swarm

Hub-based colonies, such as ants and bees, perform a variety of complex functions. One such function is the selection of a new nest site. This problem corresponds to selecting the best of n choices, a task relevant to surveillance, search and rescue tasks, as well as practical considerations such as setting up a swarm's home base.

In this section, we describe our simulation of this hub-based colony. In subsequent sections, we discuss the design of human-swarm interfaces for this system.

2.1 A Model of Honey Bees

Our simulated robot swarm is based on a paper by Nevai and Passino [15], which defines a state machine and a set of differential equations that describe how scout bees in a hive of honey bees (*apis mellifera*) select a new nest site. Their finite-state machine is shown in Fig. 1a. In this model, bees transition through five different states: exploring (E), observing (O), resting (R), assessing (A), and dancing (D). Our implementation follows this model, though instead of having our robots transition at given rates, we implemented an event structure that is meant to resemble actual bee behavior. This necessitated a switch to the state-transition function shown in Fig. 1b.



(a) (b)

Fig. 1. (a) State-transition diagram for honey bees selecting a new nest, as modeled by Nevai and Passino [15]. (b) Our modified state-transition function designed for a spatial, event-driven simulation of the swarm.

Initially, all of our robots are located in a central hub and are placed in the exploring state. Explorers move randomly through the environment until they encounter a potential site, at which point they transition to an assessing state, and fly back to the hub. Upon arriving, the robots enter the dancing state, in which they move around the hub advertising their site to the other robots. The majority of communication among robots takes place at the hub. Each robot dances for a time proportional to the quality of the site, and then, in our model, returns to the assessing state and leaves to reevaluate the site. The number of times this happens is also proportional to quality of the nest site.

Once a robot has finished dancing, it enters the resting state, in which it simply waits at the hub for a period of time before entering into the observing state. Observers wander the hub looking for dancers, and upon encountering one, enter into the assessing state to begin the dance/assess process. If no dancing robots are noticed and sufficient time passes, the robot will instead enter the exploring state and begin to look for sites, or, with a small probability, enter the resting state.

When the robots make a collective decision to accept a site, they are said to have *quorumed*, which describes their movement to the new site. In addition to the base model, we decided to implement quoruming by adding two new states and a sub-state. Robots decide to quorum based on how many robots they encounter that are assessing a particular site. If the number exceeds a threshold, then the robots begin a process called piping. We model this by creating a sub-state called site-assess where the assessor robots move around the potential site for a time before returning to dance. During this state they monitor the number of robots at the site, and if it exceeds a given threshold, they enter the piping state. Pipers alert and stimulate other robots to prepare for liftoff to settle another site [20]. Robots in this state fly back to the hub and advertise their site similar to dancers, but do not re-assess the site.

The final transition to the commit state occurs when a set time has passed and the robot senses that all robots that are nearby are also piping. Once entering this state, they move to the potential site and set their hub location to the potential site location.

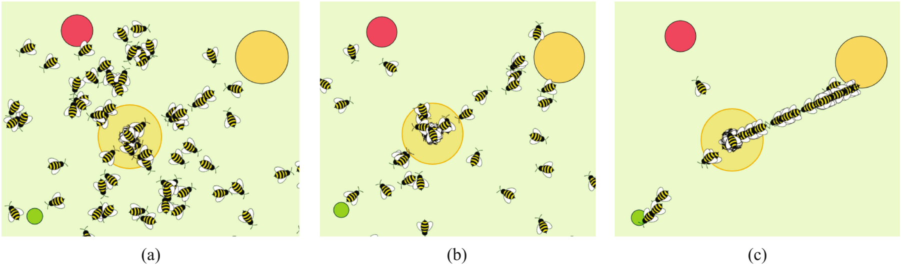


Fig. 2. Successive screen shots of a bird's-eye view of the simulated swarm as it seeks to locate a new nest site. Robots are depicted as bees, and potential nest sites are drawn as red, yellow, and green circles. (a) The robots are spread out searching for potential sites. (b) Some of the robots begin to repeated assess two of the sites. (c) The majority of the robots begin converge toward the most desirable site (Color figure online).

2.2 Simulation Results

Figure 2 shows a series of screen shots depicting the behavior of the simulated robot swarm. Initially, the robots appear to be randomly scattered throughout the world as they search for potential sites (Fig. 2a). Subsequently, some of the robots discover sites, assess these sites, and then begin to recruit others to also assess these sites (Fig. 2b), until most of the robots have selected the most desirable site (Fig. 2c).

Through repeated testing, we identified parameter settings for which the swarm tended to find the best target site without any human oversight or interaction with the swarm. We found that one important parameter for increasing the percentage of robots who committed to the best site was to decrease the robot's variation along their direction of exploration (moving outward from the hub). The robots' movement occurs at a constant velocity (barring obstacles or rough terrain) in a direction that is continually updating randomly according to a Gaussian distribution. The robots tended to explore farther from the hub and encounter better quality sites when the variance of this distribution was small.

We evaluated the performance of a 100-robot swarm in ten different environments, each with different attributes. Each environment included several target sites of varying quality, as well as obstacles, traps, and rough terrain. Table 1 shows the percentage of the swarm's robots that found each site, averaged over 30 trials for each environment. In most cases, a majority of the robots either committed to the best site or were lost (meaning they were caught in traps). However, in two of the environments (Environments 6 and 10), the robots tended to often commit to a less desirable site. These environments proved more difficult because there was an adequate site near the hub, whereas the highest quality sites were located farther away, near traps and obstacles (e.g., Fig. 3).

It is likely that various parameters of the robot swarm could be tuned to make the swarm more robust. Furthermore, a larger swarm would likely better adapt to more environmental circumstances [6]. Despite these drawbacks, the simple control technology of the swarm produces rather effective results.

2.3 Why Human-Swarm Interaction?

The simulation results shown in Table 1 confirm the ability of hub-based colonies to solve complex problems. Through local, microscopic interactions with the environment and between team members [8], the swarm produces complex, macroscopic behaviors that are extremely robust to failures. Because of this success, a natural question arises: Why is it necessary for a human operator to interact with a robot swarm patterned after hub-based colonies?

A human operator fulfills three roles in human-swarm systems patterned after hub-based colonies:

Table 1. Evaluations of the robot swarm's ability to select the best site in ten different environments. Results are averaged over 30 different trials each.

Environment	% Committed to best site	% Committed to other site	% Lost
1	58.0	20.5	21.5
2	73.7	0.0	26.3
3	60.3	0.0	39.7
4	86.9	0.0	13.1
5	81.5	1.9	16.7
6	8.8	70.2	21.1
7	79.2	0.0	20.8
8	71.0	0.0	29.0
9	99.5	0.0	0.5
10	3.3	68.0	28.7

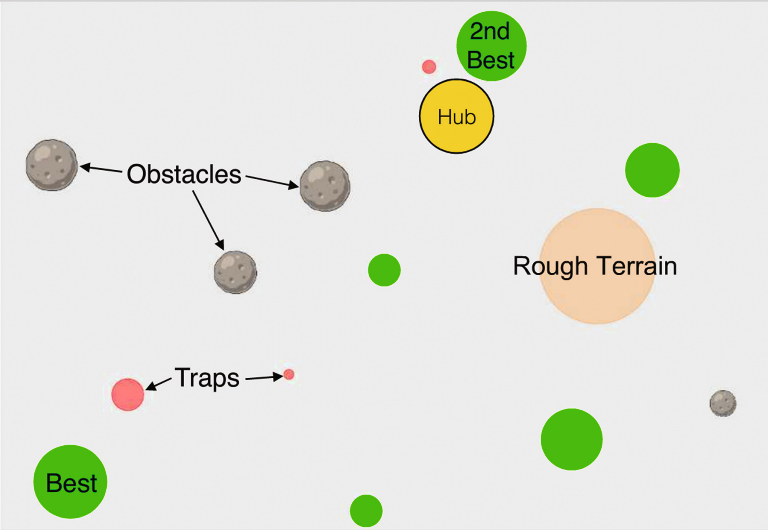


Fig. 3. Environment 10 in our evaluation studies (Table 1). Green circles are potential sites, with the site of the circle indicating a more desirable site. The swarm tended to select the second-best site in this environment due to the traps and obstacles in between the hub and the best site. (Color figure online)

1. *The human operator aligns the behavior of the swarm with strategic mission objectives.* The swarm collectively encodes information about the environment and reacts to this information. However, effectively adapting these reactions to fulfil mission objectives is often a complex and dynamic process. For instance, the mission objectives themselves may need to be adjusted or redefined to better align with a larger strategy. In this case, the operator should serve to continuously realign swarm behavior with overall goals by correcting for higher-level information the swarm is incapable of modeling or encoding.
2. *The human operator supplies information to the swarm that is not immediately available through the robot swarm's sensor and communication systems.* Other information sources may make the operator aware of information the swarm does not have. In such circumstances, the performance of the swarm can be enhanced if the operator is able to effectively communicate this information to the swarm. Table 1 indicates a specific instance where swarms would benefit from human intervention. Given appropriate abilities to influence the swarm, the substantial number of robots that get caught in traps in our simulations during the during the nest-selection process could potentially be reduced.
3. *The human operator augments the swarm when it is not properly evolved for the current environment.* While swarm dynamics are incredibly robust to failures under normal circumstances, swarms may still fail. For example, when

control parameters, optimized for particular environments, are not properly tuned for the current environment, the swarm's underlying dynamics could potentially lead to undesirable outcomes (see, for example, the results from environments 1, 6, and 10 in Table 1). Alternatively, if the swarm size becomes depleted, the swarm may require assistance, as it may not be able to approximate the true state of the environment [6]. Under such circumstances, the human operator can potentially adjust or augment the swarm.

The remainder of the paper focuses on how human-swarm systems can be designed so that a human operator can effectively play these roles without disrupting the swarm dynamics.

3 Decision-Making in Human-Swarm Systems as Shared Control

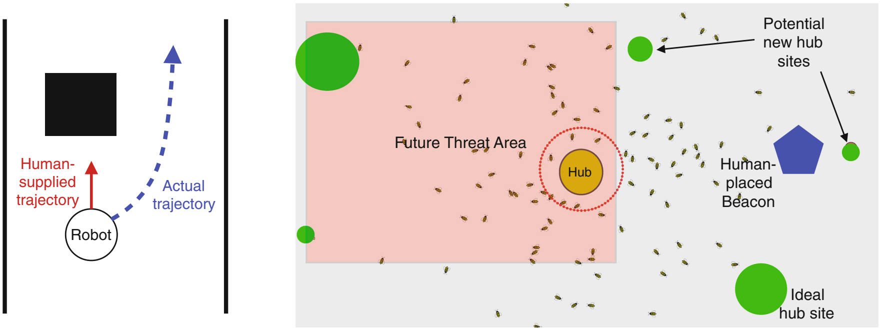
Although robotic hub-based colonies have considerable innate potential, human guidance helps to ensure their compliance with mission goals. Nonetheless, a human element has the potential to override a swarm's desirable features if the operator does not possess correct knowledge of the operational environment. In an attempt to maximize the advantages of both human and swarm decision-making, we argue for *shared control*. In this paradigm, the swarm should accept human input as additional information to be acted upon according to the swarm dynamics. In this way, the robust and fault-tolerant nature of the swarm can be maintained while considering human input.

In this section, we discuss this shared-control paradigm for human-swarm systems. We then consider how this control paradigm impacts the trade-off that emerges between operator control and the swarm's fault tolerance. Finally, we discuss how the information and control elements of the human-swarm interface can be designed to achieve an effective balance between operator control and fault tolerance.

3.1 Robustness Through Shared Control

The concept of shared control has been used in many kinds of human-robot systems, particularly in teleoperation systems (e.g., [3,9,22]). In these systems, the human typically expresses high-level intent through the control interface. The robot is charged with finding a low-level behavior that both satisfies acceptable performance criteria and conforms to the high-level intent expressed by the human operator. For example, in teleoperating a robot through a corridor, an operator may tell the robot to move in a particular direction, and leave the actual path planning (i.e., navigation around obstacles) to the robot (Fig. 4a). In this way, the operator controls the high-level behavior, while the robot controls the low-level behavior necessary to achieve human intent and performance constraints (e.g., avoiding obstacles).

Shared control in human-swarm systems works similarly. As an example, consider a scenario in which the operator has been notified of a future spatial



(a) (b)

Fig. 4. (a) An example of shared control teleoperation, in which the human provides a general direction it intends the robot to move. The robot then generates low-level behavior that moves in this general direction but avoids obstacles. (b) An example requiring shared control in a human-swarm system, wherein the operator seeks to move the hub out of a threat area. The human specifies the intended direction to move the hub with a beacon (blue pentagon), and the swarm is then responsible for finding an new nest site (green circles – larger circles indicate better sites). (Color figure online)

threat at the location of the swarm's hub. In this case, the hub must be moved away from the potential threat area, even though the robots cannot yet sense the threat (Fig. 4b). In this case, the operator can initiate a "move nest" behavior, but must then influence the search in a particular direction (potentially via an attracting beacon) to influence the robots to search outside of the threat area. This general expression of intent is then satisfied as the swarms finds an ideal site using its underlying dynamics coupled with the influence of the beacon (which acts on the swarm's dynamics).

The use of a beacon to attract robots to particular locations in the example illustrated in Fig. 4b highlights an important trade-off. If the beacon exercises too much influence over the swarm, the swarm will fail to find the ideal new hub location in the bottom right corner of the figure. Rather, the robots will focus their search exclusively near the beacon, thus making it likely that the swarm will converge to the undesirable site near the beacon. On the other hand, if the beacon has too little influence over swarm dynamics, the robots could potentially converge to the highly desired target site in the upper left corner, a site that is still in danger of the anticipated future threat. This illustrates the important trade-off between operator control and the fault tolerance that is caused by the use of shared control in human-swarm systems.

3.2 The Trade-Off Between Fault Tolerance and Operator Control

Another potential benefit of working with hub-based colonies is resistance to single points of failure. However, when introducing a human controller into the swarm, that human becomes a new potential single point of failure. This further motivates the concept of shared control, but also poses the question of how to balance the control between the human and the swarm. The human should have enough leverage to affect the colony, but not so much as to negate its beneficial, fault-tolerant dynamics. In short, how much control is enough, and how much is too much?

There are many ways that the operator could affect the swarm. It is reasonable to assume that one could design a control scheme for the swarm such that human control is sufficient for the needs of the mission, but limited enough to preserve the beneficial behaviors of the swarm. We desire to create a measure for potential control schemes that specifies how different levels of control impact the swarm's fault tolerance. Because *controllability* is already well defined and our measure suggests a spectrum of control, we instead refer to this measure as the level of *influence* a control scheme gives to the operator.

Thus, we believe that for any robot swarm modeled after hub-based colonies, the higher the influence the human has over the colony, the lower the fault tolerance of the swarm will be. The actual relationship that exists between the two concepts is likely dependent on many aspects of the swarm, including the type and form of control given to the human operator. We would like to design a framework that would allow us to rigorously study these terms and their relationships, but for now we can only project some possibilities. An ideal case for this relationship would be something like the blue (solid) line in Fig. 5, where there is a level of operator influence that does not substantially sacrifice the swarm's fault tolerance. However, if swarm dynamics and operator control are not carefully designed, other trade-offs between operator influence and the swarm's fault tolerance are likely. For example, the green (dashed) line in Fig. 5 suggests an equal loss of fault tolerance to gain in operator influence, while the red (dotted) line suggest a substantial loss in fault tolerance even for low levels of operator influence.

We hypothesize that human-swarm systems are likely to have desirable trade-offs between influence and fault tolerance when they are guaranteed to maintain certain properties. For example, Millonas [14] stated five principles of collective intelligence that a swarm should maintain. Specifically, the swarm should be able to (1) perform simple space and time computations, (2) respond to quality factors in the environment, (3) avoid allocating all of its resources along excessively narrow channels, (4) avoid reacting to every fluctuation in the environment, while (5) having the ability to change behavior when doing so is worth the computation price. We anticipate that human-swarm systems that maintain these swarm principles will lead to desirable trade-offs.

Understanding the dynamics between operator influence and fault tolerance may help in the design and evaluation of interaction frameworks for robotic swarms by providing measurements of swarm capabilities in the presence of human control. Such understanding could potentially allow for the design of control frameworks that are more resistant to human error, cyber-attacks, and, as already noted, single points of failure. In the next subsection, we begin to discuss various design decisions for human-swarm interfaces that likely impact these dynamics.

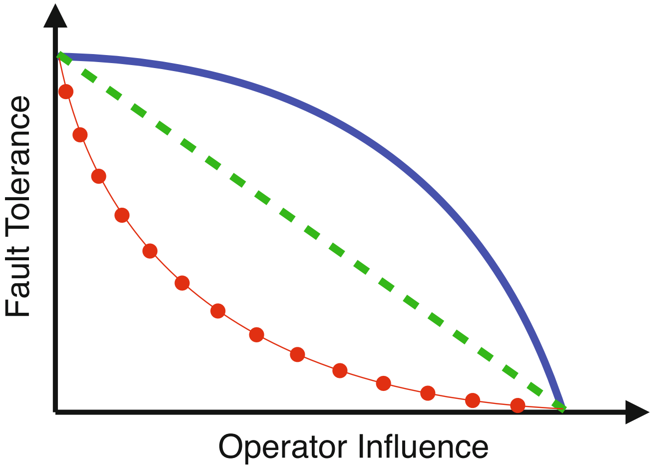


Fig. 5. Hypothetical trade-offs between fault tolerance and operator influence. The blue solid line represents a desirable trade-off in which fault tolerance is maintained for moderate amounts of operator influence, whereas the green (dashed) line and red (dotted) lines represent less desirable trade-offs. (Color figure online)

3.3 Characterizations of Human-Swarm Interactions

The human-swarm interface defines the interactions between the operator and the swarm. We now characterize broad notions of interactions that interfaces could potentially support. We divide these characterizations of the human-swarm interface into three categories: the levels of engagement of the operator, the categories of control mechanisms provided to the operator, and the elements of observation given to the operator. We discuss each in turn.

Levels of Engagement. A human operator can potentially engage with the swarm at two different levels: *swarm-level engagement* and *mission-level engagement*. In swarm-level engagement, the operator observes and adjusts the state of the swarm. The operator is interested in how the swarm evolves, what it is doing, and how it is doing it. On the other hand, mission-level engagements focus on strategic mission objectives. In mission engagement, the operator is concerned with articulating the strategic objectives of the mission to the swarm and determining whether or not these objectives have been or are being accomplished.

There is not a general answer for the question of which level of engagement is ideal. In many scenarios, both levels of engagement should be possible. Several factors contribute to this design decision. For example, how much of the swarm's behavior and state can reasonably be communicated to the operator? If communication bandwidth does not permit rich understanding of the current state of the swarm, mission-level engagement might be more effective. Likewise, how much operator influence should be supported? Lower influence will typically

Table 2. Four categories of control mechanisms, each of which represents a different way of providing input to the robot swarm.

Category of control	Primary level of engagement	Brief description
Parametric control	Swarm	The operator changes parameters governing individual robot behaviors
Control by association	Swarm	The operator directly controls members (or virtual members) of the swarm, who then influence the rest of the swarm via interactions
Environmental control	Swarm	The operator "modifies" the swarm's environment, potentially via virtual environmental features that the robots can sense
Strategic control	Mission	The author gives mission-level feedback or high-level instructions to the swarm

relate to mission-level engagement, whereas high influence will typically support swarm-level interactions.

The levels of engagement supplied by an interface appertain to both the control mechanisms and observation elements of the interface.

Categories of Control Mechanisms. Table 2 summarizes four different categories of control mechanisms that can be used in human interaction with robot swarms modeled after hub-based colonies. We refer to the first category of control mechanisms as *parametric controls*. This category refers to controls that modify parameters that govern the individual behaviors of robots, including the rate at which robots perform particular functions, how quickly the transition between states, how broadly they explore, etc. For example, for the robot swarm described in Sect. 2, the operator could potentially change the amount of time each robot spends exploring, dancing, or resting. Such changes can produce dramatic changes in the overall swarm behaviour.

Parametric controls are desirable because operators can make a single set of parameter changes that require neither line of sight nor significant subsequent supervision of the swarm. For example, suppose that a human operator oversees a swarm in an environment where visibility is limited. If robots repeatedly campaign for poor quality nest sites, the operator can respond by decreasing the time permitted for dancing. Even though the operator cannot see where quality nest sites are, he can compel the robots to continue searching until they have found an acceptable site. One disadvantage of this method is that managing a swarm is non-intuitive. An operator must understand how the different rates of change affect swarm state, and think clearly enough to produce a desired outcome. Hence, these methods may not always be appropriate for novice users.

The second category of control mechanisms listed in Table 2 is *control by association*, wherein the operator directly controls members (or virtual members) of the swarm, who then influence the rest of the swarm via interactions.

Many studies suggest that a human can only manage a limited number of robots efficiently [4,17,24]. Since swarms contain hundreds of autonomous robots, it is not possible to control all robots at once, nor would this likely lead to fault-tolerant swarms. However, by controlling a limited number of robots or virtual robots (e.g., [23]), the operator can impact the other robots in the swarm through association. The number of robots (or virtual robots) controlled by the operator impacts operator influence when using this form of control.

Environmental controls are a third category of control mechanisms that could be made available to human operators. Under these mechanisms, the operator does not directly influence the behavior of the swarm, but rather modifies the environment in which the swarm operates to produce desired behavior. As an example, the operator may be able to discern the strategic value of certain locations more quickly than the swarm, and can encourage or discourage exploration around those locations by placing virtual objects (which can be sensed by the robots) in the environment. The advantage of this approach is that it is more immediately intuitive; its drawbacks are that it assumes the operator has higher-quality information about the environment than the swarm. It also may require significant operator attention to be fully effective.

Strategic controls differ from the other three categories of control in that they directly pertain to controlling the mission rather than controlling the robots in the swarm. These control mechanisms include playbook style interactions [13] in which the operator selects high-level swarm behaviors (e.g., initiating a find-new-nest behavior) or reinforcing particular mission outcomes. Such interactions are desirable because they allow the operator to ignore swarm dynamics (which they may have difficulty observing anyway) and instead focus on the bigger picture. On the other hand, such controls do not allow the operator to influence low-level behaviors.

Elements of Observation. The control mechanisms available to the operator are likely contingent on what the operator can observe and perceive from the user interface. Chen et al. [2] identified three levels of transparency that could be

Table 3. Three levels of transparency that could potentially be achieved by human-swarm interfaces. Adapted from Chen et al. [2].

Level of transparency	Brief description
1	Conveys what is going on with the swarm and mission, and what the robots are trying to achieve
2	Conveys why are the robots doing what they are doing, including the robots beliefs and reasoning processes
3	Conveys what the operator should expect to happen in the future to the swarm. Communicates whether mission objectives be achieved?

communicated by the human-swarm interface (Table 3) to support situation awareness. The first level relates to information that communicates what is happening (both at the swarm and mission levels), and what individual robots are trying to achieve (swarm-level engagement). The second level relates to information about how the robots make decisions. This swarm-level engagement is often necessary to successfully implement parametric controls. Finally, the third level of transparency relates to information that indicates future swarm and mission states.

The question arises as to the degree to which each level of transparency should be portrayed to swarm operators. Given limitations in communication bandwidth, it is unlikely that all levels of transparency could be communicated for individual robots. However, various aspects of transparency would likely be useful at the swarm or mission level. Regardless, transparency requirements should be carefully considered when selecting which control mechanisms are implemented in the human-swarm interface.

4 A Human-Swarm Interface (Preliminary Design)

In the previous section, we advocated that human-swarm systems should use appropriate shared-control methodologies to adequately balance operator influence and fault tolerance. We also enumerated a variety of different methodologies for controlling a swarm, each of which must be supported by appropriate transparency requirements. In this section, we describe a preliminary design for a human-swarm interface to support operator interactions with hub-based colonies. In so doing, we describe both the information and control elements to be supported in this interface.

4.1 Information Display

As stated in Sect. 3.3, the human-swarm interface should provide appropriate transparency [2] both in terms of the state of the robot swarm and the state of the mission. First, we propose supporting level-1 transparency through radial displays of both the mission and swarm state (Fig. 6). Given the limited capabilities of individual robots to communicate what they learn and to sense the environment, only limited and somewhat uncertain information will be available to the operator. Furthermore, given the vast number of robots in the swarm, knowledge about individual robots would overwhelm the operator. Our information display, which is based on radial visualizations [5], communicates the swarm's state and the overall state of the mission and environment, rather than displaying the state of individual robots (Fig. 6).

The radial display allows the user to see which direction each robot left the hub. This gives the user the (level-3) transparency of seeing the predicted directions of where each robot is headed and where they may end up. The hub also allows the user to predict the behavior of the robots by showing the projected amount of robots leaving the hub in each direction after the user has excited or inhibited the swarm in each direction (described in more detail in Sect. 4.2).

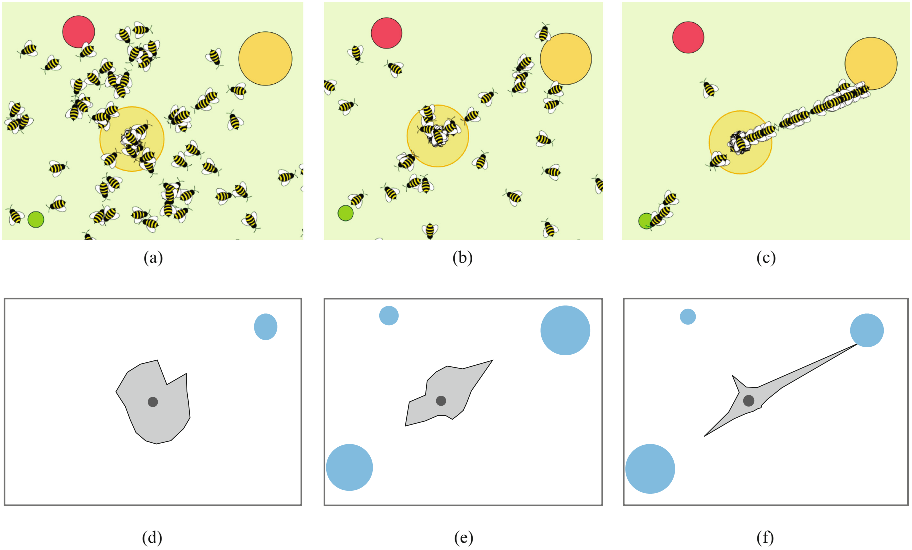


Fig. 6. (a-c) Three screenshots of a bird's-eye view of the robot swarm foraging for a location for a new hub location. Robots are depicted as bees, and potential nest sites are drawn as red, yellow, and green circles. (d-f) Mock-ups of the corresponding visualizations of the swarm state (grey) and potential nest sites (blue), where bigger circles are thought to be better sites, for the three scenarios depicted in (a-c). (Color figure online)

We also intend for the interface to use data from the robots to predict other behaviors. The hub can use the velocity of the explorers as well as their direction to predict when the explorer should return to the hub. If the explorer does not return to the hub in time, the hub takes note. If this happens repeatedly, the hub should notify the user that many robots leaving in a certain direction are disappearing, and there is likely something dangerous in that direction.

The information display also communicates relevant information about the mission state. As the robots collect information about various sites, the estimated quality of each site is displayed. Together with the radial display showing the swarm state and an understanding of swarm dynamics, the operator can infer the likely future state of the system.

4.2 Controls

In Sect. 3.3 (Table 2), we identified and briefly discussed four different categories of control mechanisms: parametric control, control by association, environmental control, and strategic control. Our preliminary interface design is intended to implement one or more example mechanisms from three of the four control types. Table 4 lists these example mechanisms, which we discuss in turn.

Table 4. Control mechanisms in our preliminary human-swarm interface design.

Category of control	Specific mechanisms
Parametric control	Rate control – Control how robots transition between states
	Exploration control – Control directions robots leave the colony
Control by association	None. Due to communication constraints which provide limited knowledge of the movements of individual robots in the swarm, we chose to not implement this control mechanism
Environmental control	Bug bait – Environmental cue that attracts robots
	Bug bomb – Environmental cue that repels robots
Strategic control	Playbook – Whole swarm behaviors focused on task achievement
	Quality attribution – Ability to modify how sites are valued

Parametric Controls. We are considering two different forms of parametric control: rate control and exploration control. Rate control refers to real-time modifications to parameters that control how robots transition through their states. In some cases, the operator may observe that the swarm appears to be converging too quickly or too slowly to a solution. In this case, the operator can inhibit or excite state transitions by interacting with an information display showing the distribution of robots thought to be in each state.

The exploration control is done by interacting with the radial display showing the swarm's state. Recall that this display is formed by recording the direction that each robot leaves the swarm. We allow the user to excite or inhibit exploration in any direction by clicking and dragging on this radial display. Then, robots leaving the swarm to explore select their direction according to the distribution selected by the user. So as to not provide too much influence to the operator (in the spirit of shared control), this suggestion is maintained for only a limited amount of time.

Environmental Controls. We implemented two different environmental controls in our system, which we call *bug bait* and *bug bomb*. The bug bait acts as an attractor; it attracts robots to it. On the other hand, the bug bomb is a repellent, as it drives robots away from it (Fig. 7). To use these tools, an operator specifies a location and whether or not exploring robots should be attracted to or repelled from the point. Several features have been built into the tool to limit operator influence. Firstly, the attraction and repulsion mechanisms are probabilistic. Robots have a chance of ignoring an attracting or repelling influence and continuing to explore normally. Second, these attractors and repellents also eventually expire so that they no longer influence the robot's movements. To force the robots to stay in or move away from an area, attractors and repellents must be continually replaced by the operator. Lastly, the operator must wait for a “cooldown” period after using the tool before using it again.

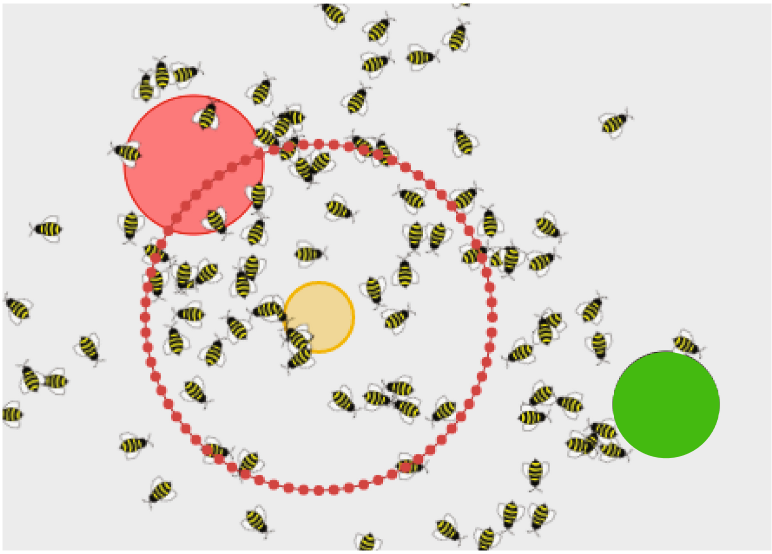


Fig. 7. The operator can drop a *bug bomb* (red circle) in the virtual environment to drive robots away from a particular location. When this repellent is placed, robots probabilistically scatter away from the specified location. (Color figure online)

These mechanisms were created to balance the benefits and potential hazards of human influence on the swarm. A human operator using these tools can shepherd robots towards an area they would otherwise be unlikely to explore, or push them away from a poor quality site they might settle on. Conversely, a distracted or even absent operator cannot cause the robots to become perpetually “stuck” because of the finite lifespan of the attractors and repellents. Malicious or erring users should similarly find influencing the robots to converge on a suboptimal site to be difficult, since robots have a chance of ignoring an attractor or repellent, they may end up discovering a high quality site despite misguidance from the operator.

Strategic Controls. Strategic controls are similar to parametric controls, but function on a “mission” level, rather than a “swarm” or “robot” level. Strategic controls in our simulation are still under development. Currently, sites are assigned arbitrary quality values during simulations, but a more realistic scenario might have sites that feature several different qualities based on distance, safety, size, or strategic import. One potential strategic control, called *quality attribution*, would allow for a different level of importance to be assigned to each feature, depending on the strategic objective. Changes to the importance of various features would in turn cause robots to evaluate a site's overall quality differently.

A second strategic control we intend to implement are playbooks [13] for hub-based colonies. In Sect. 2, we described a swarm system for selecting the best of n alternatives. This behavior constitutes a play. A variety of other plays could be created, including foraging, hub merging and splitting, etc. Once defined, the operator need only select the play, and the swarm would then automatically transition to a new set of behaviors.

These operations, functioning at a much higher level of abstraction, are more immediately intuitive than parametric controls and require far less micromanagement than environmental controls. However, they introduce much more risk for fault into the system, as the swarm, without any sort of model for strategy, cannot correct for poor operator decisions.

5 Conclusions

Human-swarm systems modeled after hub-based colonies, such as ants and bees, can potentially have very attractive properties. However, one of the challenges of implementing these systems is determining how the human should engage with the swarm to ensure that strategic mission objectives are met without, at the same time, compromising these properties. In this paper, we have advocated that the ideal way to do this is through shared control, wherein the human operator and the underlying situated dynamics of the swarm share the burden of decision-making. We have also discussed different ways in which shared control can be realized in such human-swarm systems, a discussion which culminated in a description of a our preliminary design of a human-swarm system.

In future work, we plan to evaluate and refine this system via user studies and further design, with the goal of continuing to develop generalizable principles for the design of fault-tolerant, human-swarm systems.

References

1. Brown, D.S., Goodich, M.A., Jung, S.-Y., Kerman, S.: Two invariants of human-swarm interaction. J. Hum. Robot Interact. **1**(5), 1–31 (2016)
2. Chen, J., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.: Situation awareness-based agent transparency. Technical report ARL-TR-6905, Army Research Lab (2014)
3. Crandall, J.W., Goodrich, M.A.: Characterizing efficiency of human robot interaction: a case study of shared-control teleoperation. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 2, pp. 1290–1295. IEEE (2002)
4. Crandall, J.W., Goodrich, M.A., Olsen Jr., D.R., Nielsen, C.W.: Validating human-robot systems in multi-tasking environments. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. **35**(4), 438–449 (2005)
5. Draper, G.M., Livnat, Y., Riesenfeld, R.F.: A survey of radial methods for information visualization. IEEE Trans. Vis. Comput. Graph. **15**(5), 759–776 (2009)

6. Gordon, D.: Ant Encounters: Interaction Networks and Colony Behavior. Princeton University Press, Princeton (2010)
7. Harriott, C.E., Seiffert, A.E., Hayes, S.T., Adams, J.A.: Biologically-inspired human-swarm interaction metrics. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1471–1475. SAGE Publications (2014)
8. Hinchey, M.G., Sterritt, R., Rouff, C.: Swarms and swarm intelligence. Computer **40**(4), 111–113 (2007)
9. Hu, H., Li, J., Xie, Z., Want, B., Liu, H., Hirzinger, G.: A robot arm/hand teleoperation system with telepresence and shared control. In: Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics (2005)
10. Jadbabaie, A., Lin, J., Morse, A.S.: Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Trans. Autom. Control **48**(6), 988–1001 (2003)
11. Kolling, A., Sycara, K., Nunnally, S., Lewis, M.: Human swarm interaction: an experimental study of two types of interaction with foraging swarms. J. Hum. Robot Interact. **2**(2), 103–128 (2013)
12. Kolling, A., Walker, P., Chakraborty, N., Sycara, K., Lewis, M.: Human interaction with robot swarms: a survey. IEEE Trans. Hum. Mach. Syst. **46**(1), 9–26 (2016)
13. Miller, C., Funk, H., Wu, P., Goldman, R., Meisner, J., Chapman, M.: The playbook approach to adaptive automation. In: Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (2005)
14. Millonas, M.M.: Swarms, phase transitions, and collective intelligence. In: Proceedings of Artificial Life III, pp. 417–445. Addison-Wesley (1993)
15. Nevai, A.L., Passino, K.M.: Stability of choice in the honey bee nest-site selection process. J. Theor. Biol. **263**(1), 93–107 (2010)
16. Olfati-Saber, R.: Flocking for multi-agent dynamic systems: algorithms and theory. IEEE Trans. Autom. Control **51**(3), 401–420 (2006)
17. Olsen, D.R., Wood, S.B.: Fan-out: measuring human control of multiple robots. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI) (2004)
18. Reynolds, C.W.: Flocks, herds and schools: a distributed behavioral model. ACM SIGGRAPH Comput. Graph. **21**, 25–34 (1987)
19. Scharre, P.: Robotics on the Battlefield Part II: The Coming Swarm. Center for a New American Security, Washington (2014)
20. Seeley, T.D., Tautz, J.: Worker piping in honey bee swarms and its role in preparing for liftoff. J. Comp. Physiol. A **187**(8), 667–676 (2001)
21. Seiffert, A.E., Hayes, S.T., Harriott, C.E., Adams, J.A.: Motion perception of biological swarms. In: Proceedings of the Annual Cognitive Science Society Meeting, pp. 2128–2133 (2015)
22. Shamaei, K., Che, Y., Murali, A., Sen, S., Patil, S., Goldberg, K., Okamura, A.M.: A paced shared-control teleoperated architecture for supervised automation of multilateral surgical tasks. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2015)
23. Su, H., Wang, X., Lin, Z.: Flocking of multi-agents with a virtual leader. IEEE Trans. Autom. Control **54**(2), 293–307 (2009)
24. Velagapudi, P., Scerri, P., Sycara, K., Wang, H., Lewis, M., Wang, J.: Scaling effects in multi-robot control. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2121–2126, September 2008

25. Walker, P., Nunnally, S., Lewis, M., Kolling, A., Chakraborty, N., Sycara, K.: Neglect benevolence in human control of swarms in the presence of latency. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 3009–3014. IEEE (2012)
26. Werfel, J., Petersen, K., Nagpal, R.: Designing collective behavior in a termite-inspired robot construction team. Science **343**(6172), 754–758 (2014)

The Evaluation of Remote Tower Visual Assistance System in Preparation of Two Design Concepts

Maik Friedrich[1(✉)], Stefan Pichelmann[2], Anne Papenfuß[1], and Jörn Jakobi[1]

[1] Institute of Flight Guidance,
DLR German Aerospace Center, Braunschweig, Germany
`Maik.Friedrich@dlr.de`
[2] Institute of Psychology, University of Bern, Bern, Switzerland

Abstract. In the last couple of years, the interest in remote tower operation systems has increased, as a concept to provide cost-efficient air traffic control service to airports. Attempts are made to stretch the concept to enable multiple remote tower control. A couple of challenges have been identified that need to be mitigated to make the concept feasible. For instance, studies revealed the disadvantage of reduced quality of the video panorama and an increased dispersion of information gathering. Design concepts were derived from the top-down and bottom-up mechanisms of human information processing to support multiple remote tower. These concepts could potentially mitigate the negative effects of a reduced quality of the video panorama and also the dispersion of information gathering. This paper focusses on the reduced quality of the video panorama and uses these design concepts to implement three visual assistance systems. These systems were tested within a laboratory experiment with 40 participants for effectiveness, efficiency, satisfaction, situation awareness, and workload. The microworld FAirControl was used to simulate an air traffic control monitoring task at a single airport. First, the visual assistance systems were compared against the baseline to assess the improvement of each system. Second, in an explorative approach, the visual assistance systems were compared to identify the advantages and disadvantages against each other. The results show that visual assistance systems can mitigate the influence of the quality-reduced video panorama. The results also indicate that usability of the visual assistance systems is crucial. Participant selective information overlay had the best overall performance.

Keywords: Remote tower operations · Visual attention · Workplace design · Interaction design · Visual assistance systems

1 Introduction

The German Aerospace Center (DLR) project RapTOr (Remote Airport Tower Operation Research, from 2005 to 2007) focused on the feasibility of remote tower operation (RTO) [1]. Remote tower operation was pushed forward by a joint venture

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 285–300, 2017.
DOI: 10.1007/978-3-319-58472-0_22

project (Remotely Operated Towers from 2006 to 2008) of the Swedish Civil Aviation Administration (LFV) and SAAB [2]. SAAB also coordinated the EU-Project ART (Advanced Remote Tower 2007–2009) [3] focusing on single remote tower control with possible extension to multiple airports. In the last couple years, the interest in remote tower operation (RTO) systems has taken a leap. This depends on an increased acceptance of the concept in the ATM community. RTO includes all technical equipment and procedures that are necessary to perform air traffic service (ATS) from a location where provision of ATS by direct visual observation is not possible. Instead a display shows the areas of responsibility controlled by the remote tower ATS Unit. The visual presentations can be transmitted anywhere as a video panorama. Following this concept, the idea of multiple RTO was proposed. This concept defines that one air traffic controller controls traffic at multiple airports in parallel. Multiple RTO for two small airports has already been successfully validated by LFV [4]. Nevertheless, the concept raises questions on how to best support monitoring tasks of the Air Traffic Control Officer (ATCO) in a multiple RTO setting in the future.

Monitoring is a main task of ATCOs in the tower. They use it to detect deviations between the preplanned and the real traffic situations. At the moment, two major influences on visual attention occur with the transition from conventional ‘single’ tower operation to multiple RTO. First, the quality from the out of window (OTW) view to the video panorama at the remote workplace is reduced compared to the quality of the human eye. Studies [5, 6] revealed that, since the RTO concept aims at reducing costs, the video panorama has some disadvantages in comparison to the OTW view, e.g. the resolution of the image. Second, visual attention of the ATCOs under multiple RTO conditions is spread. Möhlenbrink and Papenfuß [7] addressed the feasibility of controlling two airports at the same time, with a focus on visual attention. Visual attention in search tasks is mainly guided by top-down processes, namely experience and expectations [8]. In general, their results indicated that the human cognition is able to handle the applied concept for multiple RTO. Papenfuß and Friedrich [9] discussed two design concepts as support for the ATCOs to deal with the spread of visual attention. Due to the theoretical state of the concepts, Papenfuß and Friedrich [8] were not able to verify the technical feasibility. The work presented in this paper evaluates these design concepts by deriving, three visual assistance systems (VAS) and implementing these into a microworld. The conducted lab study evaluates VAS and their influence on the ATCOs performance in an air traffic control task, focusing on the first major influence. The VAS are evaluated with non-experts in a lab environment.

2 Design Concepts

Papenfuß and Friedrich [8] derived two remote design concepts from the results of their study and a workshop with an ATCO that helped interpreting the feasibility of their concepts. The basis for both concepts is that the ATCO developed in a traditional single airport environment demonstrated a good understanding of heuristics scheduling their tasks and attention. For example, a runway on an airport restricts that either one take-off or landing at the same time, and also automatically prioritizing the attention to this single event. This heuristic is challenged taking the influence of reduced quality and spread visual attention for multiple remote towers into account. Especially the reduced quality of the video panorama increases the work load for monitoring and search task. The following design concepts aim to mitigate the influences of RTO by providing additional support that is not available at the tower today.

2.1 Attention Control and Guidance

The first design concept is based on guiding the attention of ATCOs to areas of interest. Especially, the mental change between airports could be supported via attention guidance assistance to the ATCO. This could help to focus the diversion of visual attention and support the top-down processes, guiding visual attention during a search task. The attention control and guidance (AttConG) concept increases the salience on information that is eminent important for the ATCO. The AttConG foresees an analysis of the current situation to identify what is important. Therefore it could use input and data of interaction of the ATCO. For instance, the interaction between the ATCO with the electronic stripe system provides information about the aircraft and the task that is currently in focus, e.g. landing clearance for DHL 2345.

Papenfuß and Friedrich [9] proposed to increase salience on the geographical areas at the video panorama or aircraft that have increased task relevance. For instance, in the electronic flight strip display the according strip with the landing aircraft could be highlighted and in connection the aircraft itself could be highlighted in the video panorama.

A supplementary application is the support of the monitoring task. An eye tracking system could identify the aircraft that the ATCO is looking at and then additional information for this particular aircraft could be provided. It also could support the visual search by highlighting aircraft within the video panorama that are close to the eye direction of the ATCO. An advantage of the video panorama is that it can be augmented with additional information directly. Therefore the visualization of bounding boxes around an area or aircraft is simple.

2.2 Integration of Information into External View

"Head-up Only" is the second design concept to mitigate the negative influence of visual attention. Head-up Only distributes the same information to several attention areas to minimizing the spatial distribution of information. The implementation focus is thereby the integration of all information into the video panorama. As a main part of the concept the radar and stripe information are duplicated into the video panorama. Therefore, the important information from head-down displays, like aircraft type or clearances given to the flight, are accessible whilst looking at the video panorama. This is expected to increase the visual attention spend on the video panorama. By accessing the relevant information within the video panorama, the ATCO attention can focus on the situation at hand and his peripheral attention should support him in detecting unforeseen events even though his foveal attention is on other information.

For instance, the implementation of radar screen information could be realized by bounding boxes around all aircraft within the control zone. These bounding boxes could be extended by the stripe information (e.g. call sign, speed, height) connected to them. Following this example, the danger of information cluttering [10] has to be considered with growing amount of traffic.

2.3 Evaluation of Design Concepts

AttConG and Head-up Only are different approaches to support the ATCO especially working with an video panorama performing RTO. In the interest of transferring them to VAS, Papenfuß and Friedrich [9] proposed further research for the two concepts to sharpen their influence on monitoring behavior. Considering the different aspects within Sect. 2 for AttConG (support monitoring with eye tracking versus highlighting task relevant information) and Head-up Only (increase the information distribution within the video panorama), a stepwise approach should allow insight into the potential of each concept.

The implementation of different VAS covering only one aspect from either AttConG or Head-up Only is only effective if a set of criteria helps to determine the impact of each concept. Evaluation of the VAS should also focus on the individuals' performance during the ATC task. Huber [11] described how to evaluate individuals' results for such kinds of complex problems: the difference between the optimal final state of the target variables and the achieved final state of the target variables as a measure of "performance quality". The air traffic control (ATC) task's target variables are safety and efficiency. These two variables correspond to the terms of effectiveness (error rate) and efficiency (required time) as defined by the International Organization for Standardization [12]. But these two criterias only measure the final outcome. International Organization for Standardization additionally demands user satisfaction as an usability criteria which is defined as "freedom from discomfort, and positive attitude to the use of the product" [12].

Both design concepts aim at influencing the monitoring behavior of the ATCOs. Monitoring is strongly related to the concept of situation awareness [13, 14]. According to Endsley's model of situation awareness in dynamic decision making [15], situation awareness is part of a loop of continuous interaction between the environment and a person. Hence, achieving a high level of situation awareness is demanding on cognitive resources. If a VAS makes achieving situation awareness easier, more cognitive resources are available for the ATCO for other tasks. Even if the system does not help to ease the achievement of situation awareness it might help to increase the level of understanding of the situation.

Current research in ATC [16] emphasizes the importance of a reduction of ATCO workload. For example, Metzger and Parasuraman [17] found that increased automation reduced ATCO workload. Comparable to situation awareness, mental workload is also influenced by the current situation and available information and in turn influences ATCO performance. Therefore, workload was also selected as a criteria.

Based on the presented theoretical background of the ATC task, the following five criteria seem appropriate: two performance measures (effectiveness and efficiency), satisfaction, situation awareness, and workload. As the evaluated VAS focus on an easier and faster overview of the airport situation, the research hypotheses for each criteria assume a positive impact of the implemented VAS compared to a baseline without any VAS. As an explorative research question the criteria results for each VAS are also compared to each other to identify advantages or disadvantage between the different implementations.

3 Method

3.1 Participants

In total, 40 participants took part in the experiment. Their average age was 24.73 years (SD = 3.19 years) and 25 were male and they were students. The recruitment did not allow people with a color vision deficiency because of the usage of color by all VAS. 17 participants relied on optical aids. Every participant received a compensation of 20 €.

3.2 Visual Assistance Systems

The microworld FAirControl [18] was selected as implementation platform for the developed VAS. Figure 1 shows, how FAirControl represents the areas of a real airport. Each arriving aircraft must be given six clearances (Landing, Taxi In, Taxi In Apron, Push Back, Taxi Out, and Take Off) before it departs. A participant gives clearances by clicking on one of the six buttons (Fig. 1 center below). Participants select an aircraft by selecting the flight strip. Aircraft shown in the left flight strip bay only request and execute Landing, Taxi In and Taxi In Apron clearances, whereas aircraft shown in the right flight strip bay only request and execute Push, Taxi Out and Take Off clearances. Each flight strip shows the time, the aircraft identifier (Callsign), the necessary clearances, and whether the clearance is still requested or was already deployed.

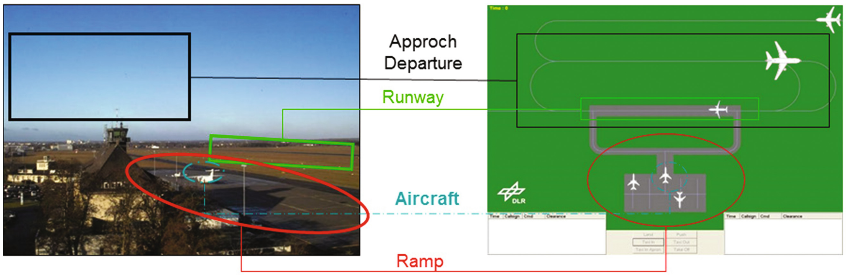


Fig. 1. Schematic representation of FAirControl (Color figure online)

FAirControl consists of three components: a human model, an airport process model, and an interaction model [10]. The human model was replaced by the participants, using a graphical user interface to interact with the airport. The airport process model controls the airport. Aircraft within FAirControl request and execute clearances based on their current status and position. The interaction model describes which information is presented to the ATCO and which clearances are given to the aircraft. Therefore, the VASs are implemented into the interaction model.

In total, three VAS were tested. Each of them aimed on improving salience with augmented information. All of them marked the position of aircraft on the screen by providing colored bounding boxes for the aircraft. Colors were chosen to provide high contrast to the surrounding area. All VAS used at least two colors: yellow and blue. A yellow frame around an aircraft symbolized a pending request of this aircraft, while a blue frame showed that no command was necessary. Throughout the experiment the same monitor was used for all participants to avoid effects due to different color representations.

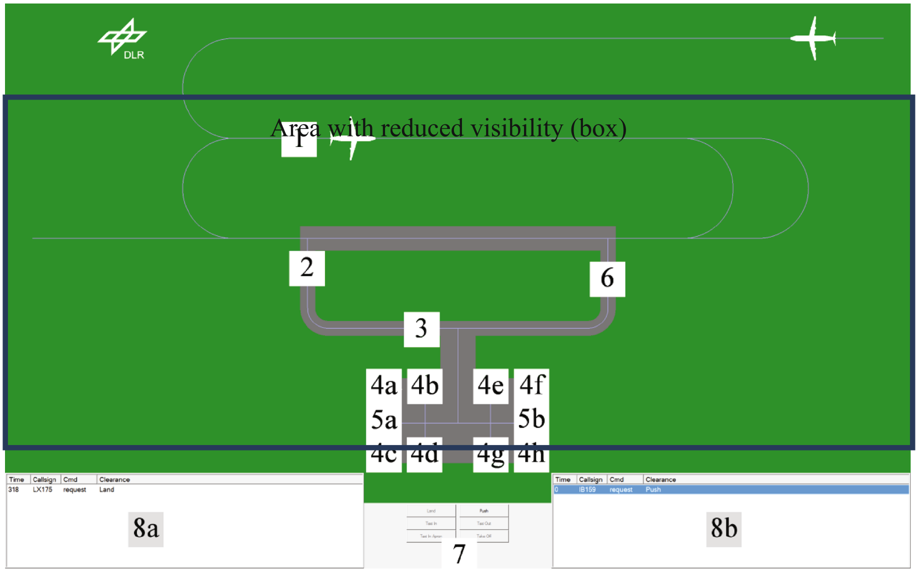
The AttConG concept was implemented in two versions. The first version is called VAS-SEL and surrounds only aircraft with a bounding box that are selected via theirs flight strip. It is assumed that these aircraft are most relevant for the ATCO, e.g. aircraft that are currently requesting clearance from the participant.

The second version is called VAS-EYE and incorporated eye tracking technology. VAS-EYE works similar to VAS-SEL but uses eye tracking to select the aircraft with highest task relevance. An algorithm identified the aircraft that was currently closest to the focus of the participant's eye gaze position. A low saturated orange frame was shown around this aircraft. If the participant pressed the space bar, the aircraft was selected and a blue or yellow frame was shown around it, additionally selecting the flight strip at the same time. The implementation uses the Tobii EyeX eye tracker with a 60 Hz capturing rate.

The Head-up Only concept was implemented as VAS-ALL. It showed bounding boxes around all aircraft that were currently depicted on the screen. VAS-ALL enhanced the visibility of all aircraft, independent from their importance to the current situation.

3.3 Scenario

The scenario consisted of 27 aircraft (). At the beginning, one aircraft is always on the apron and two arrive through the approach area within 20 s. If two aircraft request landing at the same time, other aircraft arrival was delayed automatically. The remaining 24 aircraft enter the airport roughly every 20 s depending on participants' behavior. The scenario finishes once the last aircraft takes off. The reduced visual quality of the video panorama was simulated by a partially transparent layer. The contrast was reduced stronger (91%) in the approach and departure area and less for the ramp. Hence, aircraft and airport visibility was highly diminished to the degree that aircraft were still visible but harder to spot and identify. The partially transparent layer had no effect on the visibility of the bounding boxes of the VAS (Fig. 2).



Time	Callsign	Cmd	Clearance
318	LX175	request	Land

Time	Callsign	Cmd	Clearance
0	IB159	request	Push

Fig. 2. User interface of FAirControl: request areas: Land (1), Taxi In (2), Taxi In Apron (3), Push (4a–h), Taxi Out (5a, b), Take Of (6). Command buttons (7) and tables containing the current aircraft (8a, b)

3.4 Design

A 1-factorial within subject design was applied to test all VAS and a trial without visual assistance (VAS-NONE) with each participant. Every trial used the same scenario. Participants were randomly assigned to one of four different sequences:

1. VAS-EYE — VAS-ALL — VAS-SEL — VAS-NONE
2. VAS-SEL — VAS-NONE — VAS-EYE — VAS-ALL
3. VAS-ALL — VAS-EYE — VAS-NONE — VAS-SEL
4. VAS-NONE — VAS-SEL — VAS-ALL — VAS-EYE

As mentioned above, the five criteria effectiveness, efficiency, satisfaction, situation awareness, and workload were used to determine the differences between the VAS: The criteria were measured either by internal calculation of FAirControl or by questionnaires. FAirControl recorded the required time to finish a trial as well as the amount and kind of mistakes. Mistakes were derived from actual ATC rules concerning the proximity between aircraft, e.g. at every time only one aircraft is allowed on the runway. The reciprocal of the mistakes quantity was used as a measure for effectiveness and the reciprocal of the required time as a measure for efficiency.

Questionnaires were applied to measure satisfaction, situation awareness, and workload for each trial. Satisfaction was measured using a single item “Please rate now how satisfied you have been with each of the assistance systems”. User answered on a 7-point scale. Situation awareness was measured with the situational awareness rating technique SART [19]. The SART uses three scales to calculate a general score of situation awareness: demands on attentional resources (by the situation), supply of attentional resources (by the user), and understanding of situation. Participants answered on a 7-point scale. Workload was measured with the NASA - task load index (NASA-TLX, 20-point scale) [20].

3.5 Procedure

After arrival, participants got a brief introduction to FAirControl which included an explanation of the simulated airport and a description of how to manage the aircraft. Furthermore, they were instructed which mistakes to avoid (e.g., collision between two aircraft). The participants completed a 10 min training with one aircraft already at the apron and another two aircraft arriving shortly after at the approach area. Afterwards, participants completed the pairwise comparisons used to calculate the NASA-TLX weights [13] and answered a demographic questionnaire.

Following the design, each participant completed four runs depending on the sequence they were randomly assigned to. The task of the participants was to control the arriving and departing traffic using the six clearances. Each run took approximately 11 min to complete. After every run, participants answered the NASA-TLX and SART. Satisfaction was measured once a participant finished the final run. Furthermore, participants were asked to write down any technical problems which they experienced.

4 Results

4.1 Effectiveness

Effectiveness scores of all runs were pooled by the VASs. Table 1 shows the descriptive statistics after outliers were removed. To ease interpretation, these statistics show the reciprocal effectiveness (i.e., the total amount of mistakes per run). Hence, higher values represent less effectiveness. In the VAS-ALL and VAS-SEL conditions, participants made fewer mistakes. In contrast, participants in the VAS-EYE condition made more mistakes than participants without any VAS.

Table 1. Descriptive statistics of Effectiveness^{-1} (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	38	17.82	6.71
VAS-EYE	39	21.92	5.78
VAS-ALL	38	18.82	6.41
VAS-NONE	37	20.49	7.10

All participants with outliers in any VAS condition were removed from the sample and a repeated measures test was conducted to test for significances. The Friedman test was used instead of the repeated measures one-way ANOVA, because the Shapiro-Wilk test showed significant deviations from a normal distribution for all VAS conditions. The Friedman test showed a significant effect of VAS conditions on effectiveness (n = 34, $\chi^2(3) = 11.43$, $p = .010$). The post-hoc tests between different VAS conditions were conducted according to Dunn [21] and p-values were adjusted with the Bonferroni correction. Table 2 shows the results for the Dunn test. The VAS-EYE condition yielded significant lower effectiveness scores than the VAS-ALL and VAS-SEL conditions, corresponding to medium negative effects. The difference between VAS-ALL and VAS-SEL was close to zero. VAS-ALL and VAS-SEL yielded small positive effects compared to VAS-NONE but these effects were not significant. In contrast, VAS-EYE showed a small negative but not significant effect on participants' effectiveness scores when compared with VAS-NONE.

Table 2. Results of the Dunn test for effectiveness differences between VAS conditions (n = 34; [a]values adjusted using Bonferroni correction; $^{*}p_{adj.} < .05$)

VAS conditions	z	p	$p^{a}_{adj.}$	d
VAS-NONE - VAS-SEL	1.55	.121	.727	0.38
VAS-NONE - VAS-ALL	1.22	.222	1.000	0.30
VAS-NONE - VAS-EYE	−1.46	.145	.872	−0.36
VAS-ALL - VAS-SEL	−0.33	.742	1.000	−0.08
VAS-ALL - VAS-EYE	2.68^{*}	.007	.045	0.69
VAS-SEL - VAS-EYE	3.01^{*}	.003	.016	0.78

4.2 Efficiency

As for effectiveness, efficiency scores of all runs were pooled by the VASs. Table 3 shows the descriptive statistics after outliers were removed. To ease interpretation, these statistics show the reciprocal efficiency (i.e., the required amount of time to finish the trial in minutes). Hence, higher values represent less efficiency. In the VAS-ALL, VAS-SEL, and VAS-NONE conditions, participants needed roughly eleven minutes on average. In contrast, participants in the VAS-EYE condition required twelve minutes on average.

Table 3. Descriptive statistics of Efficiency^{-1} in minutes (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	39	10.11	1.19
VAS-EYE	40	11.78	1.83
VAS-ALL	38	10.02	1.09
VAS-NONE	40	10.38	1.19

Again, all participants with outliers in any VAS condition were removed from the sample and the Friedman test was used because the Shapiro-Wilk test showed significant deviations from a normal distribution. The result was significant (n = 37, $\chi^2(3) = 32.44$, $p < .001$). The post-hoc test results are shown in Table 4. The VAS-EYE condition showed significantly lower efficiency scores than any other condition, corresponding to large effect sizes. All other comparisons were not significant and their effect sizes were close to zero.

Table 4. Results of the Dunn test for effectiveness differences between VAS conditions (n = 37; [a]values adjusted using Bonferroni correction; $^*p_{adj.} < .05$)

VAS conditions	*z*	*p*	$p^a_{adj.}$	*d*
VAS-NONE - VAS-SEL	0.36	.719	1.000	0.08
VAS-NONE - VAS-ALL	0.14	.893	1.000	0.03
VAS-NONE - VAS-EYE	-4.46^*	< .001	< .001	−1.21
VAS-ALL - VAS-SEL	−0.23	.822	1.000	−0.05
VAS-ALL - VAS-EYE	4.59^*	< .001	< .001	1.26
VAS-SEL - VAS-EYE	4.82^*	< .001	< .001	1.35

4.3 Satisfaction

Again, scores of all runs were pooled by the VASs. Table 5. shows the descriptive statistics. Participants reported highest scores for the VAS-SEL condition. In contrast, participants showed lowest satisfaction when no VAS was available.

Table 5. Descriptive statistics of Satisfaction (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	40	5.83	1.28
VAS-EYE	40	3.83	2.04
VAS-ALL	40	4.53	1.55
VAS-NONE	40	3.25	1.93

The Friedman test was used for Satisfaction because the Shapiro-Wilk test showed significant deviations from a normal distribution. The result was significant (n = 40, $\chi^2(3) = 33.60$, $p < .001$). The post-hoc test results are shown in Table 6. Participants showed significantly higher satisfaction in the VAS-SEL condition than in any other condition. Two differences in participants' satisfaction corresponded two large positive effects: VAS-SEL versus VAS-NONE as well as VAS-SEL versus VAS-EYE. Participants' satisfaction with VAS-ALL was moderately smaller than their satisfaction with VAS-SEL and moderately higher than their satisfaction with VAS-NONE but only the difference to VAS-SEL was significant. Furthermore a small positive effect for VAS-EYE was found when compared with VAS-NONE, but the effect was not significant.

Table 6. Results of the Dunn test for satisfaction differences between VAS conditions (n = 37; ^{a}values adjusted using Bonferroni correction; $^{*}p_{adj.} < .05$)

VAS conditions	z	p	$p^{a}_{adj.}$	d
VAS-NONE - VAS-SEL	5.50*	< .001	< .001	1.56
VAS-NONE - VAS-ALL	2.43	.015	.092	0.56
VAS-NONE - VAS-EYE	1.60	.109	.655	0.36
VAS-ALL - VAS-SEL	−3.07*	.002	.013	−0.73
VAS-ALL - VAS-EYE	0.82	.411	1.000	0.18
VAS-SEL - VAS-EYE	3.90*	< .001	.001	0.97

4.4 Situation Awareness

The Situation awareness scores of all runs were pooled by the VASs. Table 7 shows the descriptive statistics. Participants reported highest satisfaction scores for the VAS-SEL condition.

Table 7. Descriptive statistics of situation awareness (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	40	6.59	1.52
VAS-EYE	40	6.03	1.88
VAS-ALL	40	6.56	1.50
VAS-NONE	40	6.03	1.53

The results of the Shapiro-Wilk test showed no significant deviations from a normal distribution and therefore Mauchly's sphericity test indicated that the assumption of sphericity was not violated ($\chi^2(5) = 5.68$, p = .338). Hence, the repeated measures one-way ANOVA was conducted without corrections. It revealed a significant effect of VAS conditions on situation awareness ($F(3, 117) = 3.08$, $p = .030$). The post-hoc dependent t-tests for paired samples were conducted to analyze which of these four conditions differed significantly from other conditions. Again, p-values were adjusted using the Bonferroni correction. Results are shown in Table 8. Participants in the

Table 8. Results of the dependent t-Tests for paired samples for situation awareness scores comparing different VAS conditions (n = 40; ^{a}values adjusted using Bonferroni correction; $^{*}p_{adj.} < .05$)

VAS conditions	df	t	p	$p^{a}_{adj.}$	d
VAS-NONE - VAS-SEL	39	2.82*	.008	.045	0.45
VAS-NONE - VAS-ALL	39	2.12	.040	.242	0.34
VAS-NONE - VAS-EYE	39	0.41	.685	1.000	0.06
VAS-ALL - VAS-SEL	39	−0.12	.906	1.000	−0.02
VAS-ALL - VAS-EYE	39	1.76	.086	.514	0.28
VAS-SEL - VAS-EYE	39	1.86	.071	.425	0.29

VAS-NONE condition had significant lower situation awareness than participants in the VAS-SEL condition. VAS-ALL and VAS-SEL yielded small positive effects when compared against VAS-EYE and VAS-NONE but the corresponding t-tests were not significant.

4.5 Workload

Workload scores of all runs were pooled by the VASs. Table 9 shows the descriptive statistics. Average workload was lowest in the VAS-SEL condition, but absolute numbers for workload are rather medium (scale from 1 to 20).

Table 9. Descriptive statistics of workload (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	40	8.21	3.72
VAS-EYE	40	9.35	3.41
VAS-ALL	40	9.01	3.71
VAS-NONE	40	9.26	3.64

The results of the Shapiro-Wilk test showed no significant deviations from a normal distribution and therefore Mauchly's sphericity test indicated that the assumption of sphericity was not violated ($\chi^2(5) = 5.00$, $p = .416$). Hence, the repeated measures one-way ANOVA was conducted without corrections. It revealed no significant effect of VAS conditions on workload ($F(3, 117) = 2.41$, $p = .071$).

4.6 Relationships of the Criteria's

Calculation of the average effect sizes of all criteria's would be misleading due to their different importance (e.g., effectiveness is more important than satisfaction for ATC). The design of the study allows a comparison between the different VAS separately for each criterion. A major aspect of the comparison within the design is comparing each VAS against VAS-NONE, because it represents the baseline. Figure 3 presents a summary for each criterion. It shows the standardized VAS scores with Effectiveness, Efficiency and Workload scores inverted, to ease the interpretation. Therefore, values closes to the outer edge represent a better result.

Except for efficiency, VAS-SEL yielded small to high positive effects when compared with VAS-NONE. Two of these effects (satisfaction and situation awareness) were significant. For VAS-EYE, only one small positive effect was found for satisfaction when compared to VAS-NONE. It was not significant. VAS-ALL showed a medium positive effect for satisfaction and small positive effects for efficiency and situation awareness in comparison with VAS-NONE but none of them were significant.

The results comparing the different VAS against each other showed for VAL-SEL small to high positive effects for all criteria's in comparison with VAS-EYE and three of them were significant (effectiveness, efficiency, and satisfaction). Between

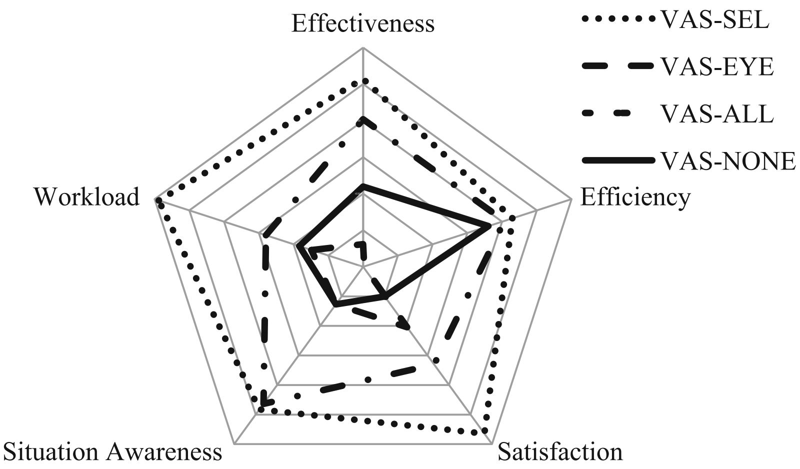


Fig. 3. Standardized VAS scores for each criterion with invented effectiveness, efficiency, and workload scores

VAS-SEL and VAS-ALL, VAS-SEL yielded a medium positive significant effect for satisfaction. Furthermore, VAS-EYE yielded one large negative significant effect (efficiency) and one small negative insignificant effect (effectiveness). Additionally, VAS-ALL yielded small to large positive effects for three criteria's (situation awareness, effectiveness, and efficiency) in comparison to VAS-EYE. Only the results for effectiveness and efficiency were significant.

5 Discussion and Conclusion

The work presented in this paper had the goal of evaluating different VAS that were derived of two different design concepts. The design concepts were defined to mitigate the problem of decreased visual quality in comparison of OTW and VR. Therefore the influence and possible support of each VAS on the ATC monitoring task was analyzed in detail. The five criteria (effectiveness, efficiency, satisfaction, situation awareness, & workload) showed only partial superiority of the VAS conditions (VAS-ALL, VAS-SEL, &VAS-EYE) against VAS-NONE.

The hypotheses of positive impact on the criteria were only significant for satisfaction and situation awareness for VAS-SEL. For effectiveness, efficiency, and workload none of the VAS were significantly better then VAS-NONE. On contrast, for efficiency VAS-EYE showed a significant negative effect compared to VAS-NONE. Nevertheless, the results for the different criteria help to rank the VAS against each other. For effectiveness VAS-EYE yielded lower results in performance than the other two VASs. The same results were found for efficiency. For satisfaction VAS-SEL showed significant medium positive effect when compared to VAS-ALL and even large positive effects when compared to VAS-EYE. Situation awareness and workload showed no significant difference between the VAS.

The VAS concentrated on an aspect of the two design concepts AttConG and "Head-Up Only". VAS-SEL and VAS-EYE use top-down mechanism thus correspond to the AttConG design concept. These concepts make use of the task relevance (VAS-SEL) and expectations (VAS-EYE) of the ATCO to superimpose additional information into the video panorama. VAS-SEL uses mouse-input via the flight strips, VAS-EYE eye gaze position on the video panorama.

VAS-SEL was significantly better than VAS-EYE concerning effectiveness, efficiency and satisfaction. The following two reasons could account for that. First, VAS-SEL used the pre-processed information of flight strips. Selection of aircraft was accurate and allowed participants to determine their exact position in the video panorama. Whilst working in VAS-EYE, participants were required to have expectations about aircraft position in the video panorama. This indicates that additional training could increase the usage of VAS-EYE. VAS-SEL also depends on additional data sources and does not cover situations where unexpected aircraft arrive. Second, the insufficient usability and accuracy of the eye-tracking system led to the unacceptable efficiency of VAS-EYE when compared to VAS-NONE. When a system or function is cumbersome to use it can account for longer handling times. The ATC task is time critically, and therefore longer handling times might also correspond with less effectiveness. On the other hand, the satisfaction with VAS-EYE was better than VAS-NONE. It is of interest to repeat the study with a more accurate eye tracking system.

VAS-ALL was derived from the "Head-Up Only" concept. Information was directly superimposed into the video panorama. Thus, positions of aircraft under control of the participants were visible. So, for instance assessment of runway status (empty or occupied) was simple. VAS-ALL did not cause information clutter, but this might be a limitation of the microworld environment.

By examination of the methodology for this paper the following three major limitations have to be considered. First, the Participants of this study were students and no ATCOs. For this implementation of the different VAS and considering the rather low level ATC task, we decided that a bigger sample size for the statistical interpretation would be more suited than a small sample size. Hence, for this specific task the performance of ATCOs and students might not differ significantly. Second, FAirControl's interface provided only visual monitoring with aircraft moving much faster than in reality. Nevertheless, FAirControl is believed an appropriate environment to test different VAS in an early stage of development. The increased speed of aircraft in FAirControl compensates for the lower workload as other duties of the ATCO are not implemented. Third, 45% of all participants mentioned problems with the eye tracker even though the calibration went without any problems and prior tests showed only negligible deviations between the traced point and the point where participants were asked to look at. People found it hard to select small and closely positioned aircraft. This might have resulted in giving commands to unintended aircraft. This could also explain the quality of the VAS-EYE results.

Therefore, the recommendation is drawn that due to the results pattern of the VAS-SEL treatment aircraft with active request should be highlighted on a realistic remote tower visual presentation.

Summarizing the conclusions for design concepts, an integrated VAS might be best to suit requirements of the real remote tower working position. The amount of information for visual assistance should also be varied to assess effects like information clutter. In this study, aircraft position was superimposed, an information that is highly relevant and mandatory for ATCOs to build up their mental traffic picture. VAS-SEL has the most positive effects for all criteria and is therefore the most promising VAS to pursuit. VAS-ALL yielded the second best results and VAS-EYE showed the lowest results. It is of interest to understand in how far VAS-SEL also mitigates diversion of attention and the detection of unforeseen events.

The research presented in this paper brings use one step closer to an integrated VAS that could support the single and multiple remote tower operation and mitigate the restrictions that come with this kind of technology. Further research should concentrate on the evaluation of VAS-SEL, VAS-ALL, and their combination in a real RTO scenario. Also to extend future research, FAirControl also implements the Integration Guideline for Dynamic Areas of Interest (IGDAI) [22] that allows a detail view on dynamics moving Areas of interest and therefore helps to identify how the visual assistance influence the eye movement distribution.

References

1. Schmidt, M., et al.: Remote airport tower operation with augmented vision video panorama HMI. In: Proceedings of 2nd International Conference Research in Air Transportation ICRAT, Belgrade, pp. 221–230 (2006)
2. Saab Security: Remotely Operated Tower - The Future of Being Present (2008)
3. van Schaik, F.J., Lindqvist, G., Rössingh, H.J.M.: Assessment of visual cues by tower controllers. In: 11th IFAC/IFIP/IFORS/IEA, Valenciennes, France, p. 19 (2010)
4. Mullan, C., et al.: OSED for Remote Provision of Air Traffic Services to Aerodromes, Including Functional Specification for Single & Multiple Aerodromes, Brussel (2012)
5. Ellis, S.R., Fürstenau, N., Mittendorf, M.: Frame rate effects on visual discrimination of landing aircraft deceleration: implications for virtual tower design and speed perception. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **55**, 71–75 (2011)
6. Friedrich, M., Möhlenbrink, C.: 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 (2013)
7. Möhlenbrink, C., Papenfuß, A.: ATC-monitoring when one controller operates two airports: research for remote tower centres. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **55**, 76–80 (2011)
8. Wickens, C.D.: Using interference models to predict performance in a multiple-task UAV environment-2 UAVs. In: Security (2003)
9. Papenfuß, A., Friedrich, M.: Head up only – a design concept to enable multiple remote tower operations. In: 35th Digital Avionics Systems Conference (2016)
10. Knabl, P., Többen, H.: Symbology development for a 3D conformal synthetic vision helmet-mounted display for helicopter operations in degraded visual environment. In: Engineering Psychology and Cognitive Ergonomics: Understanding Human Cognition (Part I), Las Vegas, USA. Springer, Heidelberg (2013)

11. Huber, O.: Complex problem solving as multistage decision making. In: Frensch, P., Funke, J. (eds.) Complex Problem Solving: The European Perspective, pp. 151–173. Lawrence Erlbaum Associates, Hillsdale (1995)
12. International Organization for Standardization, Ergonomic requirements for office work with visual display terminals (VDTs). In: Part 11: Guidance on Usability (1998)
13. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Hum. Factors: J. Hum. Factors Ergon. Soc. **37**, 65–84 (1995)
14. Wickens, C.D., Mavor, A.S., McGee, J.P.: Flight to the Future. Human Factors in Air Traffic Control. National Academy Press, Washington, D.C. (1997)
15. Endsley, M.: Towards a theory of situation awareness in dynamic systems. Hum. Factors **37** (1), 32–64 (1995)
16. Burke, K.A., Wing, D.J., Lewis, T.: Pilot subjective assessments during an investigation of separation function allocation using a human-in-the-loop simulation. In: 2013 Aviation Technology, Integration, and Operations Conference (2013)
17. Metzger, U., Parasuraman, R.: Automation in future air traffic management: effects of decision aid reliability on controller performance and mental workload. Hum. Factors: J. Hum. Factors Ergon. Soc. **47**(1), 35–49 (2005)
18. Möhlenbrink, C., Oberheid, H., Werther, B.: A model based approach to cognitive work analysis and work process design in air traffic control. In: Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter, Braunschweig (2008)
19. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: Situational Awareness: Limitations and Enhancement in the Aviation Environment, pp. 6/1–6/14. Advisory Group for Aerospace Research and Development, Neuilly-Sur-Seine, France (1990)
20. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Hum. Ment. Workload **1**, 139–183 (1988)
21. Dunn, O.J.: Multiple comparisons using rank sums. Technometrics **6**(3), 245–252 (1964)
22. Friedrich, M., Rußwinkel, N., Möhlenbrink, C.: A guideline for integrating dynamic areas of interests in existing set-up for capturing eye movement: looking at moving aircraft. Behav. Res. Methods (2016)

The Investigation Human-Computer Interaction on Multiple Remote Tower Operations

Peter Kearney[1(✉)], Wen-Chin Li[2], Graham Braithwaite[2], and Matthew Greaves[2]

[1] ATM Operations and Strategy, Irish Aviation Authority, Dublin, Ireland
`peter.kearney@iaa.ie`
[2] Safety and Accident Investigation Centre, Cranfield University, Cranfield, UK

Abstract. The aim of current research is to develop an effective human-computer interaction framework for multiple remote tower operations. Five subject-matter experts familiar with multiple remote tower operations and human performance participated in current research. The Hierarchical Task Analysis (HTA) method is used to break down activities, scenarios, and tasks into single separate operations. The step by step breakdown of multiple remote tower operations included ATCO's operational behaviors involving human-computer interaction such as interaction with EFS, OTW, RDP, and IDP during task performance were noted. Designing and managing human-computer interactions require an understanding of the principles of cognitive systems, allocation of functions and team adaptation between human operators and computer interactions. It is a holistic approach which considers distributed cognition coordination to rapidly changing situations. The human-centred design of multiple remote tower operations shall be based on a strategic, collaborative and automated concept of operations, as the associated high performance of remote tower systems in conflict detection and resolution has the potential to increase both airspace efficiency and the safety of aviation. The focus is on the human performance associated with new technology in the RTC and the supported tools used by an Air Traffic Control Officer, to ensure that these are used safely and efficiently to control aircraft both remotely and for multiple airports. The advanced technology did provide sufficient technical supports to one ATCO performing a task originally designed to be performed by several ATCOs, however, the application of this new technology also induced huge workload on the single ATCO.

Keywords: Multiple remote tower operations · Human computer interaction · Situation awareness · Workload

1 Introduction

The innovative concept of multiple remote tower operations is an evolution of the use of remote tower technologies and can maximize cost savings through the implementation of Remote Tower Operation (RTO). In recent years, Air Traffic Management has

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 301–309, 2017.
DOI: 10.1007/978-3-319-58472-0_23

had to confront difficulties of infrastructure and airspace capacity which has resulted in aircraft delays. To address these concerns, the Single European Sky initiative, a European Union project, has been established up to improve safety, minimize costs and environmental impact, and at the same time increase efficiency and capacity in order to meet the requirements of expanding air traffic (Eurocontrol 2014). A novel solution of these issues is for a single air traffic controller (ATCO) to deliver control services to multiple airports from a single working position in appropriate traffic load circumstances. The multiple remote tower operation offers the potential of providing aerodrome control services for several small airports from a remotely-located control centre, without needing direct physical presence at the airports under control The aim of multiple remote tower operations is to deliver benefits in line with SESAR's high-level objectives, to increase ATCO's situation awareness, to create productivity for training, and to enhance contingency and reduce workload. Air traffic controllers must make rapid judgments of the situation that is being presented by their respective ATM systems, and then take appropriate decisions to ensure aviation safety. Interestingly research spanning from 1977 to 2008 has demonstrated that decision errors in aviation may be contributing to up to 60% of all aviation accidents (Jensen and Benel 1977; Buch and Diehl 1984; Diehl 1991; Li and Harris 2008).

The initial remote tower control of low traffic airports has emerged as a new paradigm to reduce the costs of air traffic control service provision. The application of advanced technology suggests that air traffic controllers can visually supervise airports from remote locations using videolinks to monitor many airports from a remote tower center (RTC). It is also clear that visual features of aircraft detection, recognition, and identification by RTO go beyond that required by regulators and air navigation service providers (ANSP's) (Furstenau et al. 2014). As the concept of RTC was being researched, it became clear that it would differ fundamentally from traditional modes of local tower operation. Cameras and sensors could be placed anywhere on the field, and ATCOs would be presented a virtual picture of reality, enhanced by a number of advanced technical devices such as panoramic digital reconstruction with high resolution pan-tilt zoom (PTZ), and electronic flight strip (EFS). The design and development of the human-computer interaction (HCI) for RTC can be supported by a formal cognitive work and task analysis. The results of this cognitive work and task analysis serve as input data for the simulation of the controller decision making processes at the controller working position (CWP). The anticipated outcomes of this project will develop a conceptual framework of HCI for multiple remote tower operations based on ATCO's cognitive processes and task performance.

2 Method

2.1 Participants

Five subject-matter experts participated in six focus group sessions. The subject matter experts ages ranged between 41 and 53 year old (M = 47.2, SD = 4.5); the working experience as qualified ATCO is between 13 and 25 years (M = 17, SD = 5.9).

2.2 Apparatus

The multiple remote tower research has included following equipment.

2.2.1 The Remote Tower Centre is equipped with 2 Remote Tower Modules comprising of 15 screens in each (14 active & 1 spare). Each of the modules is equipped with the SAAB Electronic Flight Strip (EFS) and Radar Data Processing (RDP) display which is used only as a distance to touch down indication and not to provide a Radar Service. Each of the modules accommodates 2 controller positions, Surface Movement Control (SMC) and Air Movement Control (AMC).

2.2.2 The SAAB Remote Tower camera system (Fig. 1) was installed at Shannon airport and Cork airport Remote Tower Sites. The Cameras are located at suitable positions to provide the exact same viewing aspect as the current Tower. The out the window (OTW) visualization is made up of 15 full HD displays in a 220° configuration. 14 displays are normally used to present the images from the 14 cameras, while the last display is a stand-by unit in the event of equipment failure. The displays match the camera resolution of 1920 × 1080 pixels, and have a refresh rate of 60 Hz.

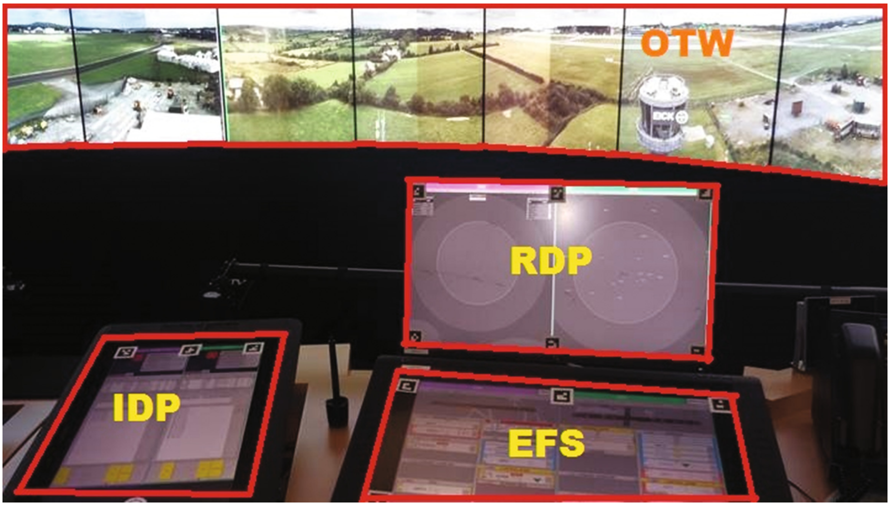


Fig. 1. The Module of multiple remote tower control centre comprised by Electronic Flight Strip (EFS), Out of the Window (OTW), Radar Data Processing (RDP), and Information Data Processing (IDP)

2.3 Scenario

One ATCO controlling a Boeing-737 landing at Shannon airport whilst controlling a Boeing-737 departing from Cork airport from a Remote Tower Centre situated 120 miles away at Dublin airport.

2.4 Data Collection Process

Five subject-matter experts familiar with multiple remote tower operations and human performance participated in current research. The Hierarchical Task Analysis (HTA) method was used to break down activities, scenarios, and tasks into single separate operations. This methodology enables a comprehensive step-by-step description of the task activities associated with the scenario described above (Annett 2004). The scenario of landing at Shannon airport and departing from Cork airport were broken down by 5 domain experts based on the principles of HTA for multiple remote tower operations. The step by step breakdown of multiple remote tower operations included ATCO's operational behavior involving human-computer interaction such as interactions with EFS, OTW, RDP, and IDP (Fig. 1) during task performance were noted. The dimensions of HCI on multiple remote tower operations including time to complete a task (sub-task and operational action), effort of cognitive workload, accuracy, and consistency were analyzed.

3 Results and Discussion

The application of HTA for analyzing the task of multiple remote tower operations was conducted at the remote tower control center, where the participants could practically use all the relevant systems and equipment to simulate task performance. This permitted an accurate assessment of the actions required and the cost of the effort and time required to complete the operational steps, such as checking the RDP to estimate the distance and timing of arrival flight, monitoring moving aircraft/vehicles on the runway by OTW, or input information into EFS. The objective was to understand the limitations of human-computer interaction on multiple remote tower operations, as a crucial first step in the certification process for multiple remote tower operations in future air traffic management.

3.1 Scope of the Task Under Analysis

The remote tower services were demonstrated in sequence for both Shannon and Cork airports during periods of low traffic density. The application of OTW visualization reinforced by RDP and EFS technology and the existing data and communications network will provide the necessary technical supports for the provision of air traffic services remotely and without any degradation to the standard of ATS provided. Safety management processes and procedures were applied to ensure that levels of safety were at least equal to those which are provided by the local control towers at both airports. The scope of task under analysis is 'one ATCO safely directing a Boeing-737 landing on Shannon airport whilst controlling a Boeing-737 departing from Cork airport from a remote tower center situated at Dublin airport'.

3.2 Data Collection Process

Data regarding the goals, standard operation procedures (SOPs), task steps involved the analysis of the technology used, human-computer interaction, teamwork, and task constraints related to a simultaneous landing at Shannon airport and departure from Cork airport were collected by 3 focus group sessions in the RTC at Dublin airport. HTA development commenced once sufficient data regarding required task performance was collected. The main focus of this data collection was on the cognitive process of ATCO's controlling aircraft landing and departing from two different airports, a critical concern related to a new configuration where one ATCO is performing a task previously designed to be done by six ATCOs', and the solution of dealing with the associated potential risks.

3.3 Define the Overall Goal

The single overarching goal is maintenance of the safety level of air traffic services without any degradation whilst a single ATCO controls two aircraft landing and departing from two different airports using new technology. It is a novel Air Traffic Control challenge for one ATCO performing two different air traffic control tasks for two aircraft at two different airports. SOPs for ATCO controlling aircraft landing and departing are considerably different. Task analysis methodologies have been defined by (Kirwan and Ainsworth 1992) as the analysis of actions and cognitive processes carried out by an operator (or a team of subjects) to reach the objectives of a specific system. The application of relevant systems along with step-by-step operational actions makes HTA particularly useful for the evaluation of task performance in response to overall goal achievement.

3.4 Determine Task Sub-goals

HTA enables a comprehensive step-by-step description of the task activity under consideration to be achieved, and has become the most extensively used of all the Human Factors methodologies available (Annett 2004; Stanton 2006; Stanton et al. 2013). The flexibility and practicality associated with the HTA technique has seen it applied in a diverse range of domains. Despite the comprehensive insight provided by the HTA methodology, the HTA output will only provide descriptive information – rather than analytical data – for the task under analysis. Consequently, the description presented by the task analysis is typically the input to additional methodologies for further analysis, such as the Human Error Identification (HEI) technique (Stanton et al. 2010). Notwithstanding the importance of the calls and visual controls which contribute to a shared mental picture of the situation between operators such as pilots, ATCOs, ground vehicle drivers and rescue teams, it was decided to give the controllers a specific version of the diagram with these steps to complete the task by interacting with the system. The sub-goals include co-ordination calls from all stakeholders, monitoring all moving vehicles/aircraft between two airports, providing line up/take off instruction, and establishing communication (such as handover to the next ATS sector and parking stand information. The subject-matter experts raised a safety concern regarding the time interval for task performance, as ATCO's will be under time pressure constantly not only due to shifting attention between two airports for many moving targets (aircraft, vehicles, birds), but they must also make decisions providing control instructions for air/surface movement control (Fig. 2). Sufficient time to complete a sub-goal is a critical safety issue of HCI with multiple remote tower operations, as it influences the efficiency of interface design, time pressure and perceived workload.

3.5 Sub-goal Decomposition

Task decomposition is a structured way of developing sub-goals into a series of detailed descriptions relating to each operational step (Stanton et al. 2008). Task decomposition involves the breakdown of a task description that explains how each step of the task is performed and the time needed for completing the task. Adding the time to complete each operational step is critical to identifying potential risk related to human performance, as there are some prioritized issues for task performance between two airports and two different tasks, landing and departing. Available time is always critical for flight operations. This allows the total information for each operational step to be decomposed into a series of statements describing the tasks of multiple remote tower operations. The sub-heading used to decompose the task steps should be chosen by the analyst based on the requirements of the analysis. The bottom level of HTA should always be an operational step and specify what action needs to be taken. For example, the task decomposition of (1.3) First contact from EINN Arrival (20 s to complete the task) can be breakdown as (1.3.1) Acknowledge call and reply (8 s); (1.3.2) Assume FLT strip for aircraft (2 s); (1.3.3) Make cross check on EICK by OTW (2 s); (1.3.4) Cross check position on RDP (2 s); (1.3.5) Utilize OTW picture to identify aircraft on approach (3 s); (1.3.6) Scan predicted aircraft track using PTZ on OTW (3 s). To maintain ATCO's situation awareness, it is not only necessary to shift attention between Shannon and Cork constantly by performing interlude sub-task (1.1, 1.3, 1.5, 1.7, 1.9 & 1.11 for Shannon, and 1.2, 1.4, 1.6, 1.8 & 1.10 for Cork), but the ATCO is also required to conduct a cross check for each sub-goal decomposition (green operational steps on Shannon, such as 1.1.3, 1.3.3 & 1.5.4; and red operational steps on Cork, such as 1.2.4, 1.4.5 & 1.6.5) (Fig. 2). The agreement among subject-matter experts was that task description should provide enough information to determine exactly what has to be done to complete each task element safely by increasing ATCO's situation awareness for air movements and surface movements at each airport.

3.6 Plans Analysis

Once all of the sub-goals and operational steps have been fully described, plans need to be added. The plans of current HTA 'one ATCO safely directing a Boeing-737 landing at Shannon airport while controlling a Boeing-737 departing from Cork airport from a

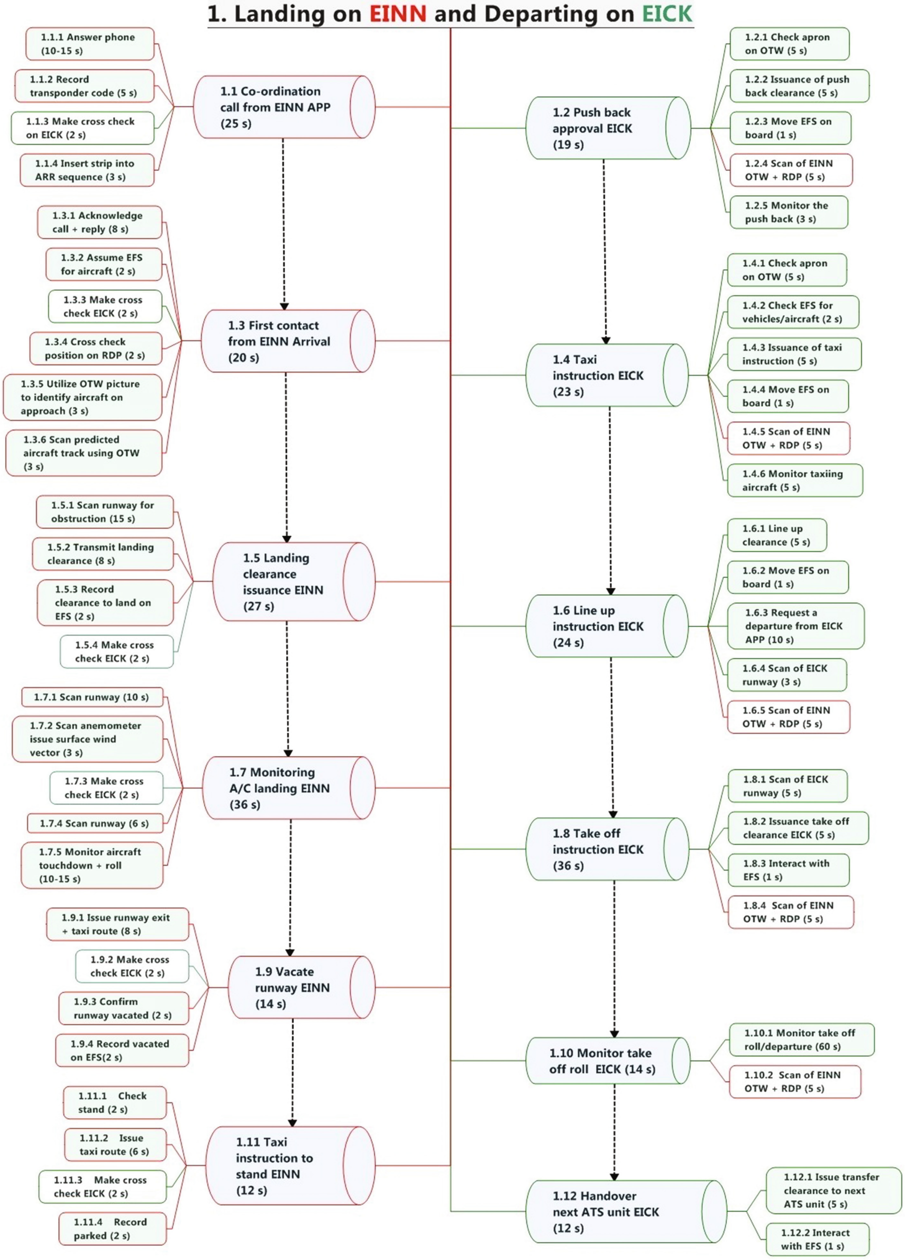


Fig. 2. The HTA of multiple remote tower operation for one ATCO commanding two B-737 land on Shannon and departing on Cork (Color figure online)

remote tower center situated in Dublin airport' is linear, for example do 1 (1.1, 1.2, 1.3…) then 2 (2.1, 2.2, 2.3…) then 3 (3.1, 3.2, 3.3…). Despite the use of the term landing and departing used to describe the scenario, there is always a prioritization in human being's cognitive process and decision-making involved in the task performance. Human operators are simply just not able to speak to two different people. Nor are they able to listen to two different flight crews describe their problems and provide suitable solutions by radio transmission to these flight crews. The linear sequence is an important application in HTA, and occasionally the linear sequence must be sufficiently flexible to manage critical urgent situations, such as an intervention in the case of an unexpected event such as a runway incursion.

4 Conclusion

HTA was originally developed in response to the need to understand cognitive tasks, it achieves this by describing the activities under analysis in terms of a hierarchy of goals, sub-goals, operations and plans. Further development of HTA involves the application of numerous other human factors analysis methods including human error identification function allocation, workload assessment, interface design, and training (Stanton et al. 2013). Designing and managing human-computer interactions require an understanding of the principles of cognitive systems, allocation of functions and team adaptation between human operators and computer interactions. It should be a holistic approach considering how distributed cognition coordination in rapidly changing situations can be safely achieved. The human-centred design of multiple remote tower operations shall be based on a strategic, collaborative and automated concept of operations, as the associated high performance of remote tower systems in conflict detection and resolution has the potential to increase both airspace efficiency and the safety of aviation (Schuster and Ochieng 2014). The focus is on the human performance associated with new technology in the RTC supported tools used by an Air Traffic Control Officer, and ensuring that these are used safely and efficiently to control aircraft both remotely and for multiple airports. The advanced technology did provide sufficient technical supports to one ATCO performing a task originally designed to be performed by several ATCOs, however, the application of this new technology also induced huge workload one the ATCO. This creates a need for further research on how to deal with the HCI issues identified, including high workload for multiple remote tower operations in the future.

References

Annett, J.: Hierarchical task analysis. In: Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick, N.A. (eds.) Handbook of Human Factors and Ergonomics Methods, pp. 329–337. CRC Press, Boca Raton (2004)

Buch, G., Diehl, A.: An investigation of the effectiveness of pilot judgement training. Hum. Factors **26**(5), 557–564 (1984). doi:10.1177/001872088402600507

Diehl, A.: The effectiveness of training programs for preventing aircrew error. In: Jensen, R.S. (ed.) Proceedings of the Sixth International Symposium on Aviation Psychology, vol. 2, pp. 640–655. The Ohio State University, Colombus (1991)

Eurocontrol: Eurocontrol Seven-year IFR Flight Movements and Service Units Forecast: 2014–2020 (Reference No. 14/02/24-43). Brussels, Belgium (2014). www.eurocontrol.int/sites/default/files/content/documents/official-documents/forecasts/seven-year-flights-service-units-forecast-2014-2020-feb2014.pdf

Furstenau, N., Mittendorf, M., Friedrich, M.: Model-Based analysis of two-alternative decision errors in a videopanorama-based remote tower work position. In: Proceedings of 6th International Conference on Human Computer Interaction, Crete, Greece, pp. 143–154 (2014)

Jensen, R.S., Bernel, R.A.: Judgment Evaluation and Instruction in Civil Pilot Training. Federal Aviation Administration, Washington, D.C. (1977)

Kirwan, B., Ainsworth, L.K.: A Guide to Task Analysis. Taylor and Francis, London (1992)

Li, W.-C., Harris, D.: The Evaluation of the effect of a short aeronautical decision-making training program for military pilots. Int. J. Aviat. Psychol. **18**(2), 135–152 (2008). doi:10.1080/10508410801926715

Schuster, W., Ochieng, W.: Performance requirements of future trajectory prediction and conflict detection and resolution tools within SESAR and NextGen: framework for the derivation and discussion. J. Air Transp. Manage. **35**, 92–101 (2014)

Stanton, N.A.: Hierarchical task analysis: developments, applications and extensions. Appl. Ergon. **37**(1), 55–79 (2006)

Stanton, N.A., Salmon, P., Harris, D., Marshall, A., Demagalski, J., Young, M.S., Waldmann, T., Dekker, S.: Predicting pilot error on the flight deck: validation of a new methodology and a multiple methods and analysts approach to enhancing error prediction sensitivity. Elsevier (2008)

Stanton, N.A., Harris, D., Salmon, P.M., Demagalski, J., Marshall, A., Waldmann, T., Dekker, S., Young, M.S.: Predicting design-induced error in the cockpit. J. Aeronaut. Astronaut. Aviat. **42**(1), 1–10 (2010)

Stanton, N.A., Salmon, P.M., Rafferty, L.A., Walker, G.H., Baber, C., Jenkins, D.P.: Human Factors Methods: A Practical Guide for Engineering and Design, 2nd edn. Ashgate, Farnham (2013)

Integrated Design of System Display and Procedural Display in Advanced NPP Control Rooms

Yiran Ma[1,2](✉), Qin Gao[1,2], Fei Song[1,2], and Yufan Wang[1,2]

[1] Tsinghua University, Beijing, China
mayiran94@126.com, gaoqin@tsinghua.edu.cn
[2] Shanghai Nuclear Engineering Research and Design Institute, Shanghai, China

Abstract. When dealing with abnormities, operators' display-management workload takes up much attention resource, bringing about safety issues. This study is based on the idea of a procedure-integrated display design, which has passed preliminary testing in a previous study. The new design introduces navigation system that coordinates the two types of displays, and adapt this procedure-integrated display to real NPP HSI requirements. The procedure-integrated display has been realized on a simulated control room platform under two scenarios. The fully-developed system emulates all functions and effects in a real NPP control room. We invite two highly-experienced operators for usability testing, and four NPP experts for review. Evaluation result shows that this procedure-integrated display is applicable to help operators improve performance and reduce mental workload. Testing also reveals that the disorientation level is low in the new navigation system, and operators report high acceptance and learnability towards the new design. By analyzing the results from the two parts of evaluations, this paper discusses an improved design.

Keywords: Integrated display · Operation procedures · Nuclear power plant · Advanced control room · Human reliability · System development

1 Introduction

After the computerization of human-system interface as well as operation procedures, control room of nuclear power plant has faced with the challenge of maintaining human reliability.

On one hand, advanced NPP control rooms have computerized instrumental and operation panels which were originally spread across the walls. The new form of interface can support huge amount of information and has the advantage of being flexible and powerful in assisting operators. The drawbacks, however, include the following three aspects: (a) Information overload; (b) Restricted view [1]; (c) Huge number of pages.

On the other hand, advanced NPP control rooms have made the transition from paper-based operation procedures to computerized procedures. This has made it easier to retrieve information and improved performance of operators. Nevertheless, operators have to switch between operation procedures, system displays and control displays,

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 310–318, 2017.
DOI: 10.1007/978-3-319-58472-0_24

while making decisions to solve the emergencies. In this process, they have to maintain the working memory of operation steps, which will take up the attention resources for primary tasks [2].

Previous researches pointed out that, by integrating system display and operating procedure, there would be a reduction in operators' mental workload and an improvement in their performance [3–6]. In one of the study, Chen designed a unit-based integrated display. It was proved that this unit-integrated design can improve performance under emergent scenarios. However, her design was only tested on a simplified single display and was not integrated into the control room system. Meanwhile, after integrating operation procedures, the system will be organized in task-based structure rather than a system-based structure, creating a challenge to the navigation system. Previous studies have not yet covered this issue.

The contributions of this study are:

- We improved the design of a procedure-integrated display, introduced navigation system to adapt the integrated design to real NPP control room norms and tasks;
- We implemented the design on a simulated NPP control room platform;
- We conducted evaluation on the procedure-integrated design and proposed a future plan for improvements.

2 Related Work

2.1 HSI Design of NPP Control Rooms

In the review guidelines published by NRC, interface management was defined as the behavior of the operator moving within the organization structure to obtain information. The guidance noted that interface management task would increase workload, therefore operators would be cautious about opening new windows especially when workload is already at a high level [1]. Other researches showed that operators should maintain accessibility to higher-level functional structures to support diagnosing anomalies [7–9].

An outstanding problem in present control room HSI design is the conflict of system-based navigation and task-based navigation. When handling emergent tasks under time pressure, operators would have huge difficulties searching for information in a conventional system-based navigation system. Previous studies found that task-based navigation improved performance [10, 11], but would be difficult to adapt to conventional tasks.

2.2 Navigation Design of Other Complex Information System

NPP control room HSI uses hyperlinks in many contexts (including text, graphics). Hyperlinks often lead to disorientations, as users will navigate to a new page without knowing where they have just been, where to go next, and their current position [12]. A study focused on computer-aided learning [13] found that in the context of using hyperlinks, using non-linear hyperlink organization (any interface can be linked to other levels of interface) would take up more working memory and have more negative impact on performance than using linear hyperlink organization. Navigation aid system, on the other hand, would effectively alleviate the problem of the users' disorientation problem [14]. Although the navigation aid would be helpful in reducing loss, redundancy of multiple navigation systems within the same system can cause greater distress [15, 16].

2.3 Previous Work on Integrated Design

In Chen's study on unit-based procedural integrated display (Fig. 1), she defined the representations for different states of operation unit on the system display, as well as the operation history on the procedural display. In the procedure list, sub-steps could be expanded by clicking on the higher level step. The current step was shown in a yellow label and expanded on default, whereas the future steps were in gray labels and could not be expanded.

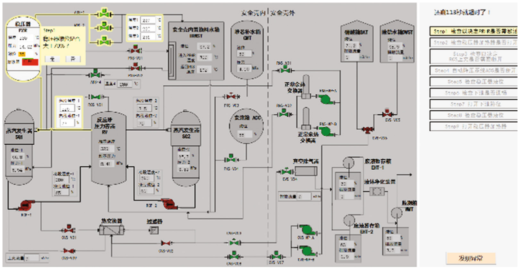


Fig. 1. Unit-based procedural integrated display (Chen et al. in press) (Color figure online)

3 Research Methods

3.1 Use Case Analysis

We carried out a use case analysis based on the information acquired from literature and in-field study at SNERDI. Operators carry out operation procedures through the following use cases:

- Judge, monitor, and operate in the order of the procedure;
- Jump out of the sequence to inspect another step and associated display;
- Operate on another open display of interest;
- Return to the current step after jumping out of the sequence;
- Review the action on each step and export the report;
- Go to another procedure and return at any time;
- Monitor a parameter, and on reaching a set point, jump out of the current sequence, and perform another step.

3.2 Design and Iterations

The first stage focused on optimizing the design proposed by Chen, and introducing dual navigation system to support task-based and system-based navigations in real settings. In stage 2, we adapted the design to the norms and constraints of the control room DCS platform. We further modified the page layout after static implementation and identifying problems through expert review in stage 3. Finally, we redesigned the dynamic properties through another round of expert review. The final design is as shown in Fig. 2.

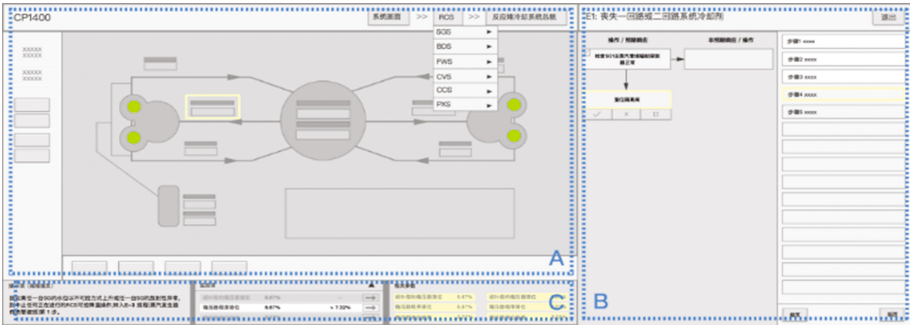


Fig. 2. Section A: System display; Section B: Procedural display; Section C: Reminder, monitor and relevant parameter area

3.3 Implementation

Based on the stage 4 design, the integrated display is developed using VBScript on the DCS platform.

4 Evaluation

4.1 Usability Testing

Introduction

We implemented the stage 4 design on a simulated NPP control room platform. We then conducted a usability testing, which was set up in a lab modeled after real NPP control room. The task was the emergency boronization procedure. This procedure involved moderate number of systems as well as complexity for small-scale experiments. We invited two operators from a related nuclear power station to participate in the usability test. Both of them were senior operators, with 17 years and 11 years of operational experience respectively.

Measurements

Performance Level. We used errors and inefficient behaviors as indicators for performance. Error referred to the deviation from the procedure or invalid operation; inefficient behavior referred to obvious hesitation or duplication. Operators' performance was studied using screen recorders and post-test interviews.

Task Workload. The standard NASA-TLX scale [17] was used to assess operators' workload.

Disorientation. This consisted of perceived disorientation [18] and extra navigation steps.

Perceived Ease of Learning. The procedural-integrated display changed the structure of the original HSI and adopted innovated interactions. Therefore, this study measured how well operators learn to use the design.

Situational Awareness. We used SAGAT method [19] to measure situational awareness. Two freezing points were determined by NPP experts.

Acceptance. We assessed acceptance using the technology acceptance model (TAM) proposed by Davis et al. [20].

Procedure. First, operators went through a 10 min training session and were given sufficient time to get familiar with the new design. We then carried out a 15 min interview and probed into their learning experience. During the testing, the operator performed emergency boronization procedure. Two SA freezes each took 2 min. After the experiment, operators were asked to fill out questionnaires for task workload, disorientation and acceptance. Finally, a 20 min interview probed into the problems found during the test and gathered overall attitudes.

Results

Performance. Screen behaviors (mouse tracking and clicks) and operators' think-aloud data was recorded for performance study. Each of the operator had seven errors/inefficiencies throughout the experiment (Table 1), yet each of them was recovered quickly and had not hindered the smooth completion of the task. The total time was in a reasonable range.

Table 1. Usability testing performance result

Indicators	Description-problems found
Inefficiency	Not recognizing the highlighted component in the picture and searching [P1, P2]
	Forgot to activate a sub-step and searched for components manually in the screen before realizing [P1]
	Not knowing where to look for a parameter (which was in the monitor section) [P1]
	Wanted to add a parameter to monitor section which turned out to be disabled [P1, P2]
Error	Advancing in the procedure without following the steps rigorously [P1]
	Failed to notice an important parameter before prompted [P1]
	Forgot to mark a step as *finished* [P2]
	Mistakenly judged the status of a component based on the *normal* color coding without noticing the *abnormal* parameters [P2]
	Attempted to operate on a component outside the highlighted area [P2]
Total time	27 m 58 s [P1] 25 min 40 s [P2]

Task Workload. Operators reported low level of frustration and time pressure during the task, and high satisfaction rate towards their performance. Mental demand was high for both operators and interviews found out that this was because they are trained to invest all their attention and memory resources when performing operation procedures (Fig. 3).

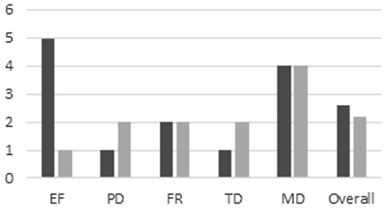


Fig. 3. Workload result

Table 2. Disorientation

Indicator	Score	
	P1	P2
I feel confused in the test	3	1
I feel circumventing through the test	1	1
I find it hard to return to a previous page	1	1
I find it hard to navigate in the system	1	1
I do not know which page to go after a series of actions	4	1
I feel lost and disoriented in the system	3	1
Weighted average	2.3	

Disorientation. The result showed that the two operators experienced low level of disorientation during the task, and the navigation design provided good support for operators (Table 2). During the interview, the two operators pointed out that they were well aware of the current position in the system even without the global navigation aid. Operators have higher orientation ability than we originally expected.

Perceived Ease of Learning. In the interview, both operator said that for qualified operators, the new design would be an assistant tool to the original system and should be very easy to learn and become adapted to.

Situational Awareness. The results of situational awareness measurements are shown in Table 3 and Fig. 4.

Table 3. Situation awareness result

SA Level	Questions	Score	
		P1	P2
L1	Q1	10	6.5
	Q2	10	7
	Q3	10	6
	Q4	10	5
L2	Q1	10	3.5
	Q2	10	7
L3	Q1	10	4
	Q2	10	3.5

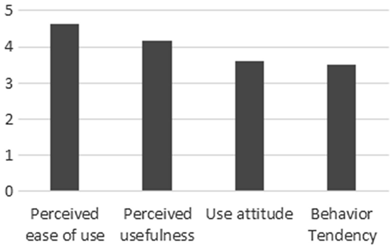


Fig. 4. Acceptance

Acceptance. Operators rating showed a high level of acceptance toward the new design. *Behavior Tendency* and *Use Attitude* had lower ratings, and we found out from the interview that the two operators were conservative towards the introduction of uncertainties. Both operators acknowledged the significance of the integration of system display and procedural display. Both operators thought that the new design was easy to use and effective in reducing errors.

4.2 Expert Review

Introduction. We invited experts from four different fields (Control Room Human Factors, NPP Process System, NPP Information and Technology, NPP Accident Analysis) for a review to identify problems and help improve the design. The experiment settings and operation procedure used for expert review were the same as in the usability testing. The review started with a 10 min introduction/training session. Experts then tried out the design for approximately 15 min. In this process, researchers observed and recorded experts' comments and think-aloud data. After that, experts were asked to take a questionnaire to measure disorientation, which was followed with an interview.

Results

Disorientation. Two of the four experts reported confusions. In one case, the slow reaction rate of the system led to confusion. In another case, the highlight did not stand out visually.

Design Problems.

- Color: the gray color should represent the status of completion while the white color as incompletion because white would invite more attention. Experts also suggested that the confirmation status include redundant color coding to provide more visual aids.
- Layout: the monitor area is too small and not distinctive compared to the Relevant Parameter area; the reminder texts should be displayed along with the procedure as they should be read side by side.
- Dynamic properties: experts suggested that the operation dialog box should also be able to get activated through the operation procedure.

Overall Evaluation of the Integration Screen from Expert Review. The integration and dynamic link between the procedure and the system display could improve efficiency for NPP operators.

5 Improvement Plan

Based on the result of usability testing and expert evaluation, this study came up with an improvement plan (Fig. 5).

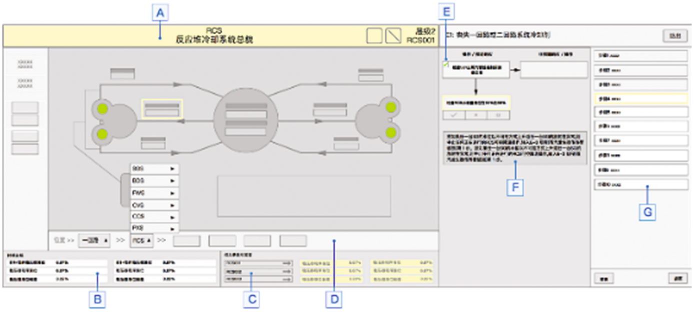


Fig. 5. Improvement plan

A	Title of the system display
B	Monitor area
C	Entrance to relevant system displays
D	Breadcrumb navigation bar at the bottom
E	Two-way coding of completion status. Single click on the step triggers inspection mode on the step, while double clicks lead to operation mode, in which action box and operation dialog will appear
F	Attention and reminder texts included in the operation procedures
G	Operation procedure steps, with uncompleted steps represented in white.

6 Conclusion

This study optimized the design of procedure-integrated display. Based on previous studies and actual requirements of NPP control room, this paper proposed a design that combined system-based and task-based navigation, allowing operators to locate information easier and reducing interface management tasks. After four rounds of iterations based on field study and expert reviews, the final design was realized on a simulated control room platform. This study invited two operators and four NPP experts for usability testing and expert review respectively. The result showed that this procedure-integrated display would be applicable to help operators improve performance and reduce mental workload, and lead to low disorientation level as well as high acceptance and learnability. This paper also discussed an improved design.

References

1. O’Hara, J., Brown, W.: Human-system interface management: human factors review guidance. Brookhaven National Laboratory, Upton, New York (2001)
2. Wickens, C.D.: Engineering Psychology and Human Performance. HarperCollins Publishers, New York (1992)

3. Strand, S., Svengren, H., Nihlwing, C., Kristiansen, L.I., Andresen, G., Meyer, B.: Task-based displays—prototype extensions and the second user test (HWR-841). OECD Halden Reactor Project, Halden (2007)
4. Xu, S., Song, F., Li, Z., et al.: An ergonomics study of computerized emergency operating procedures: presentation style, task complexity, and training level. Reliab. Eng. Syst. Saf. **93** (10), 1500–1511 (2008)
5. Hwang, F.-H., Hwang, S.-L.: Design and evaluation of computerized operating procedures in nuclear power plants. Ergonomics **46**(1–3), 271–284 (2003)
6. Yoshikawa, H.: Human-machine interaction in nuclear power plants. Nucl. Eng. Technol. **37** (2), 151 (2005)
7. Burns, C.M., Skraaning Jr., G., Jamieson, G.A., et al.: Evaluation of ecological interface design for nuclear process control: situation awareness effects. Hum. Factors J. Hum. Factors Ergon. Soc. **50**(4), 663–679 (2008)
8. Vicente, K.J., Christoffersen, K., Pereklita, A.: Supporting operator problem solving through ecological interface design. IEEE Trans. Syst. Man Cybern. **25**(4), 529–545 (1995)
9. Janzen, M.E., Vicente, K.J.: Attention allocation within the abstraction hierarchy. Int. J. Hum Comput Stud. **48**(4), 521–545 (1998)
10. Heslinga, G., Herbert, M.: Experience with the introduction of new system s for human-machine interaction from a human factors perspective (CORECH-HMI Summary Report). KEMA, Netherlands (1993)
11. Heslinga, G., Herbert, M.: Experiences with advanced systems for human-machine interaction. In: Proceedings of the 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems. MIT, Cambridge (1995)
12. Chalmers, P.A.: The role of cognitive theory in human-computer interface. Comput. Hum. Behav. **19**(5), 593–607 (2003)
13. McDonald, S., Stevenson, R.J.: Disorientation in hypertext: the effects of three text structures on navigation performance. Appl. Ergon. **27**(1), 61–68 (1996)
14. Allinson, L., Hammond, N.V.: A learning support environment: the Hitch-Hiker's guide. In: McAleese, R. (ed.), Hypertext: Theory into Practice, Intellect, Oxford (1989)
15. Edwards, D., Hardman, L.: Lost in Hyperspace': cognitive mapping and navigation in a hypertext environment. In: McAleese, R. (ed.) Hypertext: Theory into Practice, Intellect, Oxford, pp. 90–105 (1989)
16. Gupta, M., Gramopadhye, A.K.: An evaluation of different navigational tools in using hypertext. Comput. Ind. Eng. **29**, 437–441 (1995)
17. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
18. Ahuja, J.S., Webster, J.: Perceived disorientation: an examination of a new measure to assess web design effectiveness. Interact. Comput. **14**(1), 15–29 (2001)
19. Endsley, M.R.: Situation global assessment technique (SAGAT). In: Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, NAECON 1988, vol. 3, pp. 789–795. IEEE (1988)
20. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
21. Chen, Y., Gao, Q., Song, F., Li, Z., Wang, Y.F.: Procedure and information displays in advanced nuclear control rooms: experimental evaluation of an integrated design. Ergonomics (in press). doi:10.1080/00140139.2017.1288929

Design and Evaluation of an Abstract Auxiliary Display for Operating Procedures in Advanced NPP Control Rooms

Yahui Ma[1], Xiang Jiang[1], Qin Gao[1](✉), Haitao Lian[2], and Qiuyu Wang[2]

[1] Tsinghua University, Beijing, China
gaoqin@tsinghua.edu.cn
[2] Shanghai Nuclear Engineering Research and Design Institute, Shanghai, China

Abstract. Accident investigations have revealed that human errors are the most likely reasons of major accidents in the nuclear domain. As the nuclear power plant system become more computerized and automated, operators pay more attention to following the operating procedures, rather than why they are conducting a certain operation, or what results will the operation lead to. On the other hand, due to complexity of the nuclear system as well as accident scenario, operators are overwhelmed by detailed operating procedures and component parameters, thus may lose awareness of system status while conducting operations. This is very dangerous especially when emergency or unanticipated events occurs.

Researches show that displaying higher information abstraction hierarchy on the nuclear system interface can help operators achieve higher efficiency as well as situation awareness. This approach has been used in interface design in the nuclear domain, and has been proved to be helpful with operators understanding of the system, as well as the connection between component parameters and the system status. However, existing applications of abstraction hierarchy are system-oriented, and do not explicitly supply procedures operation.

This project is an exploratory research to improve the functional displays design for a 3rd generation nuclear power plant. Our objective in this research, is to design and evaluate an auxiliary display with higher abstraction hierarchy, for supplying emergency operating procedures in the emergency scenario. We use task analysis methods to obtain a list of system components and parameters that are related to the emergency operating procedures. Work-domain analysis is used to obtain goal-means connections among system components as well as their parameters. With this list of elements and their connections, we design an auxiliary display, referring to the current system interface design handbook. The display design is implemented on design-support DCS platform. User testing and expert evaluation are conducted to evaluate the display design in two aspects: improvement of operation efficiency, as well as reduction of human errors.

Keywords: Display design · Human error · Situation awareness · Abstraction hierarchy

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 319–328, 2017.
DOI: 10.1007/978-3-319-58472-0_25

1 Introduction

1.1 Research Background

As more and more emphasis is paid upon clean energy, nuclear energy is being used to meet the needs of more industries. However, from Three Mile Island to Fukushima nuclear power plant accident, the safety of nuclear power plants has been the bottleneck of the development and wide acceptance of nuclear energy. On the one hand, serious consequences of accidents could be caused in the field of nuclear power; on the other hand, there is a high proportion of human error among all errors. With the development of nuclear power technology and the improvement of system automation, human error has become the most likely cause of accidents in the field of nuclear power (Endsley and Garland 2000). This is very dangerous especially when emergency or unanticipated events occurs (Burns 2008). Thus, a lot of research has been done on how to improve operator performance. The digital human-machine interface is a very important part of the factors that affect operators' performance.

The digital human-machine interface includes the digitized monitoring display, the alert system and the procedure display. Our study focuses on the procedure display of the system. Procedure is what guides people to achieve a certain purpose by taking sequential actions (Niwa and Hollnagel 2002). In nuclear power plant control room, the operator follows these procedures to carry out corresponding actions. Specifically, we focus on emergency operation procedures.

1.2 Research Methods

The current procedure display includes three modules: the main step module containing procedure steps, the sub-step module containing specific operations corresponding to each main task, and the parameter module containing relevant parameters of the current system. The issue with this display is that only the detailed steps to be performed are displayed. Although operators have taken training regarding the reason to perform this step, it's hard to maintain overall situational awareness (Baddeley 1972) of the whole system, since operators are conducting operations with a high degree of concentration on the current step. Meanwhile, the problem related to situational awareness is the most likely cause to operators' errors (Endsley and Garland 2000). Due to the complexity of the emergency and the operation itself, the operator may be lost in the details of the specific task, and not able to maintain awareness of "why this task needs to be accomplished", "what is going to happen if the step is accomplished" and "what is going on in the whole system". The situation is particularly dangerous when the operator is conducting unfamiliar operation procedures.

Researches have shown that improving the level of abstraction hierarchy of information displayed on the screen can improve operator performance. This prompts us to add an abstract auxiliary display to the digitized system to improve the performance of the operator in emergency. We evaluate operators' performance from two aspects: first, the completion rate of procedures, including the completion time, error rate, workload, etc.; second, the situational awareness of the system.

On the other hand, most of the previous researches are focused on the information content and information organization of the monitoring display. In this study, we will apply these methods and design concepts to the procedure display. Since the purpose of improving the abstraction hierarchy of the information (Rasmussen 1985) in the monitoring display is to improve the operators' situation awareness of the system state, it is necessary to obtain the function information of each system components. While in the procedure display, we are concerned not only about the situation awareness of the whole system, but also of the current task being operated on. Thus, we need to extend the study to each system components and operation steps. After obtaining their functional information, we then come up with a reasonable way to organize it.

In this study, we first use the cognitive work analysis to get the information content to be presented on the display, and the second is to research the organization of information per different interface design methods. Then we implement the procedure display on the design-support DCS platform in the actual system. Experiment is then carried out to evaluate the display for its ability to support situation awareness as well as task completion.

2 Design of Procedure Display

2.1 Design Method

In "abstract auxiliary display" we concern about two aspects: "abstract" refers to the use of information abstraction hierarchy (Rasmussen 1985) in the display to show the system state-related information to help the operator improve the overall state of the system cognition; "auxiliary" refers to adding computerized procedures to support the display, to ensure that the operator can successfully complete the procedures in the display.

In terms of showing the system state, in the abstract auxiliary display each relevant component is displayed, their state being encoded by shapes and colors. Focusing on the emergency operation procedure we choose, we mainly concern on the coolant loading, pressure and temperature of the first loop. At the same time, we also show the functional connection between the components in the display.

In terms of supporting the procedure operation, the display shows various items on the paper-based procedure, including steps, sub-steps, prompts and monitoring items. And by highlighting the information associated with the current task in the auxiliary screen, we combine the procedure steps and the auxiliary screen to improve the operators' situation awareness of the current operational task.

2.2 Display Design

Our design called the connection display is shown in Fig. 1. In this display, we keep the original physical connections among components of the system, components' parameters and states are encoded by shape, area and color. The main coding methods and their corresponding meanings are described as follows:

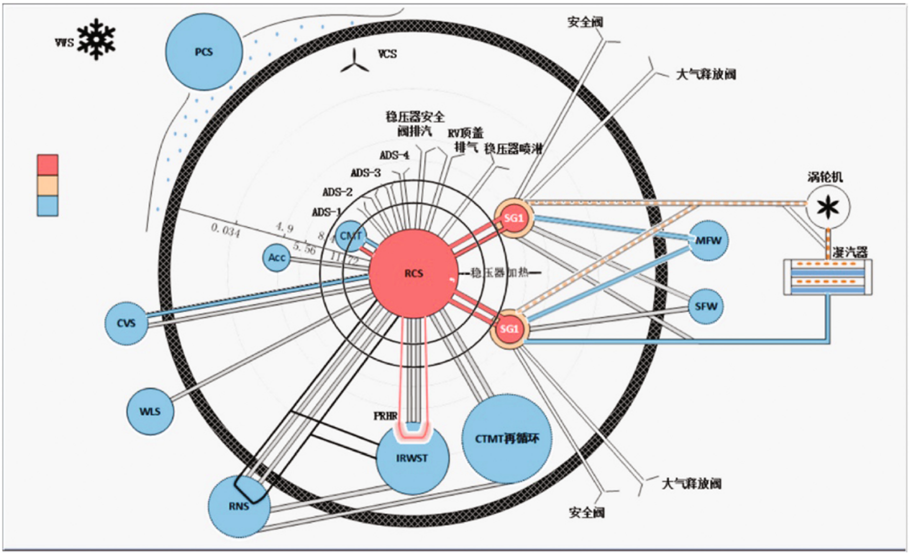


Fig. 1. Connection display design (Color figure online)

- Circles, pipes and loading

Storage and transmission of the coolant is the main support of loading, represented by circles and pipes between circles. Specifically, if there is only one-way coolant transfer between the two devices, the two circles are connected by only one pipe; if there is a two-way coolant circulation between the two devices, the two circles are by two pipes. As shown in the figure, it is a one-way transfer from ACC to RCS, and it is a two-way circulation between CVS and RCS.

In addition, we use the area of circles combined with solid/dashed contour to represent the volume of each equipment. A solid contour is used when the equipment (such as the ACC) carries the set amount of load which doesn't change with condition. And the device that can introduce more liquid from the outside of the system is represented by a dashed contour. For circles with solid contour, we use the area size of the circle to represent the size of its inherent capacity. For example, in PXS system, IRWST has the largest amount of loading, followed by CMT, and ACC has the minimum, then circle representing IRWST has the maximum area size, followed by CMT and finally ACC. Note that the size of area and the amount of the loading are only corresponded in order, not in proportion.

- Black solid lines and temperature

The direct operation on the temperature is mainly in the form of heat exchange in pipes, represented by black solid lines in our display. For example, PRHR uses coolant in IRWST to cool down the high-temperature coolant in the pipe from RCS, through the pipe heat exchange.

- Pipes with outlet and pressure

The RCS system primarily reduces the pressure by venting gas (or gas-liquid mixture), such as in ADS stages 1–3, which depressurize the RCS system by excluding high-pressure gas from the RCS system into the containment vessel. We represent this with the pipes with outlet.

Since the procedure is a cooling and depressurization process of the RCS system, we use color and concentric circles to represent the temperature and pressure in our display. And the combination of color and pipe represents whether the device is running or not.

- Colors and temperature

Three colors are used to represent the system temperature: red represents the highest temperature, which characterizes the temperature of the RCS system and the primary coolant of the SG; yellow represents the second-high temperature which characterizes that of the SG secondary side liquid, gas, and RCS gas outlet; the lowest temperature is represented by blue, as shown in the figure, such as CVS, CMT.

- Colors and equipment status

The color can be used combined with a pipe to represent the operation status of a device: for example, when a coolant is being transferred between two circles, the pipe is filled with solid red or blue. Red indicates that the coolant in the RCS flows into the other device, and blue indicates the flow from the other devices into the RCS. If the ADS first-stage is running, the ADS first-stage pipeline is filled with yellow dashed line. For heat exchange pipes, they are filled with solid black when not running, and filled with read surrounded by a halo when they are running.

- Concentric circles and the size of the pressure

From the work analysis of the system, most of the automatic cooling and pressure relief operations are related to the RCS system pressure. In the display, we use concentric circles to represent RCS system pressure. There are some representative pressure values, such as when the RCS system pressure is lower than 11.72 MPa, CMT and PRHR are in use, and when it's less than 4.9 MPa the ACC is triggered to come into use. As shown in the figure, from inside to outside, concentric circles form a pressure drop scale ring. To show the correspondence between the pressure and the devices being triggered by each pressure value, devices are placed on the corresponding concentric circles. For example, when the RCS pressure drops to a specific value, the corresponding concentric circle lights up (turns black) and the corresponding device starts to work (pipe filled).

3 Design Valuation

3.1 Operating Environment

Since the design of this abstract auxiliary display is engineering-oriented, it is necessary to extend the design scheme to all procedures. The focus of attention is on

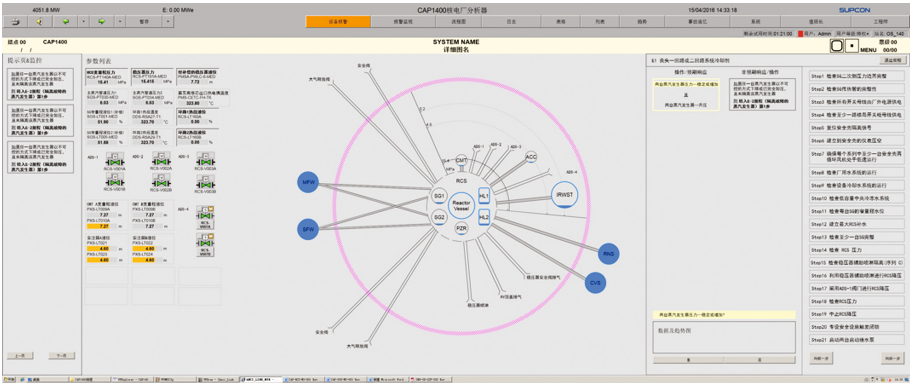


Fig. 2. Design implementation on system display

practical application in nuclear power plants. Therefore, the compatibility between the design and the original nuclear power plant system, and the operating experience are very important. Considering the actual situation of the laboratory, we combined user test and expert evaluation to evaluate the display design.

- **Procedure display.** In this study, the following two procedure displays are involved:
  - **Original procedure display:** the display that includes only the main step module, the sub-step module and the parameter module, not including the abstract auxiliary display;
  - **Procedure display with abstract auxiliary display:** the display that includes the main step module, the sub-step module and the parameter module, together with the abstract auxiliary display (Fig. 2).
- **Operational tasks.** As previous researched have shown, the structural goal-instrument information has a significant effect on performance improvement when the operator performs a non-designed task, but little help in designed tasks and the situation with interference. To solve the problem in these two cases, we use highlighting relevant information to connect procedure display and abstract display to provide operational purposes-means information. To distinguish these three cases, the following three tasks are designed:
  - **Designed benchmark task:** to conduct the emergency operation procedures;
  - **Designed benchmark task with non-design tasks**: When conducting procedures, abnormal situations happen in the system, which is about parameters relevant to the designed benchmark task, but doesn't affect the operation and final system status. For example, CMT level value is normal, but is shown abnormal and in alarming color in the abstract. If the operator raised this question during the operation or interviews, then consider the operator's non-design task completed;

- **Designed benchmark task with interference:** the interference occurs when operator is conducting procedure steps, whereby the performance of the operator after the return to the task is evaluated. Situational awareness questionnaires, which are needed to obtain data during task operation, serve as interference in our experiment.

3.2 Testing Process

In the user test section, a total of two nuclear power plant operators participated in the test. The two operators respectively go through the following processes: training, completing the learning experience questionnaire, conducting emergency operation procedure tasks (filling in the situational awareness questionnaires in the process), filling in the operation experience questionnaire and post-test interview.

In the procedure operation section, the operator conduct operations alone in the simulated control room, using the abstract auxiliary display, and the accident process and data are controlled from the simulator by staff. During operation, to ensure that the operation process is close to the actual situation of the main control room as much as possible, staff and expert reviewers observe the operator's screen and mouse operation in the observation room through the Morae recording and monitoring tools on another screen.

3.3 Operator Performance Measures

In the user test section, on the one hand we obtain subjective opinions feedbacks through the interview, on the other hand we also collect the following quantitative data (Kim and Seong 2009).

- **Task completion time:** the whole time the operator needed to complete the operation from the beginning to mandate procedures.
- **Situational awareness:** it is divided into situation awareness of system status and of procedure operations (Burns 2008), the two of them are also divided into overall perception of the process, and understanding of the current situation respectively. The overall perception is measured by the questionnaire sent out after the task is finished, at which time the participant cannot see the system display and need to fill out the questionnaire relying on their own recall. The question in the questionnaire is about the whole operation process, the change trend of local or global parameters. The understanding of the current situation involves questions like "why you are conducting this operation", "what do you expect to see when this operation is completed" and "why is this situation happening". A total of four situation awareness questionnaires were sent during the test. The first time was after the operator chose to start the operation, and the situation awareness of the system state was examined. The second and third were completed respectively by the operator after the operation of two modules of operations, and cognition of the current operation was examined. The fourth is sent when the operation is completed, once again on the cognition of system status.

- **Workload:** we use the NASA-TLX scale to examine workload, including six dimensions of mental demand, physical demand, time pressure, effort requirement, performance satisfaction and frustration, each dimension scoring 1–10, with higher scores indicating greater workload.
- **Non-design task completion:** CMT level value is normal during the test, but is shown abnormal and in alarming color in the abstract. If the operator raised this question during the operation or interviews, then consider the operator's non-design task completed.

3.4 Data Analysis

- **Task completion time.** The two operators spent 17 min and 21 min respectively completing 23 steps of the procedure. Since there is no comparison between operation with and without the abstract auxiliary display, we asked operators to evaluate the task completion time during the post-test interview. Both operators reckon that, due to the screen integration and automatic navigation, they do not need to find their relevant displays and parameters on their own, thus the use of the abstract auxiliary screen can shorten task completion time.
- **Situational awareness.** In the case of questionnaires on situational awareness, the two operators showed relatively significant differences. The first operator received a high score in four questionnaires, and the four experts who participated in the review considered the operator's response to be correct. In post-test interviews, the operator said he was familiar with the process of the accident, and can recall the accident process and direction after a brief observation of the abstract display. The second operator received a low score on the situational awareness question in the system state, especially the state of CMT component is poorly understood. In post-test interviews, the operator said he was not familiar with the process of the accident. As we can see, the abstract auxiliary display has a better effect when operators are more familiar with the situation.
- **Workload.** Both operators selected moderate workloads in the workload scales.
- **Non-design task completion.** Both operators raised question regarding CMT loading alarm in interview.

3.5 Operators' Feedback

In terms of usefulness, both operators considered the current abstract auxiliary display to include all the components, parameters, and procedural operations required for the emergency operation. And system abnormal is easy to notice using this display.

In terms of usability, the operators offered some improvement suggestions.

First, the connection relationship of internal RCS system components is better to be displayed intuitively.

Second, we can combine the auxiliary display with the original display where detailed system components and all parameters are shown, since the auxiliary works better in obtaining situation status in a short time, but more detailed information need to be provided using original system displays.

Finally, in terms of degree of computerization, it is desirable that the system can recommend the next operation in the procedure, eliminating the need for the operator's comparison and judgment steps so that the operator only needs to finalize and confirm the recommended operation from the system.

4 Results and Discussion

4.1 Results and Suggestions

From the operators' performance and interview we can that the current abstract auxiliary display design can meet the basic operation of the chosen emergency operation procedures. The future direction of improvement is to achieve the navigation connection between abstract auxiliary display and the original display, and involve all the relevant parameters from the original system flow diagram display in the abstract auxiliary display.

After the evaluation, it was found that the operators and the experts basically recognized the function of the success in procedure operation support of the abstract auxiliary display, and hoped to use the display as part of the comprehensive function display to help the operator complete the procedure operation and understand the system status.

The abstract auxiliary display can be used as a part of the comprehensive function display, facilitating the operator's procedural operations. It can be applied in two ways.

- One is to reduce the display size and embed it into the original system flow diagram display, to help the operator detect the overall status of the system. To make this happen, we need to adjust the current size of the corresponding parts in the display to ensure that each part and their status are still recognizable after reducing the display size.
- The second is to involve all the relevant parameters from the original system flow diagram display in the abstract auxiliary display, which can ensure that the operator can rely on this one screen to complete all the procedures operation. Considering the large number of parameters, we will need to add a new layer, and the operator can freely choose which part of the display parameters to be shown.

4.2 Limitations and Future Works

First, due to constraints of time, equipment and personnel, we were not able to carry out sufficient usability testing after completing the design and implementation of the abstract auxiliary display. Thus, the design also requires follow-up engineering experiments and further improvements before being introduced in the nuclear power plant system.

Second, when applying this design to the whole system beyond the chosen emergency operation procedure, the method of our study can be used, which is, work analysis based on work domain, display design using original connections and various coding methods, and implementation on the design-support DCS system.

Finally, to fully prove the help of the abstract auxiliary display to the operator and the integration with the original display, we need to carry on the engineering experiment after systematic application of the picture, and obtain the feedback from the operators and make the next iteration.

Acknowledgment. This study was supported by the National Natural Science Foundation of China (Project No. 71671102).

References

Baddeley, A.D.: Selective attention and performance in dangerous environments. Br. J. Psychol. **63**(4), 537–546 (1972)

Burns, C.M., Skraaning, G., Jamieson, G.A., Lau, N., Kwok, J., Welch, R., Andresen, G.: Evaluation of ecological interface design for nuclear process control: situation awareness effects. Hum. Factors J. Hum. Factors Ergon. Soc. **50**(4), 663–679 (2008)

Endsley, M.R., Garland, D.J. (eds.): Situation Awareness Analysis and Measurement. CRC Press, Boca Raton (2000)

Kim, J.H., Seong, P.H.: Human Factors Engineering in Large-Scale Digital Control Systems, pp. 163–195. Springer, London (2009)

Niwa, Y., Hollnagel, E.: Integrated computerisation of operating procedures. Nucl. Eng. Des. **213** (2), 289–301 (2002)

Rasmussen, J.: The role of hierarchical knowledge representation in decision making and system management. IEEE Trans. Syst. Man Cybern. **15**(2), 234–243 (1985)

Authority Pathway: Intelligent Adaptive Automation for a UAS Ground Control Station

Derek McColl[1](✉), Kevin Heffner[2], Simon Banbury[3],
Mario Charron[4], Robert Arrabito[1], and Ming Hou[1]

[1] Defence Research and Development Canada - Toronto Research Centre,
Toronto, Canada
{derek.mccoll, robert.arrabito,
ming.hou}@drdc-rddc.gc.ca
[2] Pegasus Research and Technologies, Montreal, Canada
k.heffner@peretec.com
[3] C3 Human Factors Consulting Inc., Montreal, Canada
simon.banbury@c3hf.com
[4] Department of National Defence, Ottawa, Canada
mario.charron@forces.gc.ca

Abstract. Defence Research and Development Canada has conducted a number of human factors analysis tasks and experiments to support the Canadian Armed Force's acquisition of an unmanned aircraft system for domestic and international operations. Experiments were run on a simulation-based UAS mission experimentation testbed. Results promoted the design and development of an intelligent adaptive interface called Authority Pathway that assists crew members with maintaining situational awareness and to adhere to rules of engagement when targeting. The Authority Pathway uses intelligent adaptive automation technology to adapt to dynamically changing mission goals, provide a variety of views for different users, and allow for varying levels of automation to be consistent with future requirements. Advantages and disadvantages of the Authority Pathway are discussed.

Keywords: Unmanned Aerial Vehicle · Unmanned Aircraft System · Ground Control Station · Human machine interface · Intelligent Adaptive Interface · Intelligent Adaptive Automation · Systems requirements

1 Introduction

Unmanned Aerial Vehicles (UAVs) encompass a wide range of vehicles from hand held consumer toys like the AR. Drone by Parrot [1] to large military vehicles like the Global Hawk which has a 35 m wing span [2]. A military Unmanned Aircraft System (UAS) includes the UAV; the Ground Control Station (GCS); dedicated crews for maintenance, launch and recovery; and a crew to operate the UAV during mission execution. Military UAS are typically outfitted with a variety of sensors and used for intelligence, surveillance and reconnaissance (ISR) missions, with some UAV having targeted strike capabilities [3]. Military UAS are designed to allow for ISR missions in hostile environments for extended periods of time without putting flight crews at risk.

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 329–342, 2017.
DOI: 10.1007/978-3-319-58472-0_26

Due to the nature of UAS having its crew in a separate location from the aircraft, it is necessary that automation be employed to ensure safe control of the aircraft and its systems. It is critical that the human-automation interaction between a UAS and its crew be effective, this is especially true for military systems that are used on the battlefield where lives are at stake [4]. To promote mission effectiveness, it is important that the human crew have trust in their autonomous systems [4]. This trust comes from machine reliability and an effective Human-Machine Interface (HMI) that can explain and justify the automation's recommendations and actions [4].

The Canadian Armed Force's Joint Unmanned Surveillance and Target Acquisition System (JUSTAS) program has the goal to procure a long endurance UAS for domestic and international operations [5]. As a stakeholder, JUSTAS has mandated DRDC to identify critical UAS GCS HMI information requirements and provide recommendations on GCS crew resource management. DRDC has additional stakeholder requirements to identify UAS crew responsibilities and training requirements, develop preliminary training technologies and material, and provide guidance on GCS airworthiness certification. The GCS of the JUSTAS UAS will follow the Royal Canadian Air Force (RCAF) squadron level UAS control concept, where each UAS mission crew includes: a pilot or Air Vehicle Operator (AVO); airborne sensor operators referred to as Payload Operators (PO); and intelligence analysts for Imagery (IMA) and Electronic Warfare (EW).

Defence Research and Development Canada (DRDC) has performed a three phase research project to meet these stakeholder requirements [6]. Phase 1, called stakeholder analysis, included the development of a composite mission based on the JUSTAS Concept of Operations (CONOPS) [7], UAS operator characteristics analysis and a preliminary Human Factors (HF) engineering system analysis to identify equipment and function allocation following NATO STANAG 3994 [8].

Phase 2 had multiple tasks to analyze UAS GCS functional requirements and training needs. The first task was to develop a functional requirements document that lists all UAS operator functions to accomplish the missions identified in the JUSTAS CONOPS. Another main task for this phase was a Cognitive Task Analysis (CTA) [9] to identify crew information requirements and decisions. The results of the CTA were used as the basis of a task flow analysis to develop task network models for operator performance modelling and data analysis for UAS GCS design and training requirements. Additionally HMI requirements and initial GCS user interface concepts were also derived from the CTA results. A training needs analysis was also performed to identify the knowledge, skills and abilities required for each of the UAS crew members.

DRDC is currently in Phase 3, which is focused on the development and evaluation of a simulation-based UAS mission experimentation testbed called Testbed for Integrated GCS Experimentation and Rehearsal (TIGER). Experimentation on TIGER will address stakeholder requirements on crew configuration, GCS functional requirements, workspace layouts, airworthiness certification and training requirements. TIGER consists of six crew workstations, three simulation control workstations and two experimenter workstations [6]. The six crew workstations include positions for an AVO, PO, two IMA, and two EW.

To date, preliminary trials have been conducted to assess TIGER's baseline capabilities and crew performance with six crew members [10]. Results of the trial indicated that multiple crew members often lost situational awareness (SA) due to unfamiliarity with the system and did not consider following Rules of Engagement (ROE) and Laws of Armed Conflict (LOAC) while engaging targets. Participant feedback after the trials also indicated that an automated decision aid could be helpful for improving crew adherence to ROE and LOAC. This led to the presentation of the "Authority Pathway" concept [10], an Intelligent Adaptive Interface (IAI) that assists UAS crews in following ROE and LOAC when engaging targets. An IAI is an operator interface that dynamically changes the display and/or control characteristics of human machine systems to adaptively react to the operator states and external events in real time [11]. The Authority Pathway, being developed on TIGER to address JUSTAS requirements for GCS HMI recommendations, is an IAI because it dynamically changes its HMI on TIGER in real-time in response to changes in the mission and environment as well as when it receives inputs from the UAS crew or Tasking Authority (TA).

2 Authority Pathway

The Authority Pathway was developed to support UAS crew HF studies investigating the use of Intelligent Adaptive Automation (IAA) technology [12] to assist the crew in maintaining awareness of the steps and permissions needed to conduct a lawful and successful target engagement [10]. An IAA technology is necessary for functional integration, rather than functional separation [12]. IAA technologies are adaptive because system control is shared dynamically by both the operator and automation based on their availabilities, capabilities, and limitations. IAA technologies are intelligent because they seek to restore the operator to the role of decision maker and provide safeguards for situations where time constraints or problem complexity restrict the operators problem-solving ability [12]. IAA systems proactively collect information, are goal driven, are capable of reasoning at multiple levels, and can learn from experience [12]. The goal of IAA is to enhance operator judgement, decision making, and responsibility while mitigating operator limitations [12]. The Authority Pathway uses IAA technology to assist the UAS crew in making critical target engagement decisions, while maintaining mission SA, and following proper ROE and LOAC, especially in time critical and complex situations. The Authority Pathway is agile to dynamically changing mission goals, has a variety of HMIs for different users, and allows for varying levels of automation to be consistent with the procedures and time-critical operational context.

The Authority Pathway concept is based on the steps comprising the targeting process depicted in Fig. 1. The process essentially commences once a Person of Interest (POI) has been detected, localized and retained as a potential target. The subsequent steps involve: completing a positive identification (PID) checklist; notifying the Tasking Authority (TA) and requesting authorization for the use of a kinetic response; performing a Collateral Damage Estimate (CDE) while simultaneously completing the weapon engagement planning; notifying the TA of the UAS readiness to engage and requesting weapon release authorization; lazing the target (as required); and releasing the weapon. Maintaining continuous eyes on target is a requisite parallel activity that must occur throughout the target engagement process. Battle Damage Assessment (BDA), not shown in the figure, is performed subsequent to the weapon release.

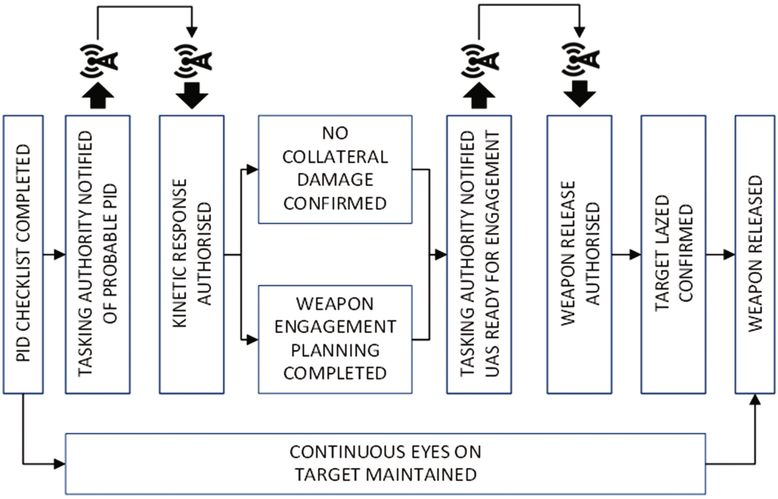


Fig. 1. Target engagement process overview.

During a sortie, the UAS may take on different roles. For instance, in the case of a call for direct fire, the UAS may be responsible for the targeting and weapon release whereas in the case of a call for indirect fire, the UAS crew may only be providing a target lazing capability for another manned or unmanned asset.

The Authority Pathway seeks to address two critical functional areas of the target engagement process: (1) *maintaining a shared mission SA* among all UAS crew members; and (2) *ensuring that all applicable, relevant ROE, LOAC are considered* in a timely manner and that the *standard operating procedures are followed*. One of the underlying premises of the Authority Pathway concept is that the judicious use of IAA applied to the GCS HMI will facilitate the individual and collaborative decision-making during the target engagement process.

Since the work reported on in [10], recent analysis and design efforts for the Authority Pathway have highlighted a number of challenges:

- the system must adapt to the mission type and UAS role, both of which may vary over the course of the sortie;
- the system must provide user-specific views (e.g. AVO, PO, IMA, role player, experiment observer); and
- the system must allow for the evolution of the level of automation consistent with future UAS requirements for additional semi-automated crew functions related to target engagement.

The following subsections address each of these challenges and describe these challenges in terms of requirements. The last Subsection 2.4, presents the advantages and disadvantages of the Authority Pathway application.

2.1 Authority Pathway Mission Agility

> "*C2 Agility is the capability of [a system] to successfully effect, cope with, and/or exploit changes in circumstances. […] Agility enables entities to effectively and efficiently employ the resources they have in a timely manner.*" [13].

The Authority Pathway will operate in an environment in which operational and emulated command and control (C2) systems have been integrated. It will support a variety of missions that will result in a specific instantiation based on mission templates. The fog of war, the unpredictability of military operations and the dynamic nature of UAS missions are all reasons why these missions are subject to change, sometimes referred to as *dynamic re-tasking*, i.e. a change in mission during a sortie. During dynamic re-tasking, the system is required to reconfigure itself to meet the evolving mission requirements.

The so-called *agility enablers* are: responsiveness, versatility, flexibility, resilience, adaptiveness and innovativeness [13]. The Authority Pathway has implicit C2 agility requirements since the target engagement requires many of these enablers. In particular, the responsiveness and adaptiveness of the system will impact the efficiency with which requests to the TA can be communicated and processed. Limitations of current systems relying too heavily on radio communications and chat for UAV operations have been documented [14]. For example, these enablers are important in the case when a Troops-In-Contact (TIC) mission preempts an ongoing Pattern-Of-Life (POL) mission. The system must be adaptive and react to this change by presenting a modified instance of the Authority Pathway that reflects the relevant set of applicable ROE and is consistent with the procedures and time-critical operational context. The Authority pathway could automatically detect this change in mission by interpreting UAS system data (e.g., C2 systems, sensor) or the Authority Pathway ROE could be changed manually by the UAS crew. These changes in ROE would result in a new or modified Authority Pathway display to inform the crew of the changes.

The need for C2 agility of the Authority Pathway is further compounded by the requirement for UAS collaboration scenarios wherein one UAS asset, for instance may provide a target lazing capability for another manned or unmanned asset that is providing a weapons fire. Collaborations among assets, whose crews are not co-located, present specific coordination and communication challenges related to maintaining mission SA while executing critical decision-making. For example, timely communication and coordination are vital to ensuring the quick response times required in the case of time-sensitive targeting (TST), in both deliberate and dynamic targeting scenarios while considering CDE, ROE, airspace and other restrictions during the targeting process [15].

2.2 Authority Pathway User Views

The Authority Pathway will be accessed by crew members, the TA, and experimenters. Different users require different functionality from their Authority Pathway views. Each *view* is a separate client application or configuration.

2.2.1 Shared SA View

This view accepts no user input. It displays the current status of the target engagement activity in the form of a graphic that indicates completed Authority Pathway steps in green and uncompleted steps in orange (see Fig. 2). The Shared SA View can be integrated into other applications, such as an overall mission SA display, as shown in Fig. 3.

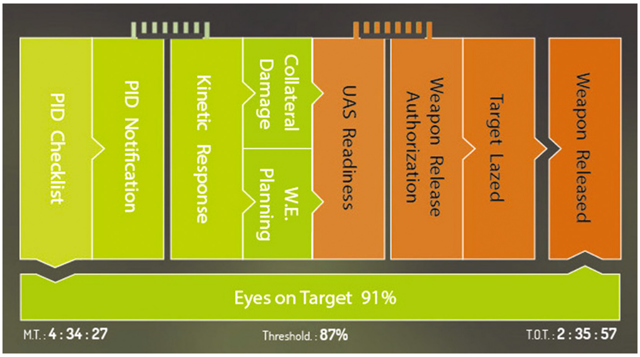


Fig. 2. Example Authority Pathway shared SA view (Color figure online)


OPERATION INTERCEPT
ID 24366
Current Time
13 : 05 : 27 zulu
Elapsed Mission Time
04 : 32 : 45

February 5, 2016
Com Auth Callsign
Gretsky
Cont Auth Callsign
Street Gang

Fig. 3. Authority Pathway shared SA view integrated into an example overall mission SA display which also includes UAS sensor feed, maps, and other important mission information.

2.2.2 UAS Crew View

This view is intended for use by the UAS crew members not performing the AVO or mission commander roles (e.g. PO/IMA/EW). In addition to the Shared SA View, this view provides information panes with a summary of the various steps (see Fig. 4). The information panes also allow a crew member to request actions from other users or set reminder alarms. When clicking on a specific information item, this view offers a drill-down capability to access increasing levels of detail. The information panes are related to another important feature of the Authority Pathway; checklists that crew members must complete and submit as part of the authorization request process.

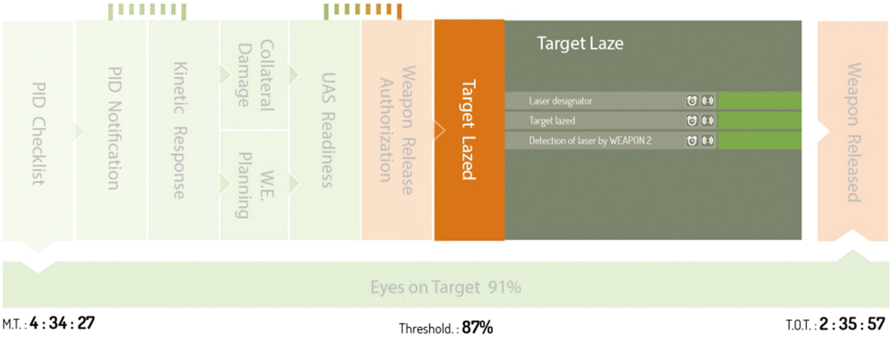


Fig. 4. Authority Pathway information pane

2.2.3 UAS Crew Commander View

This view is intended for the AVO, generally acting as crew commander, i.e. the person responsible for the mission and weapon engagements. In addition to the functionality of the UAS crew view, this view allows the crew commander to validate checklists and submit kinetic response and weapon release authorization requests to the TA.

2.2.4 Tasking Authority/White Cell View

During experiments, this view is for use by a role player acting as the TA. In addition to the read-only access to the Authority Pathway status and information panes, this view allows the user to grant or deny authorizations from the AVO. This view allows the user to review a checklist and to flag items that are incomplete, missing, or otherwise inadequate. The user can reject a request and provide instructions for rectifying it for subsequent resubmission.

2.2.5 Experimenter View

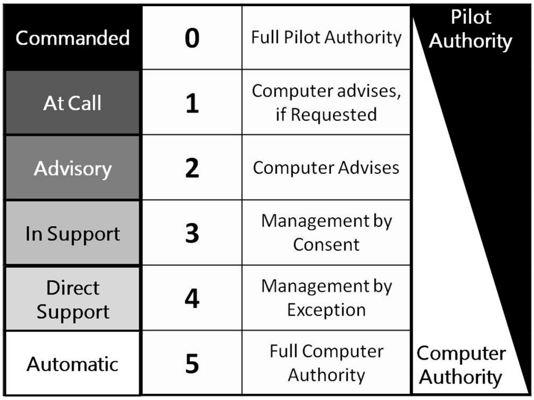
The Authority Pathway will have a record and playback capability for after-action review (AAR). Since the current focus of the Authority Pathway application is to support experiments, AAR will include observations from experimenters. The experimenter view allows the user to generate events and comments for subsequent review as part of the AAR. For example, the user can identify specific instances during which *Eyes on Target* were lost and then regained. Future experimenter view functions may include the capability to record observations concerning the operator's state. Future plans for the TIGER platform include psycho-physiological operator state monitoring capabilities that will provide data to the Authority Pathway during the experiment for the purposes of AAR and other post-experiment analyses.

2.3 Target Engagement Crew Function Automation

In the context of the Authority Pathway concept development, one of the underlying goals of integrating IAA aides within the GCS is to automate certain aspects of crew functions and thereby reduce the operator workload so that the crew can better focus on tasks that cannot or should not be automated.

The introduction of automation technology into complex environments such as UAS operations requires consideration of the risks and implications associated with a transfer of control from a human operator to a machine. The Pilot Authority and Control of Tasks (PACT) framework was initially developed to investigate the role of automation in future manned military aircraft [16]. Although there are many taxonomies to describe levels of automation when discussing human-automation interactions [12], PACT is used in this case because it gives a simple and clean description of the level of automation for Authority Pathway at a conceptual level. Applying these levels of automation (LOAs) is not always practical for systems design however it provides a conceptual understanding of the human-automation interaction goals of the system. PACT suggests that there are three basic levels of automation: *commanded (full pilot authority)*; *assisted (shared pilot/computer authority)*; and *automatic (full computer authority)*. Although fully automated control of UAV platforms currently is not a viable option, the PACT framework descriptions of assisted levels of authority are useful for the design of semi-automated UAS crew functions.

Figure 5 represents the six PACT levels of human/machine control in terms of human-machine responsibility sharing. The third column indicates a level of authority that includes the notion of automation management strategies [16]. Levels 0 and 5 are



Commanded	0	Full Pilot Authority	Pilot Authority
At Call	1	Computer advises, if Requested	
Advisory	2	Computer Advises	
In Support	3	Management by Consent	
Direct Support	4	Management by Exception	
Automatic	5	Full Computer Authority	Computer Authority

Fig. 5. PACT levels of autonomy Adapted from [16]

reference levels where the human and the pilot have complete control, respectively. Levels 1 through 4 represent shared human machine control with increasing machine control. With respect to the automation of crew functions, some functions are obvious candidates for automation while the automation of other functions may be contrary to doctrine and/or legal considerations. For example, certain aspects of the CDE activity can be greatly facilitated through the use of agent-based and other artificial intelligence technologies. Automatically signaling potential fratricide or civilian casualties and other relevant information to the crew early on in the target engagement process could facilitate crew decision-making, leading to remedial measures and alternate courses of action. To illustrate this point Fig. 6 depicts the cognitive process for determining if a target location is free of neutrals or friendly forces in the form of a decision ladder identified in the CTA analysis. The left side of the decision ladder involves primarily an analysis of the situation by the decision maker. The top of the ladder requires an interpretation or value judgment leading to a task, while the right side of the ladder is related to the planning and execution of the task. The CTA decision ladders were a key tool for the UAS target engagement task analyses and were used to identify areas where IAS technologies could be applied.

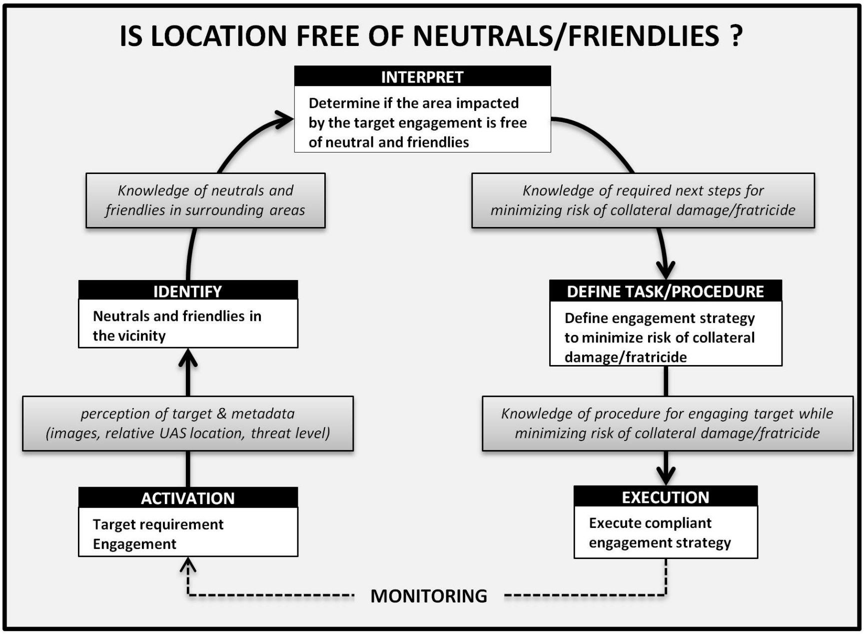


Fig. 6. Example UAS target engagement decision ladder

The decision ladders identified as pertinent to the Authority Pathway included two possible types of automation requirements: (1) IAA requirements; and (2) IAI requirements. IAA requirements are related to behind the scenes calculations and process performed *on behalf of* (or instead of) a crew member. IAI requirements involve visual and aural cues, checklists and indications that consolidate or accentuate information *directed to* the crew member. The two main areas addressed by the Authority Pathway, i.e. sharing SA and ensuring that procedures are followed, fit clearly in left and right sides of the decision ladder, respectively. For the ten steps comprising the target engagement process shown in Fig. 1, nineteen decision ladders were identified from the CTA. Table 1 presents a set of automation requirements associated with the decision ladder shown in Fig. 6. For each of the five tasks shown, two requirements have been identified. Features A, B, C and D involve the production of cues directed at the operator geared at increasing operator situation awareness concerning the existence of neutrals of friendlies in the vicinity of a target location. Feature C is of the type IAA and involves communication with other systems and automated processing in order to support feature D, the map display. These features are consistent with PACT autonomy level 2: the system is configured to inform the operator when specific conditions are met or triggers occur. The human operator is still responsible for determining if the location is free of neutrals or friendlies. Features E and F support the principal human decision-making task for this activity and also consistent with PACT autonomy level 2.

Table 1. Automation requirements from decision ladder: *Location free from friendlies/neutrals*

Task	Feature	Type
1. Target requiring engagement	A. Visual and/or audible alert. Expand window panel size	IAI
	B. EO/IR images, situation map showing UAS position and EO/IR metadata	IAI
2. Identify neutrals and friendlies	C. Identification of neutrals and friendlies within vicinity of engagement	IAA
	D. Map displaying target location and neutral/friendlies in surrounding areas	IAI
3. Determine if the area impacted by the target engagement is free of neutral and friendlies	E. Assessment of target engagement area for collateral damage impact	IAA
	F. Display area impacted by target engagement and indicated potential collateral damage/fratricide	IAI
4. Define engagement strategy to minimize risk of collateral damage/fratricide	G. Advice for minimizing risk of collateral damage/fratricide	IAA
	H. Checklist for engaging target	IAI
5. Execute compliant engagement strategy	I. Automatic execution of compliant engagement strategy	IAA
	J. Confirmation of setting changes	IAI

Feature G involves the system calculating potential strategies for subsequent actions and therefore is consistent with level 3 control. Features H and I collectively call for displaying a checklist for engaging the target and an automatic execution of compliant engagement strategy which also is consistent with level 3 – the system proposes an action that must be authorized by the operator. In the case of any changes in settings, feature J calls for confirmation by the operator, also consistent with level 3 control.

The set of automation requirements associated with the nineteen decision ladders is one of the key inputs into the Authority Pathway analysis and design. The following sections describe some of the primary design principles that enable the automation of crew functions: (1) procedures and ROE checklists; (2) authorization requests; and (3) status displays.

2.3.1 Procedure and ROE Checklists

Target engagement procedures can vary as a function of the operational context, the type of mission and the force structure. The steps involved in a TST or TIC situation are not the same as those required for engaging a target during a POL mission. Also, it can be time-consuming for personnel to identify all of the applicable and *relevant* ROE for a given situation. The Authority Pathway address these concerns by being adaptive to the specific situation and generating appropriate checklists to guide the target engagement process, including facilitating the assessment of *relevant* ROE and other items related to the LOAC.

The Authority Pathway includes mechanisms for: (1) completing a checklist among the UAS crew; (2) having the crew commander confirm that the checklist is completed; and (3) having the crew commander communicate that a checklist is incomplete with indications as to how to complete it.

2.3.2 Semi-automated Authorization Requests

Authorization requests are made by the UAS crew commander to the TA at two steps of the target engagement process: (1) after the PID checklist has been completed and confirmed, at which point the UAS crew requests an *Authorization for Kinetic Response*; and (2) once the Weapon Engagement Planning (WEP) and CDE steps have been completed and confirmed at which time the UAS crew requests an *Authorization for Weapon Release*.

The Authority Pathway includes mechanisms for communicating requests for authorization to the TA along with access to the completed checklists. The TA has the capability to approve (grant) a request for authorization or to refuse (deny) a request based on their knowledge of the ROEs and LOAC for the current mission situation. In the case of a denied request, the checklist items that were incomplete, missing or otherwise inadequate will be flagged and returned as part of the *Response for Authorization Request*. In the case of an approved request, additional information or guidance may be provided as part of the response.

2.3.3 Common/Shared Status Display

At the highest level, the Authority Pathway status display (see Fig. 2) provides an *unambiguous*, *instantaneous* view of the target engagement process status. This view serves two major purposes: (1) it contributes to the overall mission SA so that the UAS crew is aware of the current step in the process; (2) it facilitates any necessary remedial actions by clearly indicating why a given step is stalled or why a Request for Authorization has been denied.

The crew status displays are equipped with a drill-down feature that allows the operators to access increasing levels of detail concerning information items of a given checklist or response to a Request for Authorization.

2.4 Advantages and Disadvantages of Authority Pathway

The automation of alerts, notifications and warnings contribute to increasing the responsiveness and therefore the C2 agility of a UAS crew during a target engagement. At the same time, caution must be exercised to avoid an excessive number of visual and aural cues to which the crew is subjected to which case the crew is unable to respond. Automation bias is another potential disadvantage wherein the crew becomes accustomed to the system providing cues for a given event or in a given situation and therefore hesitated to identify an event in the absence of a cue from the system.

The use of automated checklists has several advantages, including: ensuring that procedures are followed; detecting/preventing human error; facilitating the communication/sharing of task status; facilitating the request for authorization process; and the possibility of recording task preparation for AAR. The adaptiveness of the system to varying operational contexts and mission types contributes to a greater C2 agility. As with automated alerts, one disadvantage of automating procedure checklists is that operators may become over-reliant on the system for knowledge of procedures and thus become vulnerable in case of system failures or when operating in a context not covered by the system or if the system is not available (e.g. a multi-national task force or exercise). As with other fields where operators rely increasingly on automation, the risks associated with automation bias must be mitigated [12].

The use of automation to formulate, communicate and track requests for authorization between the UAS crew and the TA for a variety of situations presents several advantages over traditional methods involving radio communications and chat. In particular, for POL missions, the ability to prepare, inspect, review and remediate requests for authorization to the tasking authority using semi-structured data contributes to the system flexibility (different situations) and versatility (i.e. robustness) and thus contributes to increasing the C2 agility of the UAS. However, in the case of TST situations during which a near-instantaneous response is required, in some cases traditional methods may be better suited. Also, augmenting introducing IAI capabilities into existing chat communications has been suggested as a means to bridge existing means of communication during operations with IAS technologies [14].

In the context of the Authority Pathway, the lack of a common mission SA was identified as one of the key issues during analysis of HF experiments conducted on TIGER with military personnel [6]. Therefore, providing a common/shared target engagement SA display should alleviate this issue, and will be investigated in a future trial.

The example decision ladder discussed above considers the task of CDE, which primarily involves analysis. For this reason, it is has a high potential for automation. Other crew functions and related decision-making processes, such as weapon release, likely should remain the responsibility of the human for the foreseeable future.

3 Conclusions

Through the HF analysis and experimentation of a Canadian simulation-based UAS mission experimentation testbed, it was identified that a novel UAS GCS IAI, called the Authority Pathway, could improve UAS crew SA and adherence to ROE and LOAC. The Authority Pathway must have the agility to adapt to different mission types, provide specific interfaces for specific users, and allow a level of automation that meets future requirements. The Authority Pathway will increase crew responsiveness, ensure procedures are followed, facilitate communication, and allow for comprehensive AAR. However care is needed to ensure crews do not become over-reliant and certain scenarios may require more traditional methods. These capabilities enable the Authority Pathway as a research and development tool to support stakeholder requirements: specifically identifying UAS GCS HMI recommendations.

References

1. Huang, R., Tan, P., Chen, B.: Monocular vision-based autonomous navigation system on a toy quadcopter in unknown environments. In: IEEE 2015 International Conference on Unmanned Aircraft Systems, pp. 1260–1269. IEEE Press, New York (2015). doi:10.1109/ICUAS.2015.7152419
2. Billingsley, T.B.: Safety analysis of TCAS on Global Hawk using airspace encounter models. Master's thesis. Massachusetts Institute of Technology (2006)
3. Cook, K.: The silent force multiplier: the history and roles of UAVs in warfare. In: IEEE 2007 Aerospace Conference, pp. 1–7. IEEE Press, New York (2007). doi:10.1109/AERO.2007.352737
4. Magnuson, S.: The human factor in autonomous machines. National Defense, **CI**(755), 8 (2016)
5. Garrett-Rempel, D.: Will JUSTAS prevail? Procuring a UAS capability for Canada. R. Can. Air Force J. **4**(1), 19–31 (2015)
6. Hou, M.: Testbed for integrated GCS experimentation and rehearsal (TIGER) development summary I. DRDC Technical report, DRDC-RDDC-2015-L282 (2015)
7. Canadian Air Division: JUSTAS Concept of Operations (CONOPS), Canadian Armed Forces (2013)
8. North Atlantic Treaty Organization: NATO Standardization Agreement 3994: Application of Human Engineering to Advanced Aircrew Systems (2007)
9. Schraagen, J., Chipman, S.F., Shalin, V.: Cognitive Task Analysis. Lawrence Erlbaum Associates Inc., Mahwah (2000)
10. McColl, D., Banbury, S., Hou, M.: Test-bed for integrated ground control station experimentation and rehearsal: crew performance and authority pathway concept development. In: Lackey, S., Shumaker, R. (eds.) VAMR 2016. LNCS, vol. 9740, pp. 433–445. Springer, Cham (2016). doi:10.1007/978-3-319-39907-2_42
11. Hou, M., Zhu, H., Zhou, M.C., Arrabito, R.: Optimizing operator-agent interaction in intelligent adaptive interface design. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **41** (2), 161–178 (2011). doi:10.1109/TSMCC.2010.2052041
12. Hou, M., Banbury, S., Burns, C.: Intelligent Adaptive Systems: An Interaction-Centred Design Perspective. CRC Press, Boca Raton (2015)

13. NATO STO SAS 085: C2 agility – task group SAS-085 Final report (STO Technical report STO-TR-SAS-085). NATO Science and Technology Organization, Brussels, Belgium (2013)
14. Heffner, K., Hassaine, F.: Toward intelligent operator interfaces in support of autonomous UVS operations. In: 16th International Command and Control Research and Technology Symposium (ICCRTS), pp. 1–52. Defence Technical Information Centre Report (2011). ADA547392
15. Centinkaya, O.: Architecture concerns of time sensitive targeting (TST) and NATO TST tool. In: 18th International Command and Control Research and Technology Symposium (ICCRTS), pp. 1–10. Defence Technical Information Centre Report (2013). ADA588337
16. Taylor, R.M.: Capability, cognition and autonomy. In: NATO RTO-HFM Symposium on the Role of Humans in Intelligent and Automated Systems, pp. 1–27. Defence Technical Information Centre Report (2002). ADA422249

An Evaluation of New Console Technology – Large Display – in Process Control Display

Benjamin Noah, Jingwen Li, and Ling Rothrock(✉)

Harold and Inge Marcus Department of Industrial and Manufacturing
Engineering, Penn State University, State College, PA 16801, USA
lxr28@engr.psu.edu

Abstract. The objectives of this study were to test the effect of display layout/screen type on performance in a process control task (managing a tank farm). The study compared the following two conditions: (a) 4K-resolution 55″ screen with keyboard/mouse versus (b) 6-pack screens with keyboard/mouse. A within-subject experiment was conducted among 20 college engineering students. A primary task of preventing tanks from overfilling as well as a secondary task of manual logging with situation awareness questions were designed. Primary Task performance (including tank level at discharge, number of tank discharged and performance score), Secondary Task Performance (including Tank log count, performance score), system interaction times, subjective workload, situation awareness questionnaire, user experience survey regarding usability and condition comparison were used as the measures. The 6-pack setup was found to be slightly outperformed the 4K setup in tank discharge percentage. Detection+Navigation time was approximately one second shorter in the 4K condition compared to the 6-pack condition. On the other hand, the 6-pack condition outperformed the 4K-screen in the time to enter the values by two seconds. It was also found that the total time it took participants to properly discharge tanks was not significantly different between these conditions. In terms of the subjective feedbacks, participants felt equally about the 4K-screen and 6-pack conditions. More experiments need to be conducted to resolve some of the issues and come to a clearer conclusion.

Keywords: Large display technology · Bazel · Process control display

1 Introduction

In the process control room, operators need to supervise dynamic processes, recognize unplanned disturbances and/or predict before they occur so that the proper corrective measures can be carried out in order to ensure steady state operation. Thus, process control display has attracted much attention from industry and academic since operators rely on these display design to receive information. While many challenges are being overcome through good human-centered display design and proper human-machine system evaluations, the constant arrival of new technology provides both solutions and alternative challenges. New technology may be capable of providing useful features which were not previously available, but–at the same time– usability and safety requirements need to be verified prior to implementation. For example, head-mounted

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 343–361, 2017.
DOI: 10.1007/978-3-319-58472-0_27

displays (HMD) have been shown to provide potential benefits within multiple domains [1, 2]. Heads up displays (HUD) is a similar technology which can be used for improving control in robotics [3]. Gesture control is being researched for navigation and basic control (e.g. [3, 4]). Going a step further, physiological (brain and body) techniques for control experiments have also been conducted (e.g. [5]). While these technological advancements present opportunities for improving the human-machine system, implementation of these newer technologies within the domain of process control requires careful analysis. The introduction of new forms of interaction within a field that relies on familiarity and usability could backfire if not done properly.

This study investigates the introduction of large high resolution displays that are currently available and are potentially ready for serious implementation considerations within process control. In this study, it’s tested against the traditional, ubiquitous, use of multi-monitor consoles using standard keyboards and mice for interaction. The following sections present a brief review of literature on the factors which are directly relevant to the current study, based on which we will propose our hypothesis.

Spatial layout of information, i.e., where and how to place visual objects for the users [6], is one of the key factors that has been studied to address some of the aforementioned challenges. It was determined by Vincow and Wickens’s study [7] that as more information integration was required, performance was negatively affected as the spatial distance between pieces of information increased, therefore, they suggested that items should be grouped closer together during higher levels of information integration, working memory load, and stress. This finding is in agreement with earlier, and more fundamental, studies which suggested that people are better able to recall information when presented with many attributes of a few objects, rather than a few attributes of many objects [8, 9]. Hess et al. [10] found that screen layout had a direct effect on the cognitive demands of a task, which was measured in the accuracy and response times of participants. The findings were supported by other studies like: Kandogan and Shneiderman [11] found that it is better to use hierarchically organized displays that are simultaneously presented, in a tiled layout, during dynamic task-switching work environments; Jang [12] also found that the layout of information affects performance, concluding that it is important that users are presented with multiple sources of information when integration tasks are being conducted. Simonin et al. [13] found that the radial layout outperformed the others in the visual search task. In summary, layouts need to be properly oriented in order to achieve optimum performance.

Screen/monitor arrangement is directly associated to the information layout, and both multiple-monitor and large screen setups are commonly used for control environments. Czerwinski et al. [14] demonstrated that there is a significant performance advantage to use very large, multiple monitor surfaces while carrying out complex, cognitively loaded productivity tasks. Generally, the multiple monitor layout is better in terms of inducing a cognitive layout or mental map for the user. It is assumed that users adopt a cognitive layout of the type of information to be presented and the relationships among the windows or screens and the information that they contain [15]. While it is clear that the layout of information is significant, the optimized layout is going to be dependent on the domain and the task requirements.

Within-screen Factors-Bezels: While it may at first seem that bezels are troublesome– and some research identifies potential bezel- related issues– most research has largely shown the contrary. Robertson et al. [16] found that bezels help organize work into different activities. This finding was supported by Ball and North's study [17], which proved that bezels between the monitors acted as natural dividers to help orient the participants, preventing them from getting lost in the display space in a target search task. It could be that bezels acted as dividers between different displays, and that these dividers are useful to have. However, bezels can also have some drawbacks. Going back to the study by Robertson et al. [16], they additionally found that bezels could present some issues if information is cut-off, creating visual discontinuity. This is often experienced when users are moving a mouse cursor between screens, where there is no compensation for the physical space that exists between the virtual spaces. The bezels also made reading tasks and perceiving image patterns more difficult. In a more recent study, Bi et al. [18] found that increasing the number of divisions by bezels to not effect performance in visual search and target selection tasks as long as objects are not split by the interior bezels. Bi et al. [18] found that splitting objects to have a negative effect on search accuracy and that bezels hinder performance on a straight tunnel steering task (i.e., click and drag between parallel lines between displays). Thus, interior bezels often constrain the sorts of possible layouts that are possible without splitting display objects across bezels. White space (or unused space) within a display has also been studied and found to not impact performance for search tasks [19]. In summary of the aforementioned studies, it can be suggested that bezels–or separation between displays– is important and helpful as long as the displays provide unique (individual) pieces of information. On the other hand, bezels can hinder performance if using the mouse for precision between screens is needed.

Within-screen Factors-Resolution: Transitioning to screen resolution, there is some evidence in support of using higher resolution displays. Ball and North's study [17] showed that high-resolution displays can be a benefit in that they significantly improve performance time for basic visualization tasks in finely detailed data, and they help people find and compare targets faster (up to twice as fast), feel less frustration, and have more of a sense of confidence about their responses. The same study found that there was more physical navigation (physical bodily movement) for high-resolution displays while more virtual navigation (i.e., zooming or panning in) in low-resolution displays. Also, there appeared to be a greater amount of frustration when dealing with pan + zoom as opposed to physical navigation. These findings are supported by a later study, which found that increased physical navigation on larger displays correlates with reduced virtual navigation and improved user performance [20]. These two studies favor high resolution displays. A review of using this technology within process control rooms is needed to determine if similar, and/or other, advantages exists.

Within-screen Factors-Size: Lastly, display size is another factor which has been previously investigated. The work by Andrews et al. [21] provides a very good overview and discussion of large display technology. Some have found that the larger displays should be used for higher cognitive load tasks so that less switching occurs [14, 22]. Large format displays have been shown to provide value in multiple domains, such as in military applications (e.g. [23]) and medical applications (e.g. [24]). Perceptual tasks benefit most from large displays as they allow for quicker navigation, such as navigating a map or a visualization of genes–as within the two examples cited above. The process control industry uses visualizations of similar complexities and more testing is needed to determine if these benefits can be realized within this domain. There are some challenges that exist with using larger displays. For instance, the displays typically require that they are positioned further away from the users, which can make it more difficult to control [21]. This can present issues with controlling, selecting, navigating, and linking [21]. It is common to lose sight of the mouse cursor, for example, while interacting with a large display [25]. It is suggested that, due to these interaction challenges, an alternative method to interact needs to be incorporated when using large displays [21, 26]. In this study, the use of a touchscreen is investigated with the large 4K resolution display. This may allow users to overcome some of the aforementioned challenges. It should also be noted that the current study confounds high resolution with a large display screen, simply due to the nature of the technology.

The system investigated within this study was primarily a hardware change– using a large 55 in. 4K flat screen compared with traditional multi-monitor console environments. The comparison confounds display layout with screen type as it was necessary to modify the layout of displays within the newer system, and which could not be easily replicated within the traditional system. However, effort was made to ensure that equal information was available between these systems. Therefore, the objectives of this study were to Test the effect of display layout/screen type on performance in a process control task.

2 Methods

2.1 Industrial Background and Displays

This section describes the industrial background for the primary task in this study. A customized version of the industry-leading supplier's Console Station software was developed. The primary task (A full description can be found in the *Tasks* section) involved monitoring three tank farms within three independent areas: Crude Area, Blending area, and Product area. Participants needed to prevent tanks from overfilling by detecting when a tank began filling and manually discharging the filling tanks. Three levels of displays that were used in both monitor types were described as follows, with Level 1 displays used for monitoring, Level 2 for identification of a filling tank and for navigation, and Level 3 for discharging tanks.

Level 1-Tank Farm Overview Display: As shown in Fig. 1, the Tank Farm Overview Display, providing the operator with a high-level picture of the state of the processes, shows three individual and independent tank farms: Crude Area, Blending Area, and Product area. Each area was designed with the same process structure, each containing exactly 16 individual tanks, meaning while a single tank begins to fill, and whether it overfills or is properly discharged, there was no dependency between tanks. There are 16 vertical level indicators in each area (total 48). The plot immediately below the level

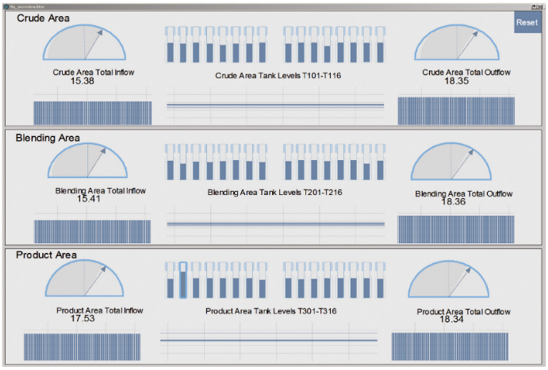


Fig. 1. The L1 tank farm overview display (Level 1, simulation is off)

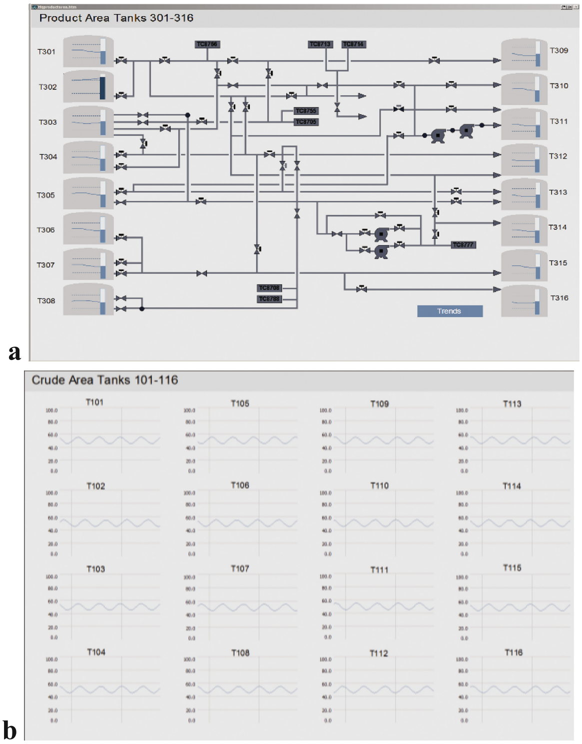


Fig. 2. Level 2 display (a) the product tank area overview (b) the crude tank area trend

indicators is a trend line chart which shows the level of all 16 tanks simultaneously, with the vertical axis representing the tank level in percent capacity (0–100%), and the horizontal axis as the time in seconds. The most current levels are plotted on the right and these levels are then updated every second, shifting the graph to the left. Total inflow and outflow indicators (including analog, dial type, flow gauge and trend bar chart) were placed on the left-hand side and right-hand side of the display respectively. The overall time range shown in both chart types is about one minute.

Level 2-Tank Area Overview Display and Trend Display: If click any of the three areas (Crude, Blending and Product) in Level 1 display, the Level 2 would be shown. Figure 2a shows the Tank Area Overview Displays for the Product and Fig. 2b shows the Tank Area Trend Display for Crude. The only differences between the three areas are the title and tank numbers (tank labels). The schematic lines and objects in the middle of the Area Overview Display (Fig. 2a) were not used in the experimental scenarios. There are 8 tanks on the left edge and 8 on the right edge, ordered numerically. Each includes two indicators: an immediate line trend chart and a level bar. These tank shapes could be clicked to access the lower level tank detail display (Level 3 display in Fig. 3). The 'Trends' button, located in the bottom right, could be clicked to access the Level 1 display for that respective area. In summary, if a tank filling event begins and is detected, the participant would need to then interact with the Tank Area Overview Display to navigate to the proper tank which is filling. The Tank Area Trend Display (Fig. 2b) shows what the trend line chart shows within the Tank Farm Overview Display (Level 1), but it separates each tank so that there are 16 individual charts. The title of each chart could be clicked to access the respective tank detail display, as an alternate navigation path to the Level 3 displays. The purpose of this display was to provide a secondary monitoring display to see what has recently happened in a single area. It could be stated that using these displays was optional as it was not mandatory to use them in order to perform well. Another type of display will be described later in the Design of Experiments.

Level 3-Tank Detail Display: Figure 3 shows a Tank Detail Display (Level 3) of Tank 101 within the Crude Area. As each area contained 16 tanks, there were 48 individual Tank Details Displays. This is the final display that users were required to navigate to in order to properly discharge a tank to prevent overfilling. On the left edge of this display, a navigation pane allowed users to click tanks within the same area

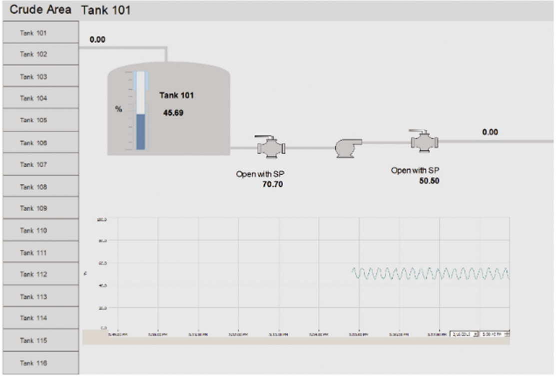


Fig. 3. The tank detail display for tank 101 (Level 3)

which would bring up the Tank Detail Display that was requested. The bottom portion of the display contains a line trend chart which shows tank level over a longer period of history than all other trend charts used in other displays. Additionally, the horizontal axis is labeled with the actual time in minute increments. The tank is shown with the tank shape in the top left, which also contains within it a level indicator and the numerical value of that level. This value is the percent that the tank is currently at with respect to its maximum capacity. Finally, to the right of the tank shape there are two valves and a pump along an exit pipe. These objects needed to be clicked, each individually, in order to call up faceplates which needed to be interacted with. Essentially, the two valves required to be opened at the set points (randomly set) shown immediately below each object using the number keypad on a keyboard. The pump needed to be turned ON using a dropdown selection panel by using a mouse. Once all three were properly opened, the tank immediately began to discharge–completing the required actions for that specific tank fill event.

2.2 Design of Experiments

Independent Variables. The study compared following two conditions: 4K large screen with keyboard/mouse versus 6 standard screens with keyboard/mouse (Fig. 4). The purpose was to investigate the effect of screen type and information layout within those screens.

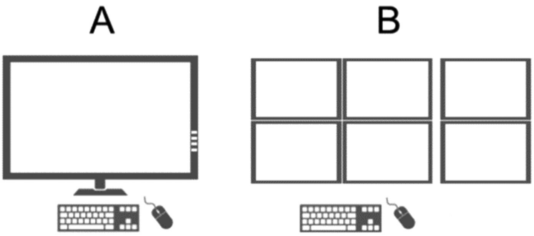


Fig. 4. Study tested case A (4K screen) versus case B (6-pack screen)

Therefore, the independent variable is the screen/layout variable, and it is a confound between screen hardware and the information layout of displays. The 4K screen used the layout that is shown in Fig. 5a. The 6-Pack screen setup used the layout that is shown in Fig. 5b. Noticing that in the 6-Pack screen layout, the right bottom one is Tank Farm Trend Display which only exist in this condition. This display shows trend line charts for all tanks in all three areas. The reason of including the display is to overcome the difference in information availability between these two conditions. As the 4K screen provided the user with the ability to view all of the three Tank Area Trend Displays simultaneously (48 tanks), the 6-monitor experimental condition only allowed for viewing of a single Tank Area Trend Display. Thus, the Tank Farm Trend Display was present in a dedicated monitor screen at all times for the 6-pack condition.

This display could not be interacted with in any form and was made to be used as a visual reference, similar to the Tank Farm Overview Display (Level 1).

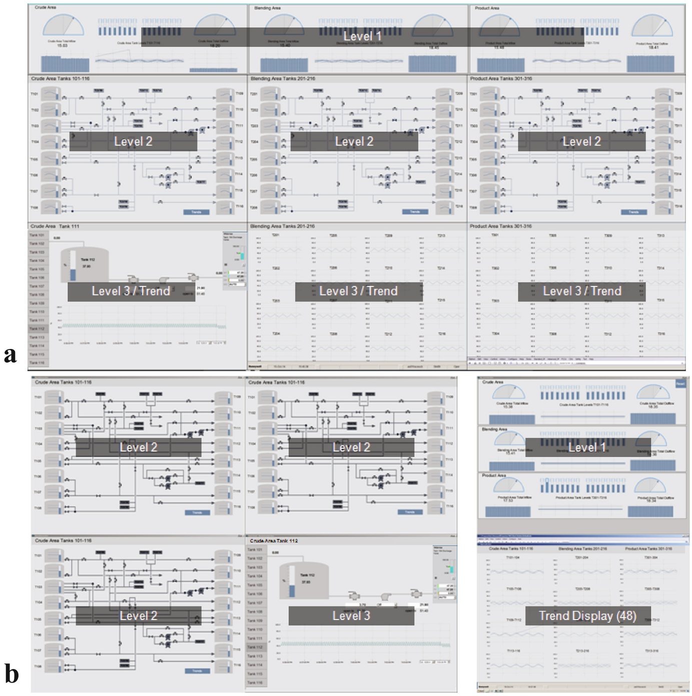


Fig. 5. Layout for two conditions (a) 4K screen (b) 6-pack screen

Tasks. Two tasks were used in the experiments. The primary task was monitoring and discharging tanks within a simulated tank farm. The secondary task was a manual logging task which used paper and pen to maintain a continuous log of tank levels. Two tasks were given equal priority.

Primary Task: The goal of the primary task was to prevent tanks from overfilling. *In* order to accomplish this, manual discharge was required through a series of actions. First, the tank that was undergoing filling needed to be identified and navigated to (to the Level 3 display). Then two valves needed to be opened and a pump needed to be turned ON. Navigating and interaction with the valves and pump required mouse clicks. Opening the valves required typing in the specified set points on the number keypad within the keyboard. Turning ON the pump required using the mouse to select the pump and then to select the 'ON' option from a dropdown menu (which only contained two choices: ON or OFF). Discharged tanks would drain down back to 50% level and the simulation automatically closed the valves and turned the pump off.

Scenario Scripts for Primary Tasks: The simulated scenarios followed pre-randomized fixed scripts with the following constraints: First, the total duration of a scenario was 20 min and scripts were broken down into 20 individual minute-blocks. Second, there were a total of 15 tanks which started filling with no overlap. Third, the time for tank filling and the time for reaction was kept within these minute-blocks. A tank fill was randomly determined to occur within the first 22 s of each minute-block. The fill rates were constant and equal between all tanks, filling from 55% level to 100% level in 38 s. Fourth, the first minute of each scenario was kept free of any event for the purpose of allowing participants to orient themselves. Fifth, four situation awareness questions were randomly preselected within individual minute-blocks, avoiding the time that a tank was filling. Figure 6 shows a breakdown of an example script, where SA refers to blocks for situation awareness questions. Scripts were created for both the experiment and also the practice scenario to train participants.

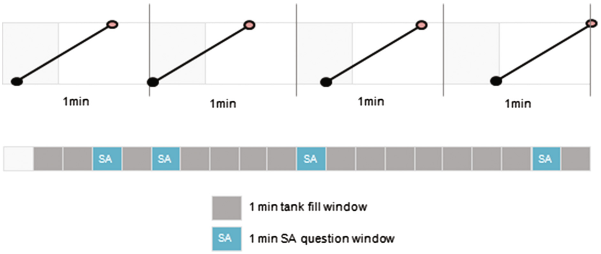
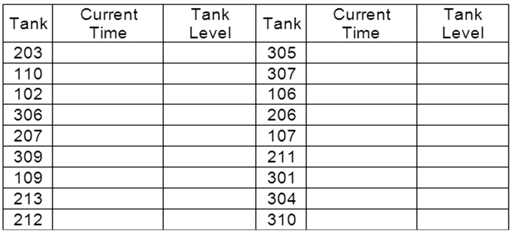


Fig. 6. An example of the structure of the experimental scenario scripts. (Color figure online)

Color indicators for primary task: The displays included warning and alarm indicators for tank fills as: no visual indicators for normal operating level range of 45%–55%; light-blue-colored, abnormal-high-warning indicators for range of 70%–79%; yellow-colored, high-alarm indicators for range of 80%–89%; red-colored, high-high-alarm indicators for range of >90%. If a tank was discharged, the colored indicators would remain present until the tank level fell back down below the aforementioned ranges. Finally, if a tank was not discharged prior to reaching 100% level, the simulation automatically reset that tank to 50% level and normal fluctuation resumed (This would be recorded as a missed event).

Secondary Task: In order to provide additional workload, the secondary task was designed as a manual logging task, which required participants to fill out a log using paper and pen as many as possible. The log listed tanks from all three areas in a randomized order (an example is shown in Fig. 7). The participants were instructed to start with the first column (the left side column) and proceed to complete each row on the log for the respective tank. The secondary task began and ended at the same times as the primary task. The final performance score on the secondary task is calculated by counting the number of tanks logged within the experimental time.



Tank	Current Time	Tank Level	Tank	Current Time	Tank Level
203			305		
110			307		
102			106		
306			206		
207			107		
309			211		
109			301		
213			304		
212			310		

Fig. 7. The manual tank logging task.

Dependent Variables. Table 1 lists all the dependent variables which were measured. For the primary task, a performance metric was calculated by subtracting the discharged level percentage from the maximum capacity (100%). Thus, higher scores are possible if tanks are discharged quicker. The score values are based on the average of all tanks discharged, which included tanks that were not properly discharged, as a zero value. The final performance score was then normalized to the 0–1 range, where 0

Table 1. The dependent variables in the study

Dependent Variable		Metric(s)	Detail
Primary task	Number of tanks discharged	Count (#)	Tanks could be discharged or overfilled
	Level % at time of discharge	Percentage (%)	As tanks fill, their level % increases and a quicker discharge ensures a lower %
	Performance score 1	Standardized score (#)	Range 0 to 1, dependent on tank level % at time of discharge
Secondary task	Tank log count	Count (#)	Number of tanks logged on log sheet
	Performance score 2	Standardized score (#)	Range 0 to 1, standardized on the individual with the highest number of tanks logged
System interaction times	Detection +Navigation time	Time (sec)	Time to reach the Tank Detail Display (Level 3) from the time when tank starts to fill
	Data entry time	Time (sec)	Time to interact with the Tank Detail Display in order to properly discharge filling tanks
	Tank discharge time	Time (sec)	Sum of (Det+Nav) and (Data entry) times
Other	Subjective workload	Index (#)	NASA TLX, scale: 0 to 100
	Situation awareness	Time to say 'ready' (sec)	SPAM Technique (see below), 4 questions per scenario, 8 questions total per participant
		Time to answer (sec)	
	Subjective situation awareness	Likert scale responses	4 subjective situation awareness questions
	System usability scale	Likert scale responses	14 usability statements (positive and negative)
	System comparisons	Binary responses	4 questions

means that no tanks were properly discharged and 1 means that all tanks were discharged immediately when filling started. For the secondary task, a performance score was calculated based on the number of logs and the individual participant who obtained the maximum number of log entries. These scores were also standardized within the 0 to 1 range. Finally, performance score 1 and 2 were summed for a total performance score, giving equal weight to each primary and secondary task.

Figure 8 represents the three times which were extracted from the experiment for the System Interaction Time metrics.

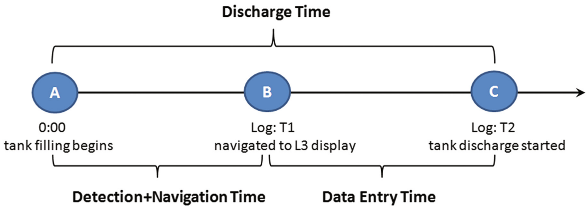


Fig. 8. System interaction times

Situation awareness was first measured with the Situation Present Assessment Method (SPAM) [27]. SPAM is an online, real-time, assessment technique to measure situation awareness. The key metric with this technique is to measure the amount of time participants take to respond to the situation awareness question. Subjective situation awareness was also measured using four Likert-scale type questions at the conclusion of each experimental scenario/condition.

2.3 Participants and Procedure

Total of 20 participants (11 males and 9 females, age range: 20–28) were recruited for this within-subject experiments. The order of the two conditions were balanced. The procedure of the experiments is shown in Fig. 9.

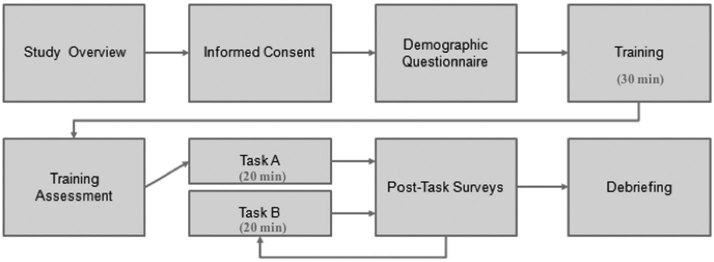


Fig. 9. Experiment protocol

2.4 Hypothesis

H1: There will be a statistical difference in performance measures between conditions

H1(a): Primary task performance scores will be different between conditions.
H1(b): Secondary task performance scores will be different between conditions.
H1(c): Detection+Navigation times will be different between conditions.

H2: There will be no statistical difference in data entry times between conditions
H3: There will be a statistical difference in subjective workload between conditions
H4: There will be a statistical difference in situation awareness between conditions
H5: There will be a statistical difference in system usability between conditions
H6: Users will prefer one condition more than the other

2.5 Data Analysis Approach

Performance data includes: primary and secondary task performance, system interaction times, situation awareness (SPAM), and subjective workload (NASA-TLX). This data was parametric and met the requirements for making the assumption of being normally distributed (tested using Anderson-Darling tests). As the comparisons were all conducted within-subject, paired t-tests were conducted to test if the means of each measure were equal (the null hypothesis). Therefore, if the paired t-tests produced results that were significant at $\alpha = 0.05$, then the null hypothesis was rejected–indicating that a difference might exist. Along with the paired t-tests, the effects size was calculated using Cohen's d parameter. From the effects size parameter, the power of the statistical test was finally calculated.

The survey data includes: the System Usability Scale, the Subjective Situation Awareness Questionnaire, and the Systems Comparisons Questionnaire. The first two questionnaires allowed for Likert scale (non-parametric) responses along a five-level scale, with the middle option always being neutral. The Wilcoxen Signed-Rank Test was used as the non-parametric version of the paired t-test to conduct similar comparisons as in the previous section (within-subject). However, two types of Wilcoxen Signed-Rank Tests were conducted. The first was a single variable test which took the median response for a question, in a single condition, and tested to see if it was equal to neutral (the null hypothesis). A significant result at $\alpha = 0.05$ would lead to the rejection of the null and indicate that the median response is not equal to neutral. Next, the Wilcoxen Signed-Rank Test was used like the paired t-test, but instead of means it compared medians between conditions. Again, the null hypothesis was that the medians were equal, and significance was determined at the same $\alpha = 0.05$ criteria. The Systems Comparison questionnaire allowed for only binary results and no statistical analysis was conducted. Instead, the direct number of responses in favor of one condition versus the other are reported.

3 Results and Discussion

3.1 Performance Results and Discussion

Results: Table 2 contains a detailed summary of the performance statistical results and they are summarized in Table 3. The statistically significant results that were better than the alternate are highlighted in green. There were four metrics which resulted in being statistically different: Tank Level at Discharge, Primary Score, Detection+Navigation Time, and Data Entry Time. No significant results were found for: performance on the secondary task, situation awareness, and workload. Below is the review of the hypothesis and the corresponding results.

Table 2. Detailed performance results

Condition	Primary Task Performance			Secondary Task Performance		Overall	System Interaction Times(sec)			Situation Awareness		Workload
	%Tanks Discharged	Tank Level at Discharge	Primary Score	Log Count	Secondary Score	Overall Score	Detection+ Navigation Time	Data Entry Time	Tank Discharge Time	(Ready) Time	(Answer) Time	NASA TLX
4K	*m*=0.97	*m*=0.83* *sd*=0.04	*m*=0.38* *sd*=0.09	*m*=69 *sd*=26	*m*=0.34 *sd*=0.13	*m*=0.72 *sd*=0.18	*m*=13.9** *sd*=2.2	*m*=12.0** *sd*=2.0	*m*=25.9 *sd*=3.1	*m*=3.2 *sd*=1.7	*m*=13.4 *sd*=3.7	*m*=54.6 *sd*=18.0
6 Pack	*m*=0.98	*m*=0.82* *sd*=0.03	*m*=0.40* *sd*=0.08	*m*=72 *sd*=24	*m*=0.36 *sd*=0.12	*m*=0.76 *sd*=0.17	*m*=15.0** *sd*=2.4	*m*=10.2** *sd*=1.8	*m*=25.2 *sd*=3.0	*m*=3.2 d=0.9	*m*=14.8 *sd*=4.6	*m*=52.2 *sd*=14.1
Paired-T Tests		*d*=0.284 *t*= 2.14 ***p*=0.0046** *Pw*=0.23	*d*=0.270 *t*= -2.16 ***p*=0.044** *Pw*=0.21		*d*=0.127 *t*= -1.10 *p*=0.284 *Pw*=0.08	*d*=0.217 *t*= -2.00 *p*=0.060 *Pw*=0.15	*d*=0.495 *t*=-2.46 ***p*=0.024** *Pw*=0.56	*d*=1.009 *t*=-6.63 ***p*=0.000** *Pw*=0.99	*d*=0.250 *t*= 1.79 *p*= 0.089 *Pw*=0.19	*d*=0.047 *t*=0.23 *p*=0.823 *Pw*=0.05	*d*=0.345 *t*=-1.38 *p*=0.185 *Pw*=0.31	*d*=0.154 *t*=0.89 *p*=0.387 *Pw*=0.10

Notes: Scores are out of 0.50 max per task, and out of 1.00 overall
** Significant difference at a=0.05, but low statistical power;*
*** Significant difference at a=0.05, with good statistical power*
d = Cohen's d; effects size; *t* = T-value; *p* = P-value; *Pw* = Power

H1: There will be a statistical difference in performance measures between conditions

Results: The data is both in support and against this general hypothesis:

H1(a): Primary task performance scores will be different between conditions. Results: The data supported this with Tank Level at Discharge and Performance Score metrics but both of these were accompanied by low statistical power.

H1(b): Secondary task performance scores will be different between conditions Results: The data did not support this hypothesis.

H1(c): Detection+Navigation times will be different between conditions. Results: The data supported this with a statistical power of 0.56.

H2: There will be no statistical difference in data entry times between conditions

Results: The data did not support this as there was a statistical difference detected, with a statistical power of 0.99

H3: There will be a statistical difference in subjective workload between conditions

Results: The data did not support this hypothesis

H4: There will be a statistical difference in situation awareness between conditions

Results: The data did not support this hypothesis

Table 3. Summary results for the System Usability Scale questionnaire.

	System Usability- (Positive Statement)							
Condition	#1 Visually appealing	#3 Easy to use	#5 Functions well integrated	#7 Learn to use quickly	#9 High confi-dence	#11 Made few errors	#13 Satisfied with system performance	#14 Good to use for a job
4K	Agree *	Agree *	Agree *	Agree *	Agree *	Agree *	Agree *	Agree *
6 Pack	Agree *	Agree *	Agree *	Agree *	Agree *	Agree*	Agree*	Agree*

	System Usability- (Negative Statement)					
Condition	#2 System complex	#4 Need tech support	#6 Inconsistent	#8 Cumbersome	#10 Had to learn a lot	#12 Not remember how to use
4K	Disagree*	s.disagree*	Disagree*	Disagree	Disagree*	Disagree*
6 Pack	Disagree*	Disagree*	Disagree*	Disagree*	Disagree*	Disagree

*Notes: * indicates value is significantly different than neutral;*
(**) indicates significant difference between conditions at a=0.05

Discussion: It was found that the 6-pack setup slightly outperformed the 4K-keyboard setup in tank discharge percentage (primary task performance). However, the statistical power of this comparison was low. In addition to that, the mean values were 0.82 (6-pack) and 0.83 (4K-keyboard)–indicating that the difference was not definitive or likely not present. It is possible that the layout of the 6-Pack setup may have enabled better performance, but it will be shown below how system interaction times likely had a larger role in this result. No statistical differences were found for situation awareness and workload. This suggests that the experimental changes (independent variables) had no effect on the user's situation awareness level or on their perceived workload. Finding that there is not much of a difference between systems in the metrics mentioned above could also be perceived as comforting to industry members, in that introducing new technology does not adversely affect these basic performance metrics (within the context for these experiments). The question at this point might be 'how does the implementation of new technology provide benefit, if any exists?'

System interaction times presented more interesting findings, which helped in answering the question that was asked above. These findings were more in line with the expectations as information retrieval, system navigation, and system interaction are all aspects which should be affected by differences in information layout and interaction method. It was found that the combined time to detect a tank filling and then navigate to the proper Tank Detail Display was faster in the 4K-keyboard condition compared to the 6-pack condition. It took participants approximately one second longer to do these actions when using the 6-pack condition. It is suggested that the layout of the 4K-keyboard condition either enabled better tank level deviation detection, or enabled users to navigate quicker. The design of this study could not differentiate between these two possibilities. On the other hand, when looking at the time it took participants to enter the values (for discharging tanks), the 6-pack condition outperformed the 4K-keyboard condition, enabling users to enter data about two seconds faster. As both conditions used keyboards/mice as the interaction method, it could be suggested that the layout differences must have played a significant role. However, as the interaction method was identical in both conditions, this came as a surprise. Fortunately, there is a reasonable explanation for what was observed. It should be noted that there was a noticeable delay in calling up displays on the 4K screen system, which would contribute to this difference[1]. The delay was primarily present when calling up the faceplate on the tank detail display, where values needed to be entered. Analysis of recorded video on both systems showed that the total delay in display call up to be between 0.5 to 1.0 s. This at least partially explains why participants took longer to enter data in the 4K screen condition. To sum up, it was found that the final time it took participants to properly discharge tanks to not be significantly different between the these conditions. So while the 4K-keyboard condition resulted in a faster Detection +Navigation time, the 6-pack condition resulted in a faster Data Entry time. The benefits of faster interaction times for each condition cancelled out one another when comparing the final discharge time, which was a summation of each system interaction time component (see Fig. 8). Had the 4K-keyboard condition not suffered from the display call up delays the overall result may have favored the 4K-keyboard condition, but further verification may be needed.

Regarding the findings in support of the layout used with the 4K screen (see Fig. 5a) compared to the layout within the 6-Pack condition (see Fig. 5b), there are some points that could be further discussed. While an equal amount of information was presented in both conditions, the organization of that information is important and can have an effect on performance as was observed within this study. Among many other researchers, Hess et al. [10] found that the layout of information affects the cognitive demands placed on individuals. The 4K screen contained columns for each of the three areas, whereas the 6-Pack screens did not have areas organized to easily recognized patterns. It could be hypothesized that one may be better than the other, but more research might be needed to verify such claims. There may be a larger effect which partially hid the effects of layout. Both conditions used a grid layout to separate the displays and this has been shown to provide real performance benefits [10, 11]. In addition, both conditions assigned specific screen space to displays, which is known to help with productivity tasks [14]. This study supports the findings of research done in different contexts and domains and only goes to further emphasize the importance of proper information layout. While much work has investigated the proper design of visualizations within displays, more work could be potentially be done in area of display layout to determine what organization of displays provides the most benefit to the console operator. A benefit of the 4K screen is that it allows for tremendous flexibility in layouts.

3.2 Survey Results and Discussion

Results: Table 3 summarizes the results for the System Usability Scale questionnaire. In summary of the statistical results, most responses generally agreed with the positive usability statements and disagreed with negative usability statements. In addition to that, no statistical differences were found in either of the studies.

---

[1] The delay in display call up was only discovered midway through data collection following comments made by some participants. The delay was likely due to a system configuration error in the 4K screen setup that is easily corrected.

Table 4 shows the summary of subjective situation awareness questions. There were no statistical differences found between two conditions. Three out of the four questions showed responses which were statistically different than neutral (see asterisks noting). Table 5 shows the responses for the Systems Comparisons. No statistical analysis was conducted on these responses. It can be seen that the study produced similar responses between conditions. There were issues with the use of: keyboard, mouse. These issues can be known from the optional comments which participants provided (are reviewed within the Discussion section). These were not technical in nature, but rather preference in usability. The relevant hypotheses proposed are reviewed below.

Table 4. Summary results for the four subjective situation awareness questions.

Condition	Subjective SA questions			
	#1	#2	#3	#4
	Aware of levels	Able to detect	Quickness to detect	Aware of 50% recovery
4K	Somewhat aware*	Somewhat able*	Somewhat/very quickly*	Neutral
6 Pack	Somewhat aware	Somewhat able*	Somewhat quickly*	Neutral

* *Indicates value is significantly different than neutral*

H4: There will be a statistical difference in situation awareness between conditions
Results: The data did not support this hypothesis

H5: There will be a statistical difference in system usability between conditions
Results: The data did not support this hypothesis

H6: Users will prefer one condition more than the other
Results: The data did not support this hypothesis

Discussion: The System Usability Scale questionnaire did not present statistical differences between the two conditions. For the most part, participants agreed with the positive usability statements and disagreed with the negative usability statements. The subjective situation awareness questions likewise, did not show any significant differences between conditions, and results were not anything out of the norm (i.e., no extreme responses at either end of the Likert scale). These questionnaire results could indicate that there are no significant concerns with any of the experimental conditions for the task that was used. Finally, the subjective comparisons questionnaire resulted in that participants felt equal preference to the 4K-keyboard and the 6-pack conditions. Eight participants did report that they had some issues with using the mouse. Most of these were related to difficulty in locating the mouse icon within the screens, for both conditions. As it can be seen, there are some useful design-related comments, but there was no single common concern among participants.

Table 5. Responses for the Systems Comparisons questionnaire.

Condition	Subjective Comparisons			
	#1	#2	#3	#4
	Condition preference	System preference	Issues with keyboard	Issues with mouse
4K	11	10	0 (yes)	8 (yes)
6 Pack	9	10	20 (no)	12 (no)

4 Limitation and Future Work

The limitations are along the lines of the design of the experiment. The first limitation is the laboratory setting and the participant pool being college students. A validation of these results should be tested in the field with console operators. The second limitation is the task used being relatively basic compared to real-world control room operations. Again, a field investigation could incorporate multiple tasks which require regular interaction with the equipment. Stating this, it is expected that the results are generalizable to industry as the significant effects are fundamental to human-computer interaction – expertise is not expected to have an effect on the system interaction times for example. Another limitation is in reference to the equipment that was used. As a delay in display call-up was encountered within the 4K conditions, some of the effects became less clear in the analysis. This can be overcome in future investigations through proper hardware/software implementations.

5 Conclusion

This study compared the 4K display with the 6-pack displays using the keyboard/mouse interaction method. The 4K display's layout seemed to allow for the faster time in detection of tanks filling and navigation to the tank detail displays. However, the 6-pack condition showed faster data entry time (time to discharge tank by opening two valves and a pump). This resulted in slightly quicker tank discharges for the tanks in the 6-pack condition, seen by a lower average for tank level percentage at the time of discharge. The time to enter data should, hypothetically, have been similar as the interaction method remained constant. There was noticeable delay in display call up using the 4K screen, which could explain this discrepancy. So while the 4K screen layout seemed to allow participants to better navigate and detect tank fillings, the display call up delay may have slowed down data entry within the 4K setup. Therefore, there is some evidence to support that a 4K-keyboard setup could potentially outperform the 6-pack setup in these performance metrics if the display call up delay is resolved.

References

1. Phipps, D.: The human–computer interaction handbook: fundamentals, evolving technologies and emerging applications. Ergonomics **56**(8), 1345–1346 (2013)
2. Ruddle, R.A., Payne, S.J., Jones, D.M.: Navigating large-scale virtual environments: what differences occur between helmet-mounted and desk-top displays? Presence Teleoperators Virtual Environ. **8**(2), 157–168 (1999)
3. Eliav, A., Lavie, T., Parmet, Y., Stern, H., Edan, Y.: Advanced methods for displays and remote control of robots. Appl. Ergon. **42**(6), 820–829 (2011)
4. Alcoverro, M., Suau, X., Morros, J.R., López-Méndez, A., Gil, A., Ruiz-Hidalgo, J., Casas, J.R.: Gesture control interface for immersive panoramic displays. Multimedia Tools Appl. **73**(1), 491–517 (2013)
5. Nelson, W.T., Hettinger, L.J., Cunningham, J.A., Roe, M.M., Haas, M.W., Dennis, L.B.: Navigating through virtual flight environments using brain-body-actuated control. In: Virtual Reality Annual International Symposium 1997, pp. 30–37. IEEE (1997). doi:10.1109/VRAIS.1997.583041
6. Lok, S., Feiner, S., Ngai, G.: Evaluation of visual balance for automated layout. In: Presented at the Proceedings of the 9th International Conference on Intelligent User Interfaces, pp. 101–108. ACM (2004)
7. Vincow, M.A., Wickens, C.D.: Spatial layout of displayed information: three steps toward developing quantitative models. In: Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 37, pp. 348–352. SAGE Publications (1993)
8. Harwood, K., Wickens, C., Kramer, A., Clay, D., Liu, Y.: Effects of display proximity and memory demands on the understanding of dynamic multidimensional information. In: Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 30, pp. 786–789. SAGE Publications (1986)
9. Yntema, D.B.: Keeping track of several things at once. Hum. Factors J. Hum. Factors Ergon. Soc. **5**(1), 7–17 (1963)
10. Hess, S.M., Detweiler, M.C., Ellis, R.D.:. The effects of display layout on monitoring and updating system states. In: Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 38, pp. 1336–1340. SAGE Publications (1994)
11. Kandogan, E., Shneiderman, B.: Elastic windows: improved spatial layout and rapid multiple window operations. In: Presented at the Proceedings of the Workshop on Advanced Visual Interfaces, pp. 29–38. ACM (1996)
12. Jang, J.: Lost in space: the surprising role of information spatial layout (2009)
13. Simonin, J., Kieffer, S., Carbonell, N.: Effects of display layout on gaze activity during visual search. In: Costabile, M.F., Paternò, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1054–1057. Springer, Heidelberg (2005). doi:10.1007/11555261_103
14. Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., Starkweather, G.: Toward characterizing the productivity benefits of very large displays. In: Presented at the Proceedings of INTERACT, vol. 3, pp. 9–16 (2003)
15. Norman, K.L., Weldon, L.J., Shneiderman, B.: Cognitive layouts of windows and multiple screens for user interfaces. Int. J. Man Mach. Stud. **25**(2), 229–248 (1986)
16. Robertson, G., Czerwinski, M., Baudisch, P., Meyers, B., Robbins, D., Smith, G., Tan, D.: The large-display user experience. IEEE Comput. Graph. Appl. **25**(4), 44–51 (2005)
17. Ball, R., North, C.: Effects of tiled high-resolution display on basic visualization and navigation tasks. In: Presented at the CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 1196–1199. ACM (2005)

18. Bi, X., Bae, S.-H., Balakrishnan, R.: Effects of interior bezels of tiled-monitor large displays on visual search, tunnel steering, and target selection. In: Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 65–74. ACM (2010)
19. Zdralek, J.: White space: how much nothing should there be? National Library of Canada = Bibliothèque nationale du Canada (2004)
20. Ball, R., North, C., Bowman, D.A.: Move to improve: promoting physical navigation to increase user performance with large displays. In: Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 191–200. ACM (2007)
21. Andrews, C., Endert, A., Yost, B., North, C.: Information visualization on large, high-resolution displays: issues, challenges, and opportunities. Inf. Vis. (2011). 1473871611415997
22. Baudisch, P., Good, N., Bellotti, V., Schraedley, P.: Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In: Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 259–266. ACM (2002)
23. Gouin, D., Evdokiou, P.: A showcase of visualization approaches for military decision makers. DTIC Document (2004)
24. Hibbs, M., Wallace, G., Dunham, M., Li, K., Troyanskaya, O.: Viewing the larger context of genomic data through horizontal integration. In: Presented at the 11th International Conference Information Visualization, IV 2007, pp. 326–334. IEEE (2007)
25. Czerwinski, M., Robertson, G., Meyers, B., Smith, G., Robbins, D., Tan, D.: Large display research overview. In: Presented at the CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 69–74. ACM (2006)
26. Moreland, K.: Redirecting research in large-format displays for visualization. In: 2012 IEEE Symposium on Large Data Analysis and Visualization (LDAV), pp. 91–95 (2012). doi:10.1109/LDAV.2012.6378981
27. Durso, F.T., Dattel, A.R., Banbury, S., Tremblay, S.: SPAM: the real-time assessment of SA. Cogn. Approach Situat. Aware. Theor. Appl. **1**, 137–154 (2004)

Use of Graphic Imagery as a Mean of Communication Between Operators and Unmanned Systems in C3Fire Tasks

Tal Oron-Gilad(✉) and Ilit Oppenheim

Ben-Gurion University of the Negev, Beersheba, Israel
orontal@bgu.ac.il

Abstract. Digitization in the battle field enables the creation of new communication means among distributed elements. We aim to examine how the use of graphic imagery and means of graphic communication facilitate performance, shorten the OODA loop (Observe–Orient–Decide–Act), and enhance operators' situation awareness. Previous research (Oron-Gilad and Oppenheim 2015) looked at bi-directional graphic communication on video feed derived from an unmanned aerial system (UAS), allowing a dismounted commander and the operator of the UAS to view and communicate on the video imagery. It was concluded that the commanders favored the ability to use non-stationary markings on the video feed relative to no markings at all or anchored-fixed markings. Furthermore, the verbal communication between the team mates changed as the team gained experience over consecutive trials, and with the use of graphic communication.

In the current study, we examined the added value of using still images, one-way, graphic communication from an operator, represented by Coral, tablet or UAS observers to a ground fighter, represented by either Tank (mounted) or Spike-MR (dismounted) attackers, while collaborating to acquire stationary targets in a rural built area. The abilities to communicate by using still images with or without additional markings (annotations) were aimed to be compared to a basic communication form based on 'coordinates' and to a future solution of augmentation of the target's location on reality. Difficulties in accurately implementing the augmentation have severely hindered the accuracy of performance in this communication mean. Therefore, comparisons were made relative to 'coordinates' as a baseline and to the use of still images (with annotations or without). Verbal communication within the teams (two-way) was available at all times. Eight teams of three; 'attacker', 'observer', and 'controller' (as the higher echelon commander) participated in the experiment. Participant soldiers were assigned to a simulated work station according to their expertise. The dependent variables were (1) objective measures - execution, response time and accuracy, and (2) subjective evaluation of – usability, stress, and the quality of communication, and (3) verbal communication.

The results show that in the Ground-Ground teams, operators benefited from the use of still images especially in the time to acquire a target and the accuracy distance from target. In contrary, still images from the UAS were less beneficial to the ground operators relative to 'coordinates' in terms of time to acquire a target and executions. Overall, observer-type participants favored the use of still images. When they had free choice, they chose the still image configurations

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 362–381, 2017.
DOI: 10.1007/978-3-319-58472-0_28

almost in all cases (98%). Yet, there were differences in the way the observers and the attackers perceived the quality of the graphic communication, and there were differences, between the mounted and the dismounted attackers as well. All in all, it seems that the addition of graphic communication increased the number of ping-pong chats between the team members, but, on average, shortened the duration of each chat. From the perspective of the controllers, who were supervising the teams, they noted that the Ground-Ground teams used less verbal communication compared to the Ground-Aerial teams. Furthermore, less verbal communication was required when using the still images interface (with or without markings) compared to 'coordinates'.

We conclude that; (1) Still images may improve the communication between Ground-Ground teams; (2) In some cases, adding the ability to graphically mark elements on the still image (i.e., annotations) improved the communication in terms of self-evaluation of team cooperation and performance, even more than still images alone; (3) In presence of graphic communication, the verbal communication patterns have changed with more 'ping-pong' transmissions among team members although shorter ones; and (4) When participants had free choice of communication means they preferred the still images with graphic markings over the other alternatives of 'coordinates' and still images without markings).

Further investigation is needed to test the added value of using bi-directional graphic communication (either on still images or by temporary markings on dynamic video scenes).

Keywords: Close Target Reconnaissance (CTR) OODA loop · Graphic communication · Ground and aerial views · Mounted and dismounted attackers · Observer-Attacker teams · OODA loop · Short cycle

1 Introduction

One common notion in combat performance is known as Boyd's OODA (observe, orient, decide, act) loop, and is drawn from military strategies to present real-time decision-making processes. An elaborative model was later offered by Boyd to comply with more complex forms of combat and feedback loops available in the decision-making process (see Brehmer 2005). In this elaboration, the 'Observe' stage consists of multiple information sources, both internal, derived from the unfolding of the circumstances and the immediate environment, and external, outside information provided by others. Observation is often influenced by implicit guidance of higher echelons or sources who may have broader perspective of the mission. In the 'Orient' stage, synthesis of information is accomplished; from representing the physical location of various elements in the immediate environment to a mental model of the situation. While not using the same terms, this process is somewhat similar to what is being described as generating the mental simulation in Klein's RPD model (Klein and Crandall 1995).

1.1 Use of Visualization Concepts to Facilitate "Observe and Orient"/ Sensemaking

Integrated visualization concepts have been shown to aid sensemakers (Ntuen et al. 2010) in complex information environments. Baber et al. (2010) proposed a framework representing two parallel cycles; a short Close Target Reconnaissance (CTR) OODA loop cycle and a broader cycle of recording, communicating and interpreting information that feeds into the short cycle and improves sensemaking, situation awareness, and situation understanding in context (e.g., imagery provided by cameras to supplement technologies such as night vision goggles or binoculars). Hence, the input fed into the short CTR loop should be relevant, accurate and timely and the notion that information is "a means to an end, not an end in itself" must be continuously stated. Oron-Gilad and Parmet (2016) have used the OODA loop framework to analyze the decision cycle of dismounted soldiers in a patrol mission who received video feed from an unmanned ground vehicle that was ~20–50 m ahead. They found that the addition of video feed data to the moving dismounted soldier had several detrimental effects on soldiers' orientation and response to events in their immediate environment, especially ones that were not seen by the technology. Their field evaluation highlights the costs that adding graphical information may have on the observe and orient components of the OODA loop, thus raising the need for new roles in combat-team setups and for additional training when unmanned vehicle sensor imagery is introduced to end-operators. Indeed, judgment and decision processes require shared knowledge as well as efficient communication. Technology enables teams to perform tasks together while they are in different locations and using various communication means. Yet, it also increases the need for common assessments or common mental models of situations during the decision-making process (Mosier and Fischer 2010). The concept of common ground (CG), relate to its contributions to mutual knowledge, beliefs, and assumptions that inspire social and collaborative activities. CG is required for the comprehension of normal conversational interactions and is essential for the coordination of joint actions (Cumming and Akar 2005). Different media provides different resources or affordances that shape communication (Kraut et al. 2002).

1.2 Use of Video to Facilitate Co-presence of Non-military Teams

Fussell et al. (2003) found that pairs work best when they are located side-by side and share full visual co-presence. On the contrary, pairs worked least well when they only had audio communication and they couldn't see the work area. Some of the benefits of shared visual space can be provided through technology; A scene-oriented camera showing a wide-angle view of the workspace provided significant benefit over audio-only communications. However, a head-mounted camera with eye-tracking capabilities provided little benefit. Moreover, the combination of head-mounted camera and scene camera did not enhance pairs' effectiveness over the scene camera alone, and in fact led to longer performance times than the scene camera alone. The usage of two cameras caused decision difficulties related to how to distribute attention between them and also confusion in understanding which one was in use. Such findings caution against strategies to create shared visual space through multiple video feeds. The authors concluded that providing a wide-angle static view was the most valuable form for remote collaboration on physical tasks. In another non-military study, the focus was on use of video as a communication mean for capturing the images of remote participants as they perform tasks together. In this type of remote-collaborative work domain, video feed that was used to communicate visual aspects of interaction, such as eye gaze, physical gestures, and facial expressions, had no effect on either the quality of the interaction or the outcome of the task. In situations where visual communication consisted of important content that was needed to improve coordination and collaboration (e.g., a neurosurgical procedure in an operating room) the results were more promising, hence, using video as data rather than as a concept of shared events or as an indication of who is present in an activity, was beneficial. These types of contextual video images provided means to maintain the team members' (including those who did not have an active role) attention to the operation at any given time and facilitated coordination of fast-paced activities between members of the team (Obradovich and Smith 2008).

Hew (2011) proposed a structured, graphical Data-Tracks-Actions (DTA) representation approach to analyzing C2 teams and their technologies. The graphic representation depicts where and how situation awareness is being formed, how it translates into battlespace actions, and the command roles that form and adapt the end-to-end workflows. In a somewhat similar way to the OODA loop representation for an individual, in the context of perception, sensemaking and action of a team, the graphic representation proposed by Hew can serve to detect roles and intra-team relationships from introducing new technological capabilities. Most importantly to our study, Hew separates between the various communication systems and technologies, their tempo and what they enable. Thus, voice messages can be directed or shared at a tempo of 5–10 and 20–30 s per position. Other communication technologies (e.g., video or imagery) varies in purpose and tempo. To exemplify, Hew (2011) presented a field artilery case study which demonstrated how this analytic approach helps in dealing with issues and opportunities in C2 design.

There are limitations to video systems in providing the shared visual space and those must also be considered (e.g., which visual cues are enabled, what is the field view of each participant, and what level of detail is needed about the work area). While the remote viewers of the video are physically present, they benefit also from contextual cues in addition to the data (hence creating information) from the video. This context could be lost if the team members are not co-located and are provided with the video feed but with no other means for maintaining common situation awareness about the broader context. To sum, practical implications of Obradovich and Smith (2008) study were: "Use video when an accurate and informative picture is needed". When tasks are assigned between team members, when some have control and others have the data or knowledge, coordination between teams as part of the interdependencies among activities management as well as interactions might be required (Obradovich and Smith 2008).

1.3 Use of Visualization Concepts to Improve Dynamic Decision Making in C3Fire Tasks

Artman (1998) failed to answer the question of whether a graphical or a textual database would improve dynamic decision making in C^3Fire (Communication, Command and Control) tasks. However, he concluded that the team's situation awareness composes of coordination and trust in others, and understanding in what way the team's actions (e.g., interactions or consequences) affect the dynamic system development. Granlund et al. (2011) further used the C^3Conflict simulation environment which is based on the C^3Fire microworld ("simulated environments that realistically capture important characteristics of a real system including the complex, dynamic and opaque characteristics of decision making problems"), but was tailored to the military domain by presenting situations with a military cover story and appearance. This environment is a two-sided game where one of them can be hostile. The C^3Conflict adds analysis abilities such as: the effectiveness of the teams, the information distribution in the team, and the team's work and collaboration methods.

Large-scale digital command and control systems often suffer from suboptimal performance due to Interface problems. Walker et al. (2010) found that with the relatively simplistic voice communication, more data was transformed into useful information than with the highly complex digital communication. People prefer a simple interaction that enables to do complex tasks quickly, rather than a complex interaction that only allow to do simple things with considerable effort. The digital communication layer carried significantly higher proportion of data compared to the voice layer, which in turn carried a greater proportion of information. Furthermore, "data" was received by the Brigade headquarters (96%) compared to a greater proportion of "information" leaving it (26%). Walker et al. (2010) suggested to reconsider the way that digitization should be perceived, designed, and operationally prepared - a human-centric view rather than a techno-centric view (in which capability is viewed in terms of technological advancement).

The goal of the current experiment was to examine whether adding simple one-directional graphic communication capabilities to an existing verbal communication channel between an observer and an attacker team members will improve mission performance (as defined by objective performance measures) and reduce the stress level of the participants (as measured by the DSSQ). The observers were provided with the technological capability to send still images to the attackers, in one of two forms (as taken, or with annotations). Delivery of still images was compared to two other alternatives: sending target coordinates (baseline) and pointing the target directly on the attacker's imagery (augmentation).

It was hypothesized that: (1) the availability of still images will improve the communication between the observer and the attacker; (2) adding the ability to graphically mark elements on the still image (annotations) will improve the communication and shorten the mission cycle, even more than still images alone; (3) the verbal communication pattern among team members will change due to the presence of the graphic communication; and (4) when participants will have a free choice of communication means they will prefer the still images with graphic markings on the other three alternatives (coordinates, augmentation on reality and still images without graphic markings).

2 Methodology

2.1 Participants and Allocation to Teams

Eight teams of three teammates: 'attacker' (dismounted/mounted soldier), 'observer' (operator of a remote device, aerial or ground), and experimental 'controller' (higher echelon commander). Each one of the attacker and observer participants took part in one experimental day. The same controllers participated as experimental confederates in all 3 experimental days. There were four teams of Tanks as attackers; two with Tablet observers, one with Coral, and one with UAS. There were four teams of Spike-MR attackers; two teams were Coral, one with Tablet and one with UAS observer. Participants were all soldiers or military reserve soldiers who have been on active duty in the year prior to the experiment. Each participant was assigned to a simulated work station according to his/her expertise. E.g., UAS oper-ator was assigned to operate the UAS etc.

2.2 Instruments and Apparatus

A set of questionnaires was administered at the end of the experiment. It consisted of: (1) personal characteristics and military experience questionnaire, (2) a 15 items usability questionnaire based on the SUS – System Usability Scale (Brooke 1996) using a five-point scale from 1 "Strongly disagree" to 5 "Strongly Agree", (3) a 20 items Dundee Stress State Questionnaire (DSSQ; Matthews et al. 1999, 2002) using "0" (Definitely false) to "4" (Definitely true) scale, after each round, and (4) Quality of communication questionnaire, which had two different versions; one for the soldiers and the other for the operators (based on Fiore et al. 2003).

2.3 Procedure

Participants arrived at the IDF battle lab for approximately one day of experimentation (~7–8 h total), including briefing and training (~1 h), short breaks between and within experimental rounds (2 h each, 2 rounds, 3rd round ~40 min, and a lunch break). Rounds #1 and #2 consisted of four sessions, each with different type of communication method (coordinates, augmentation on reality, still images with and without marking). The order was counterbalanced between experimental sessions. Each participant was given oral instructions about the task prior to beginning the experimental trials. The instructions included information such as the background and motivation for the experiment, its duration and pay-ment, ethics, safety and confidentiality. They had to sign an informed consent be-fore participating.

A practice session took place (Round #0, "training" scenario) before the experimental trials began. The practice session was used to introduce participants to their teammates, and to get familiar with the general tasks and tools. Participants were asked to fill in the DSSQ after each of the four rounds, and after the simulation parts ended, participants were asked to fill in all other questionnaires (Personal characteristics & military experience, Quality of communication questionnaire, and SUS - System

Usability Scale). Finally, participants were asked if they would like to provide any additional comments or concerns orally, they were debriefed and compensated and the experimental session ended.

3 Results

3.1 Objective Results (Simulation)

A total of 400 targets were programmed and planned for the three rounds and of them 305 ($\sim$76%) were executed during the experimental sessions. Hence, none of the teams completed all possible targets. Of the 305 targets, 100 targets out of 128 (78%) were executed within round #1, 152 out of 208 (73%) within round #2 and 53 targets out of 64 (83%) within round #3. Note that round #3 was shorter than rounds #1 and #2.

With regard to viewing angle difference (i.e., where the observer was located relative to the attacker in the ground-ground teams), the distribution across rounds #1, #2, and #3 was 100 targets of narrow angle (up to 30°), 132 targets of mid angle (30–80°), and 73 targets of wide angle (80–150°).

Objective results were analyzed in three aspects: execution analysis (i.e., whether there was fire or it was ceased by the controller of the experiment), response time to acquire a target, in seconds, and accuracy of execution in meters (for acquired targets). In some specific cases of the attacker device simulations, the implementation of augmentation on the target in the simulation turned out to be problematic, as there were discrepancies in the position of the augmentation and the target itself, implicating on objective performance. Therefore, this information was excluded from the statistical analysis related to execution and accuracy distance.

Execution Analysis. The execution analysis includes rounds #1 and #2 and is detailed separately for teams with the Spike-MR (dismounted) attacker and teams with the Tank (mounted) attacker. Execution was defined as a binary variable (1-for fire 0-for cease fire).

Dismounted attacker. Table 1 and Fig. 1 detail the number of executions that ended either with fire or ceasefire (i.e., time run outs), for the Spike-MR as attacker and the different observer types. Different patterns of performance can be seen for the different team combinations. Fire was executed in 73% of the cases on average. A logistic regression within the framework of the GLM (generalized linear model) was chosen for analysis. The model yielded a marginally significant effect for communication ability ($p < .08$) only for the Spike-MR+UAS configuration. No other effects were significant. Hence, for the Spike-MR+UAS team there was a trend for fewer cease fires in the coordinates (baseline condition) relative to both still images types. In contrary, in the Ground-Ground teams (Spike-MR + Coral and Spike-MR + tablet), there seemed to be more fire executions using still images (either with or without markings) compared to the coordinates and the augmentation of location communication abilities, but these were not significant differences.

Mounted attacker. Table 2 and Fig. 2 detail the number of executions that ended either with fire or ceasefire since time runs out, for the teams of Tank and the different observers. As can be seen, more events ended with fire in the case of the Tank as an attacker (92% on average). Using the logistic regression analysis here yielded no significant main effects for communication ability.

Table 1. No. of executions (fire and ceasefire) per teams of Spike-MR as an attacker, the different observers by communication means.

Teams of attacker + observer	Execution	Communication means				Total
		Coordinates	Still images	Still images with markings	Augmentation of target location on reality	
Spike-MR + Coral (2 teams)	Fire	8	14	12	7	**41**
	Cease fire	2	4	4	1	**11**
Spike-MR + tablet (1 team)	Fire	5	6	8	2	**21**
	Cease fire	2	4	2	2	**10**
Spike-MR + UAS (1 team)	Fire	5	3	3	5	**16**
	Cease fire	0	3	4	1	**8**
	Total	**22**	**34**	**33**	**18**	**107**

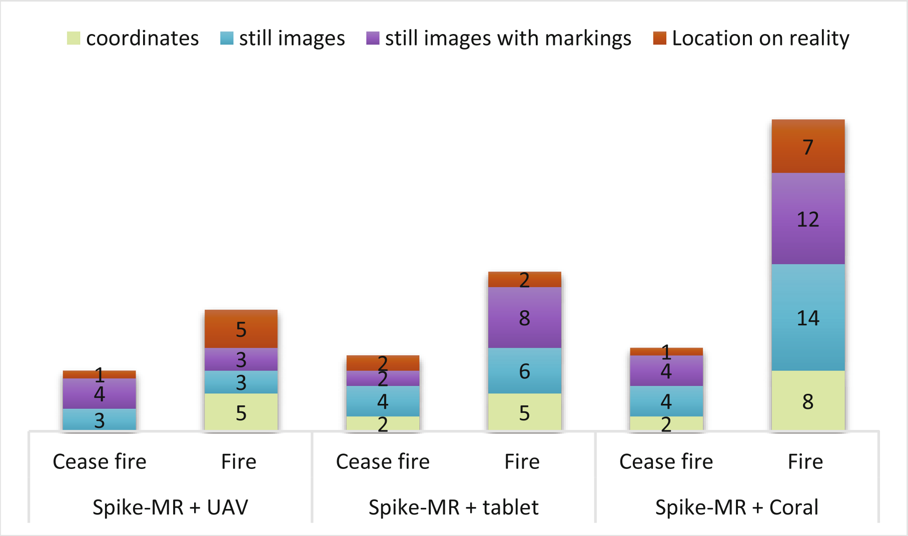


Fig. 1. Distribution of executions (fire and ceasefire) per teams of Spike-MR as an attacker and different observers and communication means.

Table 2. No. of executions (fire and ceasefire) per teams of Tank as an attacker and different observers and communication means

Teams of attacker + observer	Execution	Communication means				Total
		Coordinates	Still images	Still images with markings	Augmentation of target location on reality	
Tank + Coral (1 team)	Fire	5	13	9	4	**31**
	Cease fire	1	0	2	1	**4**
Tank + tablet (2 teams)	Fire	14	23	24	12	**73**
	Cease fire	2	1	1	1	**5**
Tank + UAS (1 team)	Fire	6	10	8	5	**29**
	Cease fire	0	0	1	2	**3**
	Total	**28**	**47**	**45**	**25**	**145**

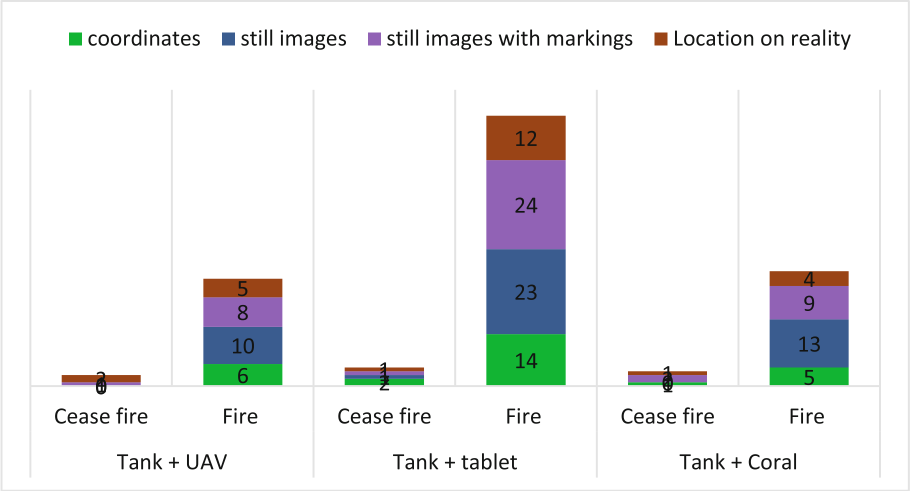


Fig. 2. Distribution of No. of fire and ceasefire per Tank attacker & different observer teams and communication means

Free Round analysis. The last round (#3) enabled participants to choose any of the four communication means (coordinates, still images, still images with markings, and augmentation on reality). Here, fire was executed in 43 out of 53 (81%) of the cases, and the other 19% of the cases ended with no fire because time has run out. The results clearly show preference for the still images communication ability with 33 (77%) fires using still images with markings and 9 (21%) using still images without markings. The ‘augmentation on reality’ communication was used only once (but recall also the comment about its implementation accuracy, hence this type of augmentation is very sensitive to the accuracy of augmentation implementation).

Response time analysis. The time to acquire a target was measured as the time from the beginning of the trial till execution. According to Parmet et al. (2014) the way most common response time analysis methods treat cases of no response (i.e., missing data) is inaccurate, and therefore they suggested using an alternative analysis technique, named survival analysis, that lead to more reliable and robust conclusions. Survival analysis is a branch of statistics which deals with death in biological organisms and failure in mechanical systems. Generally, survival analysis involves the modeling of time to event data, in our case, the event is the participant's response (or no response) to a traffic scene. Survival analyses are statistical methods and procedures that accommodate censored data. Procedures that treat differently the information gained from uncensored and censored observations.

We fitted the Cox proportional-hazards regression model Cox (1972) which is the most common tool for studying the dependency of survival time on predictor variables. The initial model included the Communication ability (Coordinates, Still images, Still images with annotations, and augmentation on reality), the angle, the interaction between the two. Interaction was not statistically significant for any of the models and therefore removed from the analysis. The main effect for communication ability was found statistically significant in the cases of Spike-MR + UAS team and the Tank + Coral team, see Fig. 3.

Accuracy analysis. The accuracy of target acquisition was measured by the distance between the target and the impact point. A logarithmic scale of the distance was used to display the data (Fig. 4).

Utilizing a GLM analysis on the log of the distance from target acquired (normally distributed) with communication ability (Coordinates, Still images, Still images with annotations) as the predicting variable, statistical main effects for communication ability were found for the Spike-MR & Coral team configuration ($F(3,37) = 2.25$, $p < .097$), Spike-MR & Tablet team configuration ($F(3,17) = 5.22$, $p < .0097$), where in both, still images only and still images with annotations, yielded shorter distances from the target than the coordinates.

With regard to the viewing angle, it was not balanced well across communication ability conditions by configurations, as can be seen in Fig. 5. Nevertheless, there is a trend showing that the accuracy distance of trials with still images were less sensitive to viewing angle than the other communication ability means. This finding needs to be replicated in future studies before making any clear statement.

3.2 Subjective Results

SUS (usability evaluation). After completing the entire experimental session and using the system in all four possible modes of communication participants had to rate the usability of the system. They were asked to record their immediate response to each of the 15 items on a 5-points Likert scale. SUS yields a single number representing a composite measure of the overall usability evaluation of the system. SUS scores ranged from 55 to 100 (out of 100). Participants' (both attackers and observers) average score

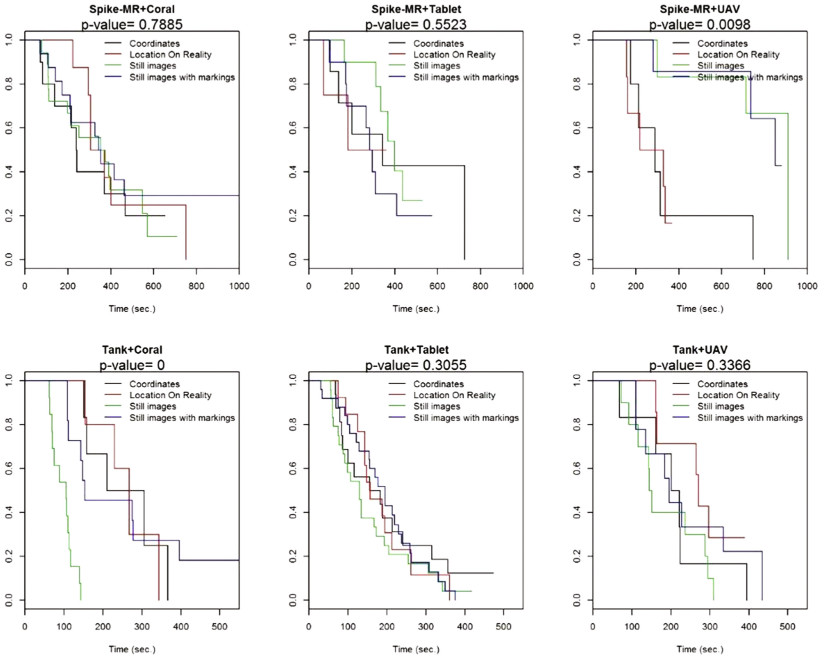
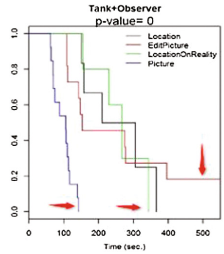


Fig. 3. *-Survival analysis for time to acquire a target. Significant effects for communication ability were found in the Tank + Coral (Observer) team configuration and in the Spike-MR (NT) +UAV team configuration.

* How to read survival analysis graphs. The Y axis represents the probability to not acquire a target, the X-axis represents the time. Hence, as the time increases, the probability that the target is acquired increases. For example, in the Tank + Coral configuration, the fastest responses were obtained with still images (blue) reaching 100% in less than 150 seconds. If there is a plateau line, and the 0 is not reached, it means some of the targets were not obtained (see the red line, about 20% of the targets were not acquired).



was 88 (SD = 9). Hence, overall participants were satisfied with the communication user interface.

DSSQ (stress evaluation). This questionnaire is concerned with participants' feelings and thoughts while performing the task (on 0–4 scale). The DSSQ measures three aspects of subjective stress; task engagement (related to task interest and focus: energetic arousal, motivation, and concentration), distress (integrates unpleasant mood and tension with lack of confidence and perceived control), and worry (composed of self-focused attention, self-esteem, and cognitive interference). The DSSQ was collected

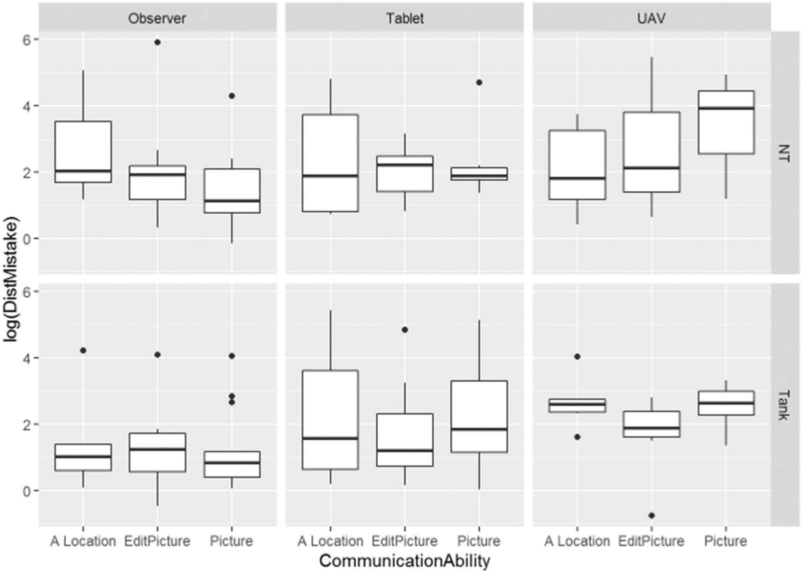


Fig. 4. Distance from target in meters (logarithmic scale) for communication ability and the various team configurations.

after each experimental round (i.e., 4 times), hence, percent of change over the course of the experimental day could be calculated. The overall averages were 24 (ranged 9–28, SD = 4), 7 (ranged 0–17, SD = 4) and, 4 (ranged 0–16, SD = 4), respectively for task engagement, distress and worry. The DSS scores after each round are detailed in Table 3. Note that the maximum 'engagement' scores which can be achieved in the DSSQ is 28. Therefore, it seems that, on average, participants were highly motivated and engaged in doing the task. The potential highest scores for 'distress' and 'worry' are 28 and 24 (respectively). This can point out that participants were pretty relaxed and became progressively even less worried as the tasks progressed.

Quality of Communication Questionnaire. In view of the problematic implementation, attributable to discrepancies in the position of the augmentation and the targets, the following analysis of communication quality excludes the 'Augmentation of target location on reality' mode. Two different versions of a subjective assessment of the communication quality were used; one for the attackers and one for the observers. The questions were divided into three groups – cooperation items (4 items in both the attackers' and the observers' versions, three of them were identical), coordination items (4 items for both attackers and observers, three of them were identical), and performance items (4 items in the attackers' version and 2 items in the observers' version).

The three shared 'cooperation' items were aimed at evaluating team work and interaction (i.e.; "Team cooperation was good", "We used the same jargon", "A unique common language was created between us"). Figure 6 presents the different patterns of evaluation among the different teams and communication means.

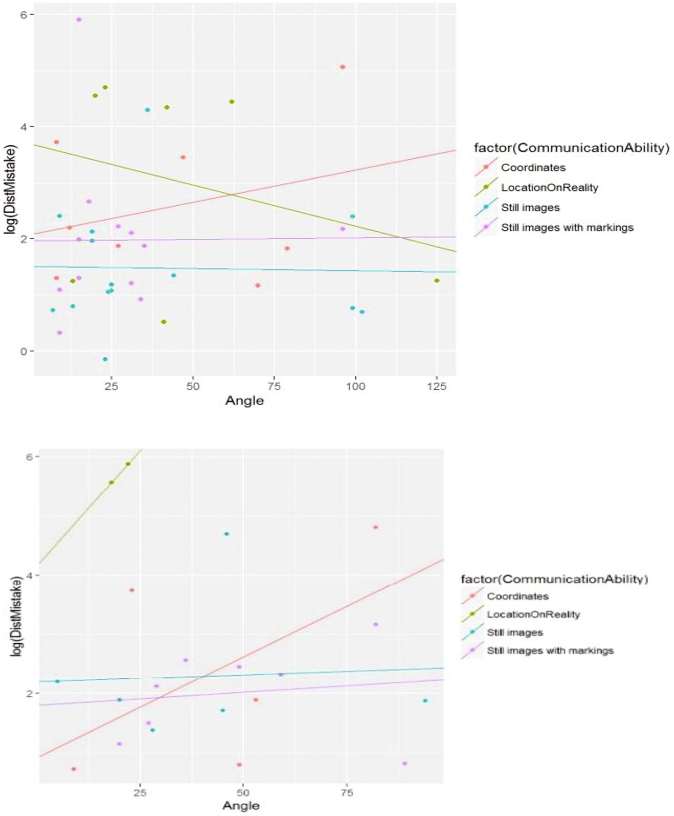


Fig. 5. Distance from target in meters (logarithmic scale) as a function of angle for the Spike-MR and Coral configuration (top) and the Spike-MR and tablet configuration (bottom).

Table 3. Average and SD results for the DSSQ by round (highest possible scores are 28, 28, and 24, respectively)

Round	Engagement	Distress	Worry
0	24 (3)	8 (4)	6 (5)
1	24 (4)	7 (4)	4 (5)
2	23 (5)	7 (6)	3 (4)
3	23 (4)	7 (4)	3 (3)
Average	**24 (4)**	**7 (4)**	**4 (4)**

The three common coordination items (i.e.; “It was necessary to use verbal communication to acquire the target”, “We worked in specific sequential order”, and “The verbal communication was based on still images”, note: the last item was relevant only within the still images comm. means) aimed at evaluating the advantages (or disadvantages) of the different communication means and the teams’ working techniques.

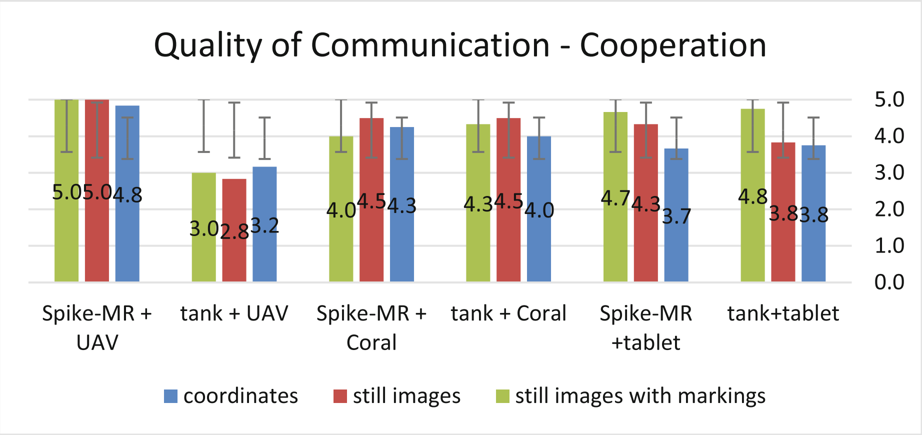


Fig. 6. Quality of Communication - Average evaluations of the cooperation items (N = 3) per teams and communication means

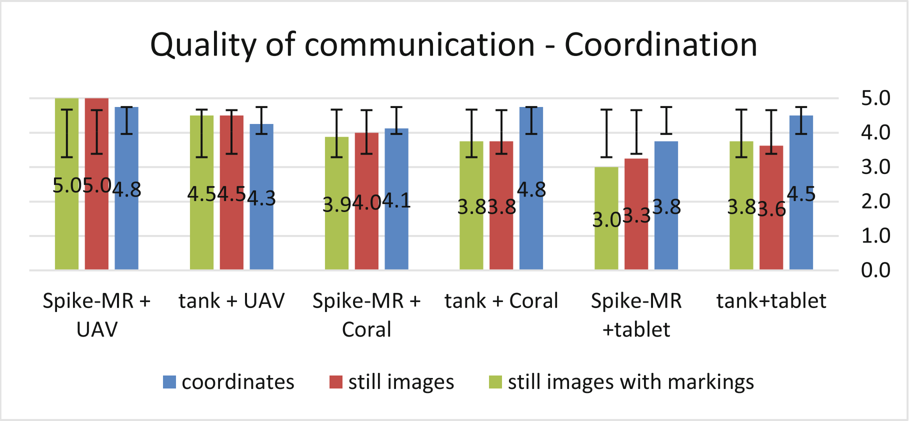
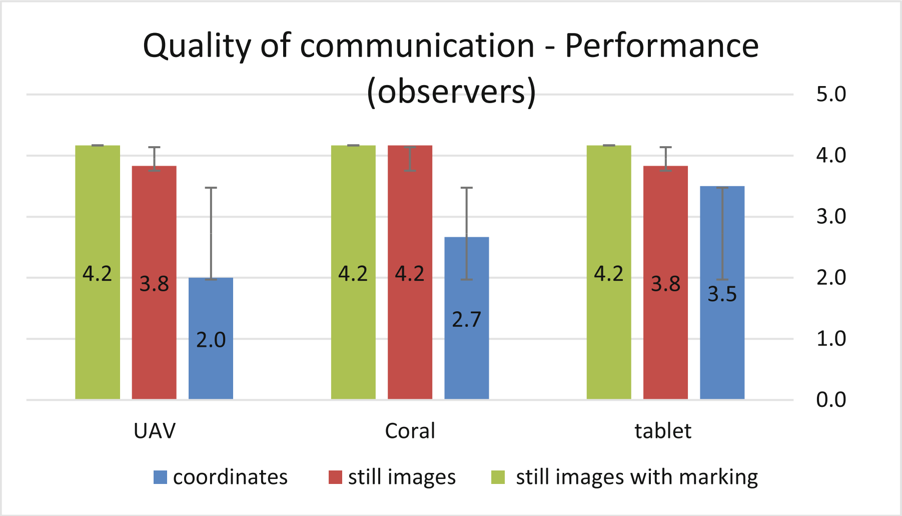


Fig. 7. Quality of Communication - Average evaluations of the coordination items (N = 3) per teams and communication means

Figure 7 presents the different patterns of evaluation among the different teams and communication means.

Participants scored low (Average = 1.2, STD = 0.9) on the fourth coordination item (attacker version: "I was overloaded and couldn't use all still images I received", and observer version: "I felt that the attacker did not use the still images I sent but trying to "figure out" by himself") indicating they favored the still images communication. The average of all 6 common items across all teams and including coordinates and still images (with and without markings) communication means was 4.1 (on a 5



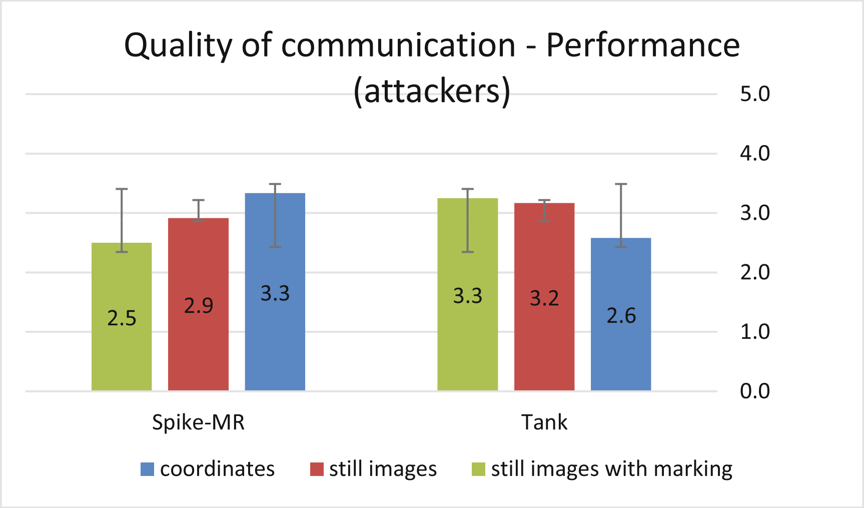


Fig. 8. Quality of Communication - Average evaluations of the performance items (N = 2) per observers (top) and (N = 3) per attackers (bottom), and communication means.

Likert scale), which means that the participants were highly coordinated and managed to create unique communication.

As for the self-performance evaluation, the analysis was done for the observers and the attackers separately. The performance items in the observer version were: “I was able to understand the attacker point of view” and “I was able to instruct the attacker based on his point of view”. The results show that the observers preferred the still images (either with or without markings) compared to coordinates (See Fig. 8 top). The average of all observers across all performance items was 3.6 (STD = 0.8). The performance items of the attackers which were included in the analysis (i.e., “It was difficult for me to acquire the target's surrounding", "It was difficult for me to acquire the target itself", and "I felt confident with the target acquisition") had an average of 3 (STD = 0.4) across all attackers. See Fig. 8 bottom for the detailed results. The attackers' scores on the fourth performance item - "I felt confident attacking the target based on pictures", which was relevant only for the still images (with and without markings) communication means, were high (Average = 4.3, STD = 1) with no differences between the still images only and the still images with markings.

3.3 Verbal Communication

The verbal communication channel between the attackers and the observers was available throughout the experiment. The data was measured by the total number of 'ping-pong' transmissions between the attackers and the observers and by the percent of time verbal communication was in use. Total of 5702 'ping-pongs' took place while 56% within the Spike-MR attacker and 44% by the Tank. In addition, 39% within the Coral, 37% via tablet and 24% by the UAS. See Figs. 9 and 10 for the detailed results and the different patterns among the various teams and communication means.

The controllers were asked to scale their impression of the necessity of verbal communication between the observers and the attackers for target acquisition. The results are shown in Fig. 11.

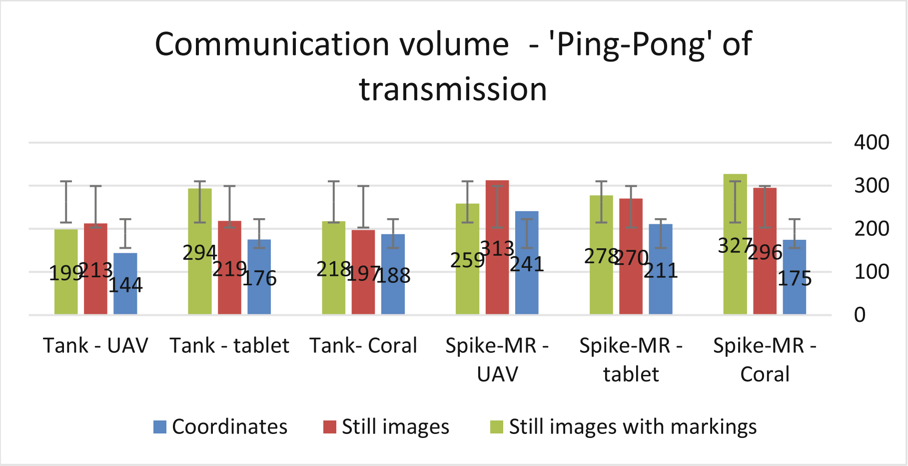


Fig. 9. Communication volume as measure by number of 'Ping-Pong' transmissions per team (attacker-observer) and communication means

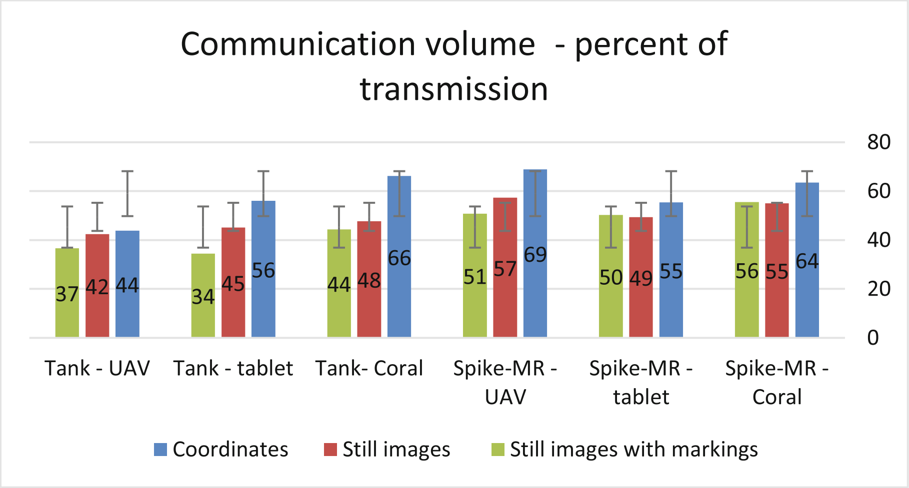


Fig. 10. Communication volume as measure by percent of transmission time per team (attacker-observer) and communication means

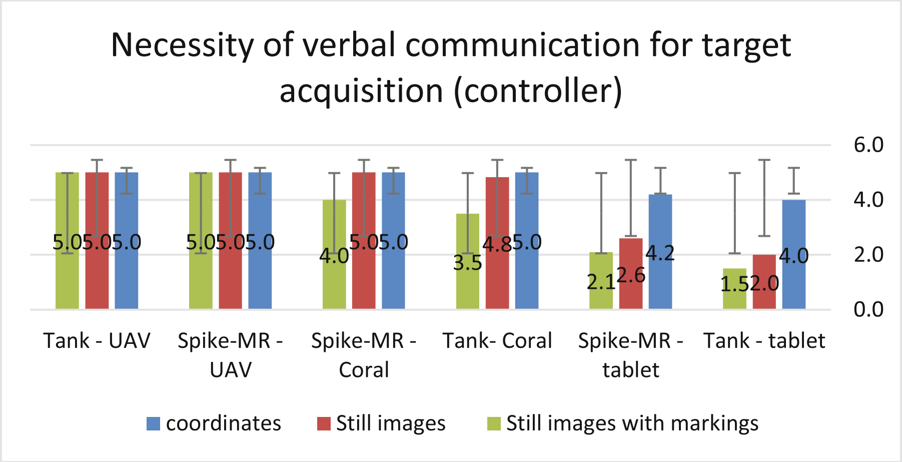


Fig. 11. Necessity of verbal communication - Average of the controllers' evaluations per communication means

4 Conclusions

In this study, we examined the added value of using one-way still images graphic communication between a ground soldier (represented by Tank (mounted) or Spike-MR (dismounted) attackers) and an operator (represented by Coral, tablet or UAS observers). The abilities to communicate by using still images with or without adding markings were compared to a basic communication based on coordinates and a future solution ('location on reality' – augmentation of target location on reality).

Objective and subjective measures were obtained. From those, it seems that in the ground-ground teams, operators benefited from the use of still images especially in the time to acquire a target and the accuracy distance (see Tables 1 and 2 and Figs. 3 and 4). In contrary, still images from the UAS were less beneficial to the ground operators relative to coordinates in terms of time to acquire a target and executions. The use of still images may improve if: a) bi-directional graphic communication will allow operators on both sides to send or annotate sent images, and b) if annotations were structured to enhance situation awareness, for example by using a fixed set of tools or annotation symbols (e.g., Granlund et al. 2011). Overall, observer-type participants favored the use of still images in the free choice round #3, they chose to use the still image configurations almost in all cases (98%).

Communication within a team was verbal (two-way) at all times and graphic (one-way) communication when applicable. As noted from Figs. 6, 7, 8, 9, 10 and 11, there were differences in the way the observers and the attackers perceived the quality of communication, and there were differences, between the mounted and the dismounted attackers as well. All in all, it seems that the addition of graphic communication increased the number of ping-pong chats between team members, but on average, shortened the duration of the chat. From the perspective of the controllers, who were supervising the teams, it seems that the Ground-Ground teams needed less verbal communication compared to the Ground-Aerial teams. Furthermore, less verbal communication was required when using the still images (with or without markings) interface compared to coordinates.

With regard to our hypotheses we can conclude that; (1) Still images may improve the communication between Ground-Ground teams; (2) In some cases, adding the ability to graphically mark elements on the still image (annotations) improved the communication in terms of self-evaluation of team cooperation and performance, even more than still images alone; (3) In presence of graphic communication the verbal communication patterns have changed with more ‘ping-pong’ transmissions between team members although shorter ones; and (4) when participants had a free choice of communication means they preferred the still images with graphic markings on the other three alternatives (coordinates, augmentation on reality and still images without graphic markings). These findings are consistent with our previous communication study (Oron-Gilad and Oppenheim, Final research report 2015) in few aspects:

- The use of rather simple communication means like temporary markings on the video feed in the previous experiment, and still images in the current one, was beneficial, at least in ground-ground settings. Which is why it is important to continue and examine the use of simple means of graphic communication that are simpler to implement.
- It is important to develop a structured interface which emulates the relevant available information, and to create a common terminology in order to simplify the graphic communication and allow convenient utilization of the new tools.
- The results confirm the inherent differences between Ground-Ground communication needs and those of Aerial-Ground teams in terms of different perspectives and shared views. This is not a new problem (See Oron-Gilad et al. 2011; Ophir-Arbelle et al. 2012 for example) but it requires attention.

Acknowledgments. This work was supported by the US Army Research Laboratory through the DCS subcontract no: APX03-S010 (B.G. Negev Technologies and Applications Ltd) under Prime Contract no W911NF-10-D-0002, Michael Barnes, Technical Monitor. This work was done in collaboration with the IDF Battle lab, Lt. Col Yaniv Minkov, Major Yisachar Shapira and Major Yoav Yulis. The work is unclassified, Approved for public release.

References

Artman, H., Granlund, R.: Team situation awareness using graphical or textual databases in dynamic decision making. In: Proceedings of the 9th Conference on Cognitive Ergonomics, pp. 149–155, August 1998

Baber, C., Fulthorpe, C., Houghton, R.J.: Supporting naturalistic decision making through location-based photography: a study of simulated military reconnaissance. Int. J. Hum.-Comput. Interact. **26**(2–3), 147–172 (2010)

Brehmer, B.: The dynamic OODA loop: amalgamating Boyd's OODA loop and the cybernetic approach to command and control. In Proceedings of the 10th International Command and Control Research Technology Symposium (2005)

Brooke, J.: System Usability Scale (SUS). Usability Evaluation in Industry. Taylor and Francis, London (1996)

Cox, D.: Regression models and life tables. J. R. Stat. Soc. **34**, 187–220 (1972)

Cumming, M., Akar, E.: The representation of common ground and its role in P2P-supported design team processes. Embracing Complex. Des. **23** (2005)

Fiore, S.M., Salas, E., Cuevas, H.M., Bowers, C.A.: Distributed coordination space: toward a theory of distributed team process and performance. Theoret. Issues Ergon. Sci. **4**(3–4), 340–364 (2003)

Fussell, S.R., Setlock, L.D., Kraut, R.E.: Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 513–520. ACM (2003)

Granlund, R., Smith, K., Granlund, H.: C3Conflict a simulation environment for studying teamwork in command and control. Naval Postgraduate School, Monterey, CA, Dept. of Operations Research (2011)

Hew, P.: Structured, Graphical Analysis of C2 Teams and their Technologies. Defence Science and Technology Organisation Canberra (Australia) (2011)

Klein, G.A., Crandall, B.W.: The role of mental simulation in naturalistic decision making. In: Hancock, P.A., Flach, J., Caird, J., Vicente, K. (eds.) Local Applications of the Ecological Approach to Human-Machine Systems, vol. 2, pp. 324–358. Lawrence Erlbaum, Hillsdale (1995)

Kraut, R.E., Fussell, S.R., Brennan, S.E., Siegel, J.: Understanding effects of proximity on collaboration: implications for technologies to support remote collaborative work. In: Distributed Work, pp. 137–162 (2002)

Matthews, G., Joyner, L., Gilliland, K., Campbell, S.E., Falconer, S., Huggins, J.: Validation of a comprehensive stress state questionnaire: towards a state big three. Pers. Psychol. Europe **7**, 335–350 (1999)

Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Warm, J.S.: Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. Emotion **2**(4), 315 (2002)

Mosier, K.L., Fischer, U.M.: Judgment and decision making by individuals and teams: issues, models, and applications. Rev. Hum. Factors Ergon. **6**(1), 198–256 (2010)

Ntuen, C.A., Park, E.H., Gwang-Myung, K.: Designing an information visualization tool for sensemaking. Int. J. Hum.-Comput. Interact. **26**(2–3), 189–205 (2010)

Obradovich, J.H., Smith, P.J.: Design concepts for virtual work systems. In: The Handbook of High Performance Virtual Teams, pp. 294–328 (2008)

Ophir-Arbelle, R., Oron-Gilad, T., Borowsky, A., Parmet, Y.: Is more information better? How dismounted soldiers utilize video feed from unmanned vehicles - attention allocation and information extraction considerations. J. Cogn. Ergon. Decis. Making **7**(1), 26–48 (2012)

Oron-Gilad, T., Oppenheim I.: Final Research Report - Task (5) Bi-directional graphic communication between a dismounted commander and the operators of a remotely operated vehicle, Technical report via DCS, Michael Barnes technical monitor, US Army Research Laboratory, December 2015

Oron-Gilad, T., Parmet, Y.: Close target reconnaissance: a field evaluation of dismounted soldiers utilizing video feed from an unmanned ground vehicle in patrol missions. J. Cogn. Ergon. Decis. Making (2016). Available online 1st November 2016

Oron-Gilad, T., Redden, E., Minkov, Y.: Robotic displays for dismounted warfighters: a field study. J. Cogn. Ergon. Decis. Making **5**(1), 29–54 (2011)

Parmet, Y., Meir, A., Borowsky, A.: Survival analysis: a fresh approach for analyzing response times in driving-related hazard perception paradigms. Transp. Res. Part F: Traffic Psychol. Behav. **25**, 98–107 (2014)

Walker, G.H., Stanton, N.A., Jenkins, D.P., Salmon, P.M., Rafferty, L.: From the 6 Ps of planning to the 4 Ds of digitization: difficulties, dilemmas, and defective decision making. Int. J. Hum.-Comput. Interact. **26**(2–3), 173–188 (2010)

Controller Intervention Degree Evaluation of Intersection in Terminal Airspace

Yannan Qi, Xinglong Wang(✉), and Xingjian Zhang

Civil Aviation of University of China, Tianjin 300300, China
1545186632@qq.com, xinglong1979@163.com,
jzhyoung2006@163.com

Abstract. The intersections in the terminal airspace is an potential factor to cause flight conflicts. The possibility of conflict can be reflected directly from air controller intervention degree. Analyzing the controller intervention degree of intersection is a precise and efficient method to locate airspace congestion location in terminal airspace. After analyzing the historical data, improved density clustering method is used to determine the distribution of intersections in terminal airspace. The method of assessing the controller intervention degree is put forward through the statistics on the changes of the flight elements within certain time and range. The experimental results show that this method can obtain the distribution of conflict and controller pressure points in the terminal area, and is quite consistent with reality. The method can provide fast and accurate optimization basis and improvement direction for the airspace planning work.

Keywords: Density based clustering method · Controller intervention degree · Intersection

1 Introduction

Intersection is an important factor increasing workload for controllers, especially in terminal airspace. Controller intervention degree demonstrate the potential conflicts and complexity of airspace which illustrate the workload of air traffic controller directly. Assessment of controller intervention degree is a good method to understand the complexity of local airspace and distribution of congestion points. Currently, the goal of intervention degree assessment is to evaluate the air traffic controller workload. Simulation and modelling are the common used methods. However, those methods are time consuming and simulation software is expensive. This paper introduces a method to evaluate the controller intervention degree on the basis of analyzing the history flight data. According to controller experience in terminal airspace, the direct result of controller intervention is the changes of flight elements (heading, speed, altitude). Surveillance data is collected and analyzed by adopting improved density clustering to dig out the intersections with the help of improved density-based clustering algorithm, the controller intervention degree of intersection is evaluated by detecting the changes of flight elements.

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 382–393, 2017.
DOI: 10.1007/978-3-319-58472-0_29

2 Determination of Intersection Distribution Based on Improved Density-Based Clustering

2.1 Problem Description and Introduction of Improved Density-Based Clustering

The basis for aircraft operating in terminal airspace is the flight procedure. However, deviation happens in actual flight because of the conflict resolution. So, the history flight data is mess, huge and uneven density. Some method should be used to identify prevailing traffic flow to obtain the location of intersections. Density clustering algorithm is the ideal way to obtain the prevailing flow and deal with data sample with arbitrary shape. So, we improved the algorithm and used it to get the location of intersections in this paper.

In order to enhance the efficiency of algorithm, improved density clustering algorithms emerged. Among them, shared nearest neighbor clustering algorithm(SNN) is not susceptible to input parameters, but is suitable to use on uniform density data set. As for this problem, improved SNN clustering algorithm is proposed in this paper. Firstly, sub-cluster of each point is generated and the density of each point is calculated. Then, with the help of optimal two-partition method, an density array of the whole data set is sequenced by grouping data with similar density into same category. Each sub-cluster is assigned a density category tag which demonstrate. Finally, SNN algorithm is used in sub dataset with same category tag to generate the average trajectory.

As air traffic surveillance data is used for experiment, one trajectory is composed with many points, each point can be expressed with a vector pi = $[x, y, z, v, \alpha, t, t_r]$, x, y *is* the latitude and longitude of point p_i, z is the altitude, v is ground speed, $\alpha$is the heading, t is the collecting time of p_i, t_r is the trajectory number of which pi belongs to.

2.2 Generation of Sub-clusters

Sub-cluster is a set [*R*] of data points that have a certain similarity relation with a certain data point. Traffic flow is a dataset of points with certain density and direction. So, the sub cluster is defined as a dataset with distance between points is within the certain scope. The similarity of trajectory points is evaluated with Euclidean distance.

Define sub-cluster of pi($C(p_i)$) as the data set contains points with distance to *pi* is less than *r*:

$$C(p_i) = \left\{ p_j \middle| \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq r, 1 \leq j \leq n \right\} \tag{1}$$

Where p_i, p_j are trajectory points, x_i, y_i is the coordinates of p_i, x_j, y_j is the coordinates of p_j, n is the number of points.

In order to reduce the amount of computation and improve the efficiency of algorithm, the search space could be limited to calculate the distance between points.

Trajectory point vector contains the latitude, longitude and heading, so the research space could be limited in a neighborhood $U(p_i)$ of the point.

$$U(p_i) = \left\{ p_j \in S \,\middle|\, \begin{matrix} \left(x_i - \frac{L\times\cos\alpha}{111}\right) < x_j < \left(x_i + \frac{L\times\cos\alpha}{111}\right) \\ \left(y_i - \frac{L\times\sin\alpha\times\cos\alpha}{111}\right) < x_j < \left(x_i + \frac{L\times\cos\alpha\times\sin\alpha}{111}\right) \end{matrix} \right\} \tag{2}$$

Where, S is the trajectory points set, L is the limited scope of neighborhood, which can be determined by the aircraft speed limit and data collection frequency in the terminal area. Under the condition where r is determined, we use $L = r$.

2.2.1 Density Distribution Array Generation

The sub-clusters are constructed from each data point in the dataset and the density of the sub-clusters is analyzed. For the data points with the same radius r, the number of track points is larger, indicating that the local density of the sub-clusters is larger. On the contrary, if the number of track points of sub-clusters is similar, their density should be basically the same.

Define the sub-cluster tag: Each sub-cluster contains a certain number of data points, recorded as d_{ps}. According to the magnitude of d_{ps}, sub-cluster are divided into s categories $(s \in Z^+)$. For a given category S_g, set its density tag as $Cat_g(1 \leq g \leq s)$. For any given sub-cluster $Cs(p_i)(1 \leq i \leq n)$ and category S_g, where the number of data points contained in $Cs(p_i)$ is d_{psi}, the density of sub-cluster $Cs(p_i)$ marked as Cat_g.

All the sub-clusters is classified into density distribution array C_{at}, which determines the density distribution of the whole data set. It is large possibility to cluster the points with similar density to same cluster. Sub-clusters with similar density are classified as same category of density distribution. How to measure the similarity of the sub-clusters density is a problem to be considered in clustering. the number of categories is small, the density distribution may be assessed incompletely, and the number of categories is too big, the density distribution is too fragmented and the clustering efficiency is low. The optimal two-partition method is used to divide the sub-clusters into appropriate number of categories. First of all, sub-clusters are sequenced by d_{sp}. The optimal two-partition is to subdivide the sub-clusters according to d_{ps} into two parts at once. The minimum total variance is target for each division, and the partitioning is carried out several times. Minimum number of categories is corresponding to overall variance mapping inflection point with the minimum total variance as target is determined (Fig. 1).

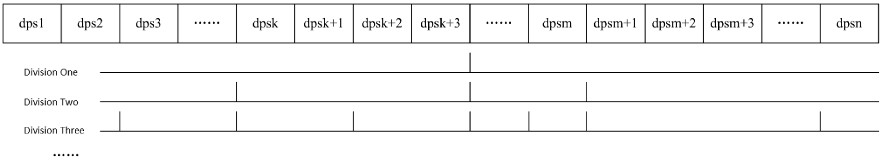


Fig. 1. Density array generation based on the optimal two-division method

At the beginning of division, the overall variance decreases fast. After reaching to a certain extent, the trend of reducing the overall variance is getting smaller and smaller. When the number of partitions is the same as the number of sub-clusters, the overall variance is zero. According to this law, the reflection point of the function of the category number and overall variance can be determined, and the reflection point is the optimal division.

2.2.2 Clustering of Sub-clustering with Same Density Tag

After marking each sub-cluster the density tag, we can cluster the sub-clusters with similar density using SNN algorithm to get the average trajectory. Sub-clusters with sparse points may be clustered into a large class without considering the direction, which does not happen in reality. So the same density category sub-clusters can be clustered according to direction into several clusters, because it is easy to separate different direction of traffic flows.

Assume ARR[*i*] is the sub-cluster set with density category is *i*, ARR[*i*].P[*j*] is the center of sub-cluster *j* with density category is *i*. The algorithm is as follows:

- Step1: judge whether ARR[*i*] is empty, if true, clustering ends, otherwise turn to Step2;
- Step2: judge whether any cluster set with same density tag is empty, if empty, $i = i + 1$, turn to step1, otherwise turn to step3;
- Step3: randomly select the sub-cluster with same density tag, insert the point whose distance with centre point ARR[*i*].P[*j*] is less than r′ to the pre-alignment queue Q, and mark the point. If marked number is 0, turn to step2, a new class is produced, otherwise turn to step4;
- Step4: judge whether Q is empty, if empty, $j = j + 1$, turn to Step2, otherwise get a point(marked as *tp*) from Q, labelled class tag, turn to step5;
- Step5: judge whether the density of *tp* is equal to ARR[*i*].Cat or ARR[$i + 1$].Cat, if true, add the points in sub-cluster of tp into Q, and delete *tp* from Q, turn to Step2; otherwise, delete *tp* from Q and unmark the *tp*, turn to step2.

Where, r' should be greater than the original r used in calculation of sub-cluster. When the density is relatively larger, r' does not work, the point number in a cluster with same density tag is decided by r' only the density is small.

3 Locate Intersections

The directional trajectory flow is formed after adopting the above algorithm. In order to determine the location of intersection, average trajectory should be generated based on the outcome the clustering. Along the direction of trajectory flow, the trajectory flow is divided into several equal length segment as a step d. The division method is: starting from a point fp_0 closest to the runway reference point, determine the center point of the segment where fp_0 belongs to, and calculate the mean value of track angle $\overline{A}$. Then, a division line is generated through the center point, the equation of division line is: $F_d(x) = kx + b$, where the slope k is the tangent of the angle between the vertical direction of $\overline{A}$ and the X axis (Fig. 2).

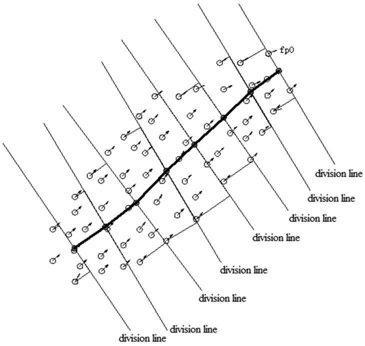


Fig. 2. The average trajectory generation method

As there are two flight status in terminal airspace, straight and turning flight. Track angle adjustment is greater than 15° is recognized as turn. For straight flight trajectory, the average trajectory is not effected by the step size *d*. However, as for turn, larger *d* will cause trajectory with sharp turn, this is quite different with actual flight, and the accurate of further analysis will be effected. So, the step size is effected by turn rate and bank angle.

Assume lv is speed limitation, α is bank angle, determine d as:

$$d = \frac{(0.321lv^2 \tan 15^\circ)}{\pi \tan \alpha} \tag{3}$$

In order to solve conflict in terminal airspace, the speed of aircraft is limited by ATC, so the speed difference of aircraft is limited, fixed value of lv is appropriate in this paper.

Points between two division lines is a segment flow, all points in each segment flow project to initial division line. Projection points defined as $[x_{pi}, y_{pi}] = prj_{Fd}(x_i, y_j)$, so the scope of projection points is determined: $[(x_{pl}, y_{pl}), (x_{pr}, y_{pr})]$, the center point $\left[\frac{(x_{pl}+x_{pr})}{2}, \frac{(y_{pl}+y_{pr})}{2}\right]$ is chosen as the average trajectory point in current segment. An average trajectory of a trajectory class is generated by connecting each average points by sequence. So, the intersections are determined.

4 Assessment of Controller Intervention Degree on Intersection

Climbing and descending are the prominent status in terminal airspace. In order to solve the conflicts, intervention instruction is issued before aircraft approaching to intersections, so the flight elements change happen in a scope outside of intersection. In this paper, controller intervention degree is assessed through calculating the statistical probability of each flight elements in certain scope of intersection.

According to standard separation minima in terminal airspace under radar control condition:

(1) vertical separation minima $Sep_v = 300$ m,
(2) lateral separation minima $Sep_{lat} = 6000$ m.

There are two traffic scenarios if there is conflict in certain scope of intersection.

(1) Two aircraft *A* and *B* from two different traffic flows converge to *P* (Fig. 3(a)). Under this circumstance, controller will justify the possibility of conflict in advance. Considering worst circumstance, aircraft *B* has reached at *P*, but aircraft *A* still maintains original flight state. When aircraft *A* enters into the area where the protection areas of two aircraft are overlapped, conflict happens.

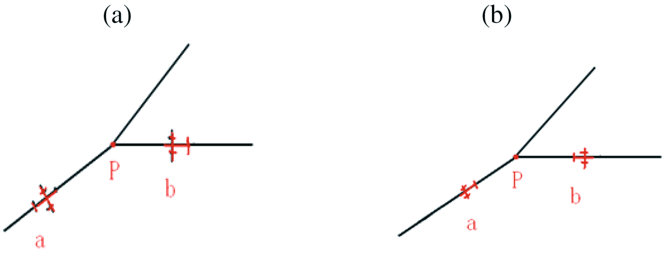


Fig. 3. (a) Scenario of convergence of flight (b) scenario of crossing after passing

(2) Two aircraft *A* and *B* crossing after passing intersection *P* (Fig. 3(b)). Still consider worst circumstance, aircraft *B* has reached to *P*, aircraft *A* still maintain original flight state. So the aircraft *A* is still in the area where the protection areas are overlapped, conflict still exists.

So, the neighborhood of *P* is defined as the scope of assessment of controller intervention degree.

$$\varepsilon_p = \left\{ (x, y, z) \in A \middle| \frac{(x - x_p)^2}{sep_{lat}} + \frac{(y - y_p)^2}{sep_{lat}} + \frac{(z - z_p)^2}{sep_v} = 1 \right\} \tag{4}$$

Where x_p, y_p, z_p is the coordinate and altitude of intersection *P*.

4.1 Assessment of Controller Intervention Degree

The change of flight elements indicates the happen of controller intervention, then define P_c as intervention degree:

$$p_c = \frac{\sum_{i=1}^{n} [f_A(p_i)|f_H(p_i)|f_V(p_i)]}{n} \tag{5}$$

Where, n is the number of trajectory points in certain scope of intersection, p_i is the *i-th* trajectory points.

$$f_A(p_i) = \begin{cases} 1 & Altitude \quad change \quad is \quad true \\ 0 & Altitude \quad change \quad is \quad false \end{cases}$$

$$f_H(p_i) = \begin{cases} 1 & Heading \quad change \quad is \quad true \\ 0 & Heading \quad change \quad is \quad false \end{cases}$$

$$f_v(p_i) = \begin{cases} 1 & Speed \quad change \quad is \quad true \\ 0 & Speed \quad change \quad is \quad false \end{cases}$$

The statistical methods for the change of flight elements are described below.

4.2 Statistical Probability for the Change of Flight Elements

Constructing Cartesian Coordinate System in Spatial Range of airspace, take neighborhood ε (Eq. (4)) as the conflicts statistics scope of an intersection, we call it as overlap area.

As the civil aircraft has certain stability, flight elements do not change immediately after maneuvering instruction. And in actual flight, the aircraft is often affected by the air flow which will cause slight disturbance. So in order to accurately detect the change of flight elements, the status change time should be analyzed.

According to the maneuvering delay time from ICAO PANs-OPS, in arrival flight segment, pilot reaction delay is 0–6 s, operating delay is 0–5 s. In departure segment, pilot reaction delay is 0–3 s, operating delay is 0–3 s. the data is relatively conservative, and autopilot is the main method in actual flight. So, 4 s flight reaction delay and 5 s maneuvering delay are adopted in this paper as the basis for detect the changes of flight elements.

4.2.1 Detection of Altitude Change

Points appearing in the overlap area are extracted by the flight number and stored separately. Each point sequenced according to time. Take $t = 3$ as change judgment step size. Assume $t1$ is the earliest time enter into the overlap area, $t2$ is the latest time enter into the overlap area. Assume $A(t')$ is the point altitude at time t'.

Check each flight points from $t1$ to $t2$:

(1) when $\mathrm{A}(t_1) \neq \mathrm{A}(t_1 + nt) \cdot (n = 1, 2, 3\ldots\ldots)$, let $t_1^{'} = t_1 + nt$, keep checking, until $\mathrm{A}(t_1^{'} + nt) = \mathrm{A}[t_1^{'} + (n+1)t]$, then let$t_2^{'} = t_1^{'} + nt$, or

(2) when $\mathrm{A}(t_1) = \mathrm{A}(t_1 + nt)\,.(n = 1, 2, 3\ldots\ldots)$, let $t_1^{'} = t_1 + nt$, keep checking, until $\mathrm{A}(t_1^{'} + nt) \neq \mathrm{A}[t_1^{'} + (n+1)t]$, then let $t_2 = t_1^{'} + nt$.

If $t_2^{'} - t_1^{'} \leq 6\mathrm{s}$, the altitude change is neglected; if $t_2^{'} - t_1^{'} \leq 6\mathrm{s}$, calculate the inequality (6).

$$A(t_2^{'}) - A(t_1^{'}) > \left| \frac{1000}{60 \times (t_2^{'} - t_1^{'} - 9)} \right| \tag{6}$$

If the inequality is true, altitude change happened in specific time period, and $f_A(p_i) = 1\,\mathrm{f_A(p_i)}$ otherwise $f_A(p_i) = 0$.

4.2.2 Detection of Heading Change

Assume point $i(x_i, y_i)$ is the earliest point where certain flight enter into the overlap area, point $j(x_j, y_j)$ is the location of point of the flight. Great circle of track angle (α) is calculated:

$$\cot\alpha = \cos(x_i)\tan(x_j)\csc(y_j - y_i) - \sin(x_i)\cot(y_j - y_i) \tag{7}$$

Check each flight points from $t1$ to $t2$. When $\alpha(t_1) \neq \alpha(t_1 + nt).(\mathrm{n} = 1, 2, 3\ldots\ldots)$时, let $t_1^{'} = t_1 + nt$. keep checking, until $\alpha(t_1^{'} + nt) = \alpha[t_1^{'} + (n+1)t)]$, then let $t_2^{'} = t_1^{'} + nt$. If $t_2^{'} - t_1^{'} \leq 6\mathrm{s}$, the change is neglected; if $t_2^{'} - t_1^{'} > 6\mathrm{s}$, calculate the inequality (8).

$$[\alpha(t_2^{'}) - \alpha(t_1^{'})] > 3 \times (t_2^{'} - t_1^{'} - 9) \tag{8}$$

If the inequality is true, heading change happened in specific time period, and $f_H(p_i) = 1\,\mathrm{f_A(p_i)}$ otherwise $f_H(p_i) = 0$.

4.2.3 Detection of Speed Change

In actual flight, flight speed is generally adjusted for every 10 knots, however, considering the change of airflow, speed of change may not be an integer multiple of 10. Troposphere is the main area for operation of aircraft, and the average speed of wind is 12.2 m/s ≈ 2knot, so we consider 8 knots as the change of speed.

Assume $C(t^{'})$ is the speed of point at time $\mathrm{t}^{'}$. Check each flight points from $t1$ to $t2$, when speed of aircraft $C(t_1) \neq \mathrm{C}(t_1 + \mathrm{nt})\,(n = 1, 2, 3\ldots\ldots)$, let $t_1^{'} = t_1 + nt$. Keep checking untile $C(t_1^{'} + \mathrm{nt}) = C[t_1^{'} + (n+1)t)]$, then let $t_2^{'} = t_2^{'} + nt$. Calculate the inequality (9).

$$C(t_2^{'}) - C(t_1^{'}) < 8 \tag{9}$$

If the inequality is true, speed change happened in specific time period, and $f_v(p_i) = 1\,\mathrm{f_A(p_i)}$, otherwise $f_v(p_i) = 0$.

5 Experimental Evaluation

In order to verify the evaluation method, 24 h ADS-B surveillance data of Tianjin airport is used to assess the controller intervention degree. Data collection step is 3 min, coordinates of trajectory points is correspond to the WGS-84 system.

5.1 Generation of Density Distribution Array

Sub-cluster of each trajectory point should be generated, before that the radius r of density distribution is determined. The size of the radius r affects the speed and the result of clustering. In some literatures, the k neighborhood algorithm is used to generate the distance d_i from the k-nearest neighborhood for each points, the r is the average of all d_i. However, large amount of data is used for analysis, and whether or not r is appropriate depends on the value of k. so as to improve it, r is determined according to speed limitation and data collection frequency in this paper. Assume the speed limitation is $V_{\max}$, t is the step size of data collection, r is determined as: $r = V_{\max} \times t$.

If t is too small, it is difficult to identify the isolated point. On the contrary, if t is too large, it is unfavorable for further clustering. In this case, we use $t = 5s$ to generate sub-cluster for each point. Trajectory points are sorted by dps_i of sub-cluster for each point. In Tianjin airport, maximum IAS is 210kt, so we use $r = 540$ m to generate sub-cluster. The optimal two-division method described in 2.2 is used to category each point. The relationship between number of segments and total variance is showed in Fig. 4.

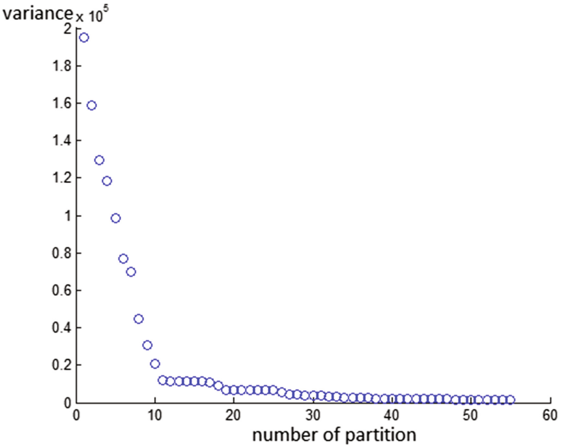


Fig. 4. Relationship between number of partition and variance

In Fig. 4, we can clearly see that the inflection point of total variance corresponds to number of segments is 11. Tag each point according to division result. In this experiment, number of elements contained in sub-cluster of each point evenly distributed from 0 to 1089. Because the existence of error happens in ADS-B data collection and analysis, noisy data should be deleted to prevent the effect to the outcome. So, points with small number (0–10) contained in sub-cluster is deleted as noise. The final segment number is 7, the density tag is 1–7, 7 means highest density.

5.2 Intersection Determination

Maximum IAS is 210kt, $lv = 390$, bank angle $\alpha = 15$ is used to generate the average trajectory. The average trajectory is showed in Fig. 5. Where, blue represents departure, red represents arrival. P1, P2, P3 and P4 are determined intersections.

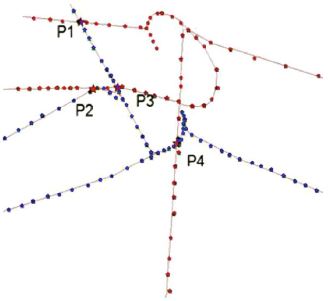


Fig. 5. Average trajectory and intersection distribution (Color figure online)

5.3 Result of Controller Intervention Assessment

The controller intervention assessment result is illustrated in Table 1.

Table 1. Statistics of cross point navigation elements changes and result of intervention degree

Intersection	Phase	No. of points	No. of altitude change	No. of speed change	No. of heading change	Controller intervention degree	Inter section	Phase	No. of points	No. of altitude change	No. of speed change	No. of heading change	Controller intervention degree
P4	Arrival	56	2	6	5	11.24%	P3	Arrival	35	5	7	1	31.47%
	Departure	155	1	1	10			Departure	65	3	2	15	
	Total	211	2	7	15			Total	100	8	9	16	
P2	Arrival	141	5	9	17	13.67%	P1	Arrival	77	1	3	1	10.34%
	Departure	201	3	1	14			Departure	37	4	1	3	
	Total	342	8	10	31			Total	114	5	4	4	

It can be seen from Fig. 5, P2 and P3 is closer than with other intersections, and the distance along same track is less than lateral minima. So, to aircraft with same flight path, enter into the overlap area of P2 also effect the aircraft around P3. However, it is impossible to know from historical data whether p2 or p3 is it's conflict point. And the time when it enter into the overlap area of is also unknown. In this paper, we consider a point between two intersections as the equivalent intersection, so that the intervention degree of the new intersection P23 is approximately equal to the average intervention degree of P2 and P3. Overlap area of P23 will be equivalent to the overlap area of P2 and P3 for aircraft with P2 or P3 as the direct collision point. This can be as a control reference in actual work of controller. The location of P23 can be located: because P2 and P3 are on the same path, so according to the intervention degree of P2 and P3, the location of P23 is on the proportional position of intervention degree of P2 and P3.

5.4 Result Analysis

In experimental airports, due to airspace restrictions, nearly 70% traffic concentrate in the west of airport, and much intervention is needed to solve the conflict, so the intervention degree is larger. The result (Table 1) shows consistency with actual operation. P2 and P3 are conflict points of departure and arrival, intervention degree of P3 is larger because from P3 the departure path divided into two branches, and both of them have potential conflict with arrival flight on the following flight (P1 and P2). And this is the reason for smaller intervention degree of P1 and P2. P4 is close to runway, the altitude of departure aircraft is low, and the arrival aircraft is high, vertical separation is guaranteed automatically, so the intervention degree is small.

As the analysis above, P3 cause large workload for controller, this can be as a reference to airspace redesign.

6 Conclusion

In this paper, we have proposed a method to evaluate the controller intervention degree through analysing the historic surveillance data. Because the amount of data used is large, so improved density based clustering algorithm is adopted to generate the average flight trajectory. Here, trajectory clustering is divided into two steps. Because the density of flight data is uneven, the efficiency and accuracy of clustering will be effected seriously. So, sub-clusters are generated first to understand the overall density distribution of data sample, and sub-clusters are divided into several categories according to density. Optimal two- partition method is used to determine the number of categories. Sub-clusters with similar density is marked with same density tag and put into same category. Then, SNN algorithm is used to cluster sub-clusters with same density tag to generate average trajectory. The intersection of trajectory is the potential conflicts point which is used for controller intervention degree calculation. Controller intervention degree is defined as statistical probability of change of flight elements. So, for the purpose of statistic, the scope of intervention area is analysed and the detection method of flight elements changes is proposed. Finally, surveillance data of an airport is used to verify the method. The experimental results show that this method can obtain the distribution of conflict and controller pressure points in the terminal area, and the result is quite consistent with reality. The controller intervention degree can be used as a reference for further airspace designation.

References

1. Meyn, L.A., Erzberger, H.: Estimating controller intervention probabilities for optimized profile descent arrivals. In: 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, 20–22 September 2011. AIAA 2011-6802
2. Gariel, M., Srivastava, A., Feron, E.: Trajectory clustering and an application to airspace monitoring. Intell. Transp. Syst. **12**(4), 1511–1524 (2011)

3. Wang, C., Xu, X., Wang, F.: ATC serviceability analysis of terminal arrival procedures using trajectory clusting. J. Nanjing Univ. Aeronaut. Astronaut. **45**(1), 130–139 (2013)
4. ICAO: Procedure for air navigation services—aircraft operation construction of visual and instrument flight procedures. Doc8168-OPS/611, vol. II (2006)
5. Bilimoria, K.D., Lee, H.Q., Oaks, R.: Properties of air traffic conflicts for free and structured routing. In: Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference, 6–9 August 2001
6. Proceedings of the AIAA Modeling, Simulation and Technology Conference, Keystone, Colorado, 21–24 August 2006 (EIWAC), Tokyo, Japan, March 2009
7. Wang, C., Han, B.C., Wang, F.: Track traffic flow identification method based on track spectrum clustering in terminal area. J. Southwest Jiaotong Univ. **49**(3), 546–552 (2014)

Implementation of a Responsive Human Automation Interaction Concept for Task-Based-Guidance Systems

Georg Rudnick(✉) and Axel Schulte

Institute of Flight Systems, University of the Bundeswehr, Munich, Germany
`{georg.rudnick,axel.schulte}@unibw.de`

Abstract. This article presents the implementation of a task-based guidance system for reconnaissance UAVs enabling simplified human-automation interaction through responsive adjustment of details. It provides a way of communicating to the system on the necessary detail level, as well as status feedback in a level of detail currently wanted by the operator. This increases control and helps reducing workload.

The concept of Manned-Unmanned Teaming (MUM-T) is currently on the rise, as it allows force multiplication by higher levels of autonomy for UAVs, while at the same time offering control to a human pilot.

A drawback of MUM-T is the high workload, resulting in pilot to UAV ratios of greater than 1. Our approach adds to already available concepts to invert this ratio to allow one pilot to control multiple UAVs.

It combines task-based guidance, in which the pilot formulates mission goals as high level tasks and a cognitive agent aboard the UAV breaks them down in achievable subtasks and executes these, with two strategies for improving the communication between agent and pilot. The first is an adjustable level of detail for the information offered to the user. The second are precise warnings for conflicting situations or violations of the rules of engagement defined at the beginning of the mission.

In order to implement such abilities of the cognitive agent, a rule engine based situation analysis was combined with an HTN planning algorithm. This enables the use of task-based guidance, guaranteeing easy communication between the pilot and the agent. The system was tested in a simulation environment with military personnel, using downscaled UAV reconnaissance missions with a single mini-UAV. Mission goal achievement and questionnaires presented to the operators were analyzed for the evaluation. The system was rated positively.

Keywords: UAV · User interface · MUM-T

1 Introduction

Manned-Unmanned Teaming (MUM-T) allows force multiplication by teaming a manned mobile component (e.g. a helicopter) with unmanned units (e.g. UAVs). The resulting formation has the potential to be more effective, as dislocation of its members is possible and to be less vulnerable for casualties, as the unmanned members can

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 394–405, 2017.
DOI: 10.1007/978-3-319-58472-0_30

execute dangerous tasks. In contrast to a fully autonomous group a human pilot can control or intervene actions of unmanned members, which is especially useful in cases of automation failures or a contact with unprecedented situations. In addition, for armed autonomous systems the “Man in the Loop” principle [1] is much easier to implement, as the decision can be made on site with shorter reaction time and better tactical awareness. The greatest advantage is the possibility to achieve better tactical behavior of the team. Tactics and military stratagem are currently difficult to implement in automated systems. The human component is able to make such decisions on behalf of the unmanned members, increasing the team effectiveness even more.

A current research project at the Institute of Flight Systems involves improving the effectiveness of MUM-T concepts. A wing consisting of a transport helicopter with a two-man crew and up to three UAVs is tasked with search and rescue or transport mission. The commander is in command of the helicopter as well as the UAVs. This results in a great increase of workload for both pilots, as the commander is now occupied with UAV mission management and has only limited capacities to support the pilot flying. Several concepts are being developed to assist the crew in reducing the workload and achieving the mission goals. The topic of this work is the command and control interface between the commander and the UAVs.

Section 2 lists the work already done in this field. In Sect. 3 the concept is presented. Section 4 describes the implementation of the system. In Sect. 5 the evaluation method is presented, followed by the results in Sect. 6. Section 7 wraps up the findings.

2 Previous Work

Many publications exist on different aspects of human-machine interaction. In the case of this article, especially the workflow between user and machine is relevant. Sheridan [2] first coined the term of supervisory control for an automated process, which works on tasks assigned by a human operator. Such supervisory controlled systems can be converted to agent supervisory controlled systems by inserting an intelligent agent between the operator and the automated process [3]. This enables higher-level cognitive functions, like planning and situation assessment, in the system and leads to a reduction of workload for the operator. Uhrmann [4] created a system using the concept of task-based guidance, where a cognitive agent interprets tasks given by a pilot, breaks them down in achievable subtasks, and presents the pilot with the resulting plan for error checking and approval. The cognitive agent can then execute the plan independently.

Uhrmann showed that the reduction of workload can enable an inversion in the span of control of current reconnaissance UAVs. In this case instead of having multiple operators controlling one UAV, a single pilot commanded three UAVs from a two-seated helicopter cockpit while simultaneously flying in a transport mission [5].

One of the success factors is the simple communication from the pilot to the automation to achieve complex mission tasks. A few button presses can easily convey his or her intent to the agent, as it is aware of circumstances influencing the situation. This reduced communication in context rich environments is similar to human interactions.

The work was done in a simplified environment, for example sensor operation was reduced and the complexity of the tactical situation was restricted. The interactions between pilot and agent were also simplified. For example, the only status information available was the current tasks and a flag indicating if the agent was planning or not. The system therefore lacks a proper communication back to the pilot, as plans are either displayed in overwhelming detail or not detailed enough, which is inappropriate for most situations. In real environments, where the pilot is responsible for the actions of the unmanned aircraft, the right amount of information is necessary. Clauß [6] used the term of etiquette, coined by Miller [7], and applied it to agent supervisory control for single UAV systems in order to identify the relevant command interface and status information for such systems. Furthermore, unexpected situation changes, which create plan failures or require re-planning, are not communicated in a helpful or easy way [8]. This kind of specific feedback is necessary especially for multi-UAV operations, as the time to process the provided information and to formulate solutions is shared between all UAVs. To expand upon this work is the content of this article.

3 Concept

The term responsive design is widely used for web interfaces, which scale their content and format to the display space available [9]. Applying this idea to human-automation interaction results in a user interface, which scales the information presented to the user according to his needs and wants. Since the user interface represents a two-way communication, information from the user to the system, which in a UAS application is commands, should be included as well as the status information flow from the system to the user. Having such an interface and an underlying system able to understand this kind of input and provide output accordingly for each UAV, allows the pilot to focus on the details where it is necessary, while maintaining an overview of the situation. Key requirements of the system therefore should be:

1. Simple command interface for each UAV, conveying the pilot’s intent
2. Detailed command interface for each UAV, allowing more fine grained control when necessary
3. Overview of the current status of all UAVs and the tactical situation
4. Easily accessible detailed status information about each UAV
5. Proactive display of information about UAV conflicts
6. Detailed information about UAV conflicts if wanted by the pilot.

The idea is to combine several strategies to fulfill all requirements. Requirements (1) and (2) are achieved by providing a system with task-based guidance and scalable autonomy [10]. This allows using the same vocabulary for mission tasks and communicating the operator’s intent, thus forming a simple and usable command interface with the ability for more detailed control. Requirements (3) and (4) are achieved by providing a user interface, which allows to scale the amount of accessible information depending on the pilots’ needs, but at the same time is able to display an overview of the tactical situation. Requirement (3) is supplemented by a fixed part of the user interface, offering a status overview of all UAVs. Requirement (5) is achieved by directing the attention of the operator to emerging problems. This is done by a three-step warning system, with the ability to access more detailed information at will, fulfilling requirement (6).

4 Implementation

4.1 Cognitive Agent

For a system capable of task-based guidance the baseline is an intelligent software agent, which can assess the current situation and develop plans in order to execute assigned tasks. The implementation was done using the Drools Expert rule engine [11] for situation assessment and a SHOP [12] like HTN-planning algorithm [13] implemented inside Drools using the Drools Rule Language (DRL). When given a task, rules apply appropriate options from a collection of preprogrammed recipes. The recipes for subtasks, also called methods, are objects in the working memory of the rule engine, which have the target task type, as well as the subtasks to add, as attributes. For example, the “Recon” task is split into the two subtasks “Calculate Recon Route” and “Select Recon Route” by the “Recon”-method. As multiple results of the route calculation are possible, the “Select Recon Route”-alternative creates an alternative for each available route, resulting in several plan alternatives, as seen in Fig. 1. Alternatives are also objects in the working memory with the target task type and the available alternatives as attributes. Each plan alternative is a copy of the original plan, where the specific task was replaced with an alternative task. A task, which cannot be broken down or replaced by alternatives, is called a primitive task. The result of the previous steps is an ordered list of primitive tasks. Special rules for each primitive task type exist. When a primitive task is activated during planning, these rules simulate the effects of applying the operator by manipulating the working memory accordingly. For example a “Fly Route”-task has the simulated effect of moving the UAV along the specified route until the end of the route is reached. Later, when a valid plan is selected and executed in the real world, operators are not longer simulating the effects, but actually changing the environment by initiating actions. In the case of the “Fly Route”-task this would be a command to the flight management system to fly the specified route. The application of methods, alternatives and operators are the basic HTN algorithm.

The planning process could continue until all available alternatives are explored. This would result in low planning performance for complex problems. Instead, as soon as an executable task is inserted into the plan, it is simulated, as described above, and the resulting changes are recorded in working memory in the planning world associated with this plan. Planning worlds leading to impossible outcomes, e.g. destruction of the UAV or physical impossibilities, are pruned and planning is continued on other plans. To improve pruning, further rules evaluate the costs associated with the tasks and an A*-algorithm is employed to restrict planning to the plan with the currently best prospects. The result is a plan tree as depicted in Fig. 2.

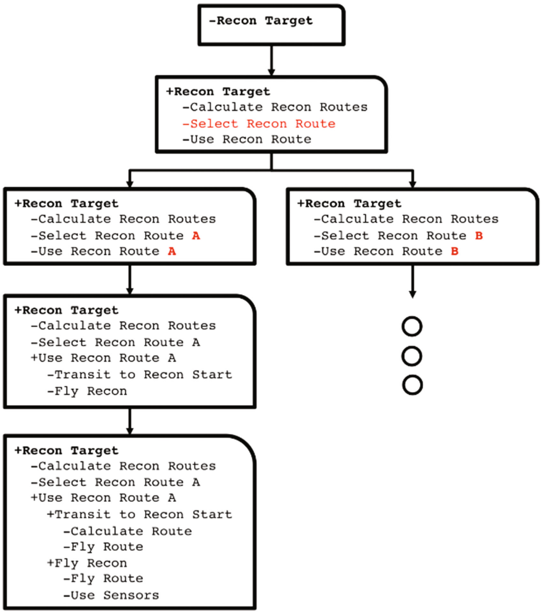


Fig. 1. Planning world tree with branches spawning from alternatives (red) (Color figure online)

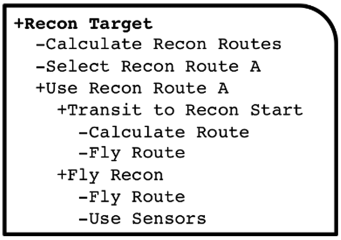


Fig. 2. Resulting plan as a tree of tasks

As soon as a plan is selected, the same simulation rules are used to extrapolate a copy of the plan into the future, while monitoring rules check for conflicts on the simulated as well as the current plan. When a conflict is detected, the simulation data is used to determine the type, the cause and the time until it becomes critical. For example a monitoring rule for proximity warnings to enemy units is activated when the simulated UAV is moved along a route by a “Fly route”-task and the distance to an enemy unit is lower then a threshold. Each activation of this rule increases the threat costs of the affected plan, which leads to a preference of other plans over this one during the planning process and to a warning to the user, when this plan is selected for execution. This feature enables proactive warnings, as mentioned previously. Since the HTN planning algorithm makes use of the same rules in order to evaluate the different alternatives, the knowledge is reused, which results in a powerful planning and monitoring architecture. The warning system is currently able to identify the following types of plan conflicts or ROE violations:

Critical:

- Ground Collision: Following the current flight plan leads to a dangerous proximity of the aircraft to the ground.
- Threat: The aircraft is currently in the firing range of an enemy unit.

Less critical:

- Trespass: Following the current flight plan causes the aircraft to enter the firing range of an enemy unit in the future.
- AirspaceViolation: Following the current flight plan causes the aircraft to violate airspace borders in the future.
- NoFlyViolation: Following the current flight plan causes the aircraft to violate no-fly-zones in the future.
- GimbalLock: The camera gimbal is currently in use by the human operator and not available for a scheduled reconnaissance task.
- DetectionLevelViolation: Following the current flight plan to an unwanted proximity of the aircraft to the reconnaissance target, which might cause detection by enemy forces.

4.2 User Interface

Overview. The main component of the user interface is a moving map with integrated sensor display. The command and control interface for the agent is implemented directly on the map display (see Fig. 3).

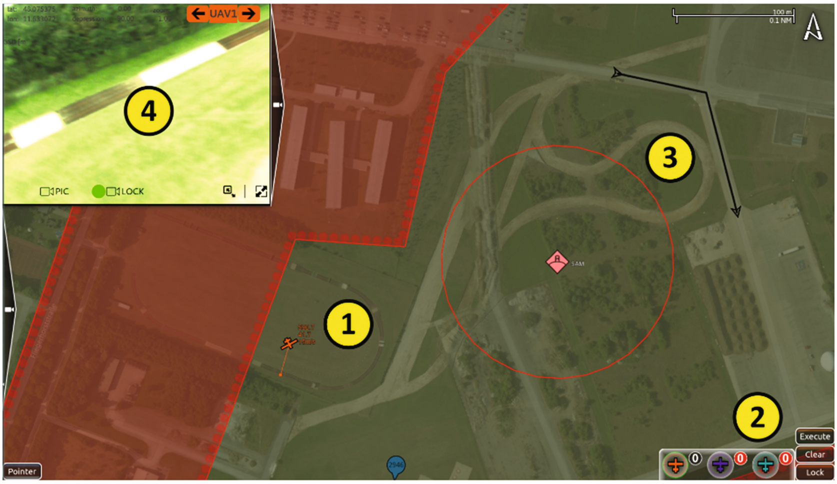


Fig. 3. Moving map interface with UAV (1), dock (2), tactical symbols (3) and sensor view (4)

Among map usability functions like a north arrow, an adjustable scale and simple drawing functions, the user interface for the UAV consisted of the following items.

Aircraft symbols (1) display position, altitude and speed information along with a point on the map, indicating the current camera viewing position on the ground.

The UAV dock allows for fast selection and overview for all available UAVs. It is also used to start, pause or abort the execution of a mission plan.

Tactical symbols on the map (3) display the current situation using the NATO Mil-Std-2525C [14] (e.g. SAM) alongside new symbols for routes and areas.

Two sensor views allow direct access to the onboard sensors, including low-level functions as locking on ground positions or adjusting the zoom.

Intent Communication from the operator to the agent is mainly done by assigning tasks through the map interface. To issue a command the operator selects a point or an object on the map and chooses an entry from the appearing context menu (see Fig. 4). This way the target and the type of the task are selected. Possible commands are “Transit”, “Recon”, “Scout” and variations. By issuing a task the operator communicates his or her intent to the agent. For example, the intent of a “Transit” task is to move the UAV to a certain location, while a “Recon” task is used to gather information about a certain object.

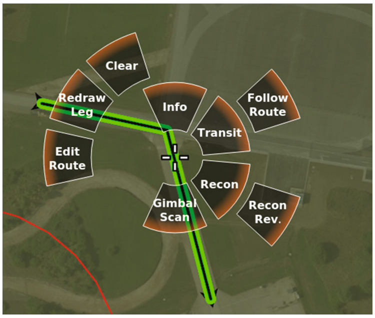


Fig. 4. Context menu on a route. Used for issuing tasks.

Constraints for tasks (e.g. Aggressiveness, Altitude, etc.) can be applied by clicking on a task arrow and choose the appropriate selection in the context menu. This allows more detailed control over the execution of the task.

Status Feedback is mostly embedded into the map. Issued tasks are displayed as arrows pointing towards the target with the task type as symbol next to it (see Fig. 5 top right). The current agenda is therefore easily accessible and embedded into the map context.

As soon as a plan is calculated, zooming in onto the tasks can access more information about them and how they are executed (see Fig. 5 bottom left and right), as task arrows exceeding a certain length on the map are replaced with their subtasks (see Fig. 5).

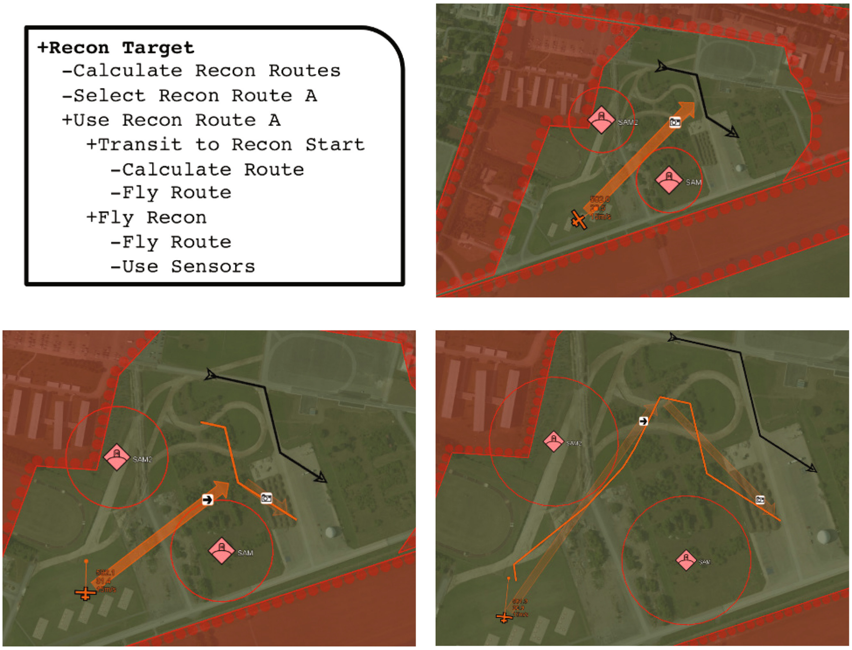


Fig. 5. Plan status: Task arrows are broken down into subtasks, if zoom level is increased.

The level of detail displayed for the plan is connected to the zoom level of the map. At first only the subtasks are displayed. With higher zoom levels, more detailed information is revealed. This function makes use of the tree structure of plans, as tasks are broken down into subtasks, as soon as the arrows exceed a certain length. The last possible step is displaying the actual flight route. This way the operator is able to adjust the amount of information. According to the situation a status overview, a detailed status report or multiple degrees in between can be accessed without cluttering the screen. The current number of tasks in the agenda is displayed next to the UAV symbol in the dock, which is displayed in Fig. 6. The background of the task count label indicates whether the UAV is executing tasks (flashing green), on hold (white) or offline (red). A running planning process is signaled by a spinning wheel at the same position. Around the UAV symbol a circular bar indicates the progress in executing the current plan.

Fig. 6. UAV Dock. Left UAV has two tasks and warnings, center UAV is planning, and right UAV is offline. (Color figure online)

A green circle indicates the currently selected UAV. The three control buttons are colored the same as the currently selected UAV and therefore signal their command target. The top button switches title, depending on the UAV status. It displays “Hold” during plan execution and “Execute” (currently not visible) when holding. The “Clear” button clears all tasks, while the “Lock” button centers the view on the UAV and causes the view to follow the UAV.

Warnings are displayed on the map as well. They contain the source and location of the conflict, the time until the conflict gets critical and offer an automatic solution if one is available. Figure 7 depicts a warning situation. Similar to the level of detail for the current plan, the warnings are at first only displayed as icons in the UAV dock, as well as next to the UAV (1). When selected, they reveal type and critical time as well as a solution button, if available (2). When the “Show” button is pressed, a description dialog appears (3) and the causes are centered and highlighted on the map (4). This approach results in a freely selectable level of detail for status and warning information and allows the operator to concentrate on the current tasks without cluttering the user interface.

Fig. 7. Three levels of warning information: Symbols in dock and next to UAV (1), detailed description (2) and info dialog (3) with flashing object indicator (4).

In this example the UAV flight plan leads through the range of an enemy SAM and the camera gimbal is locked by the human operator, thus not available for an automatic recon task. Pressing the "Replan" button would result in a flight path around the SAM, but could not release the gimbal, as only the operator can authorize this.

5 Evaluation

Goal of the evaluation was to get a first impression if the concept of responsive human-automation interaction is feasible within the context of UAV guidance and is accepted by the military users. The evaluation setup consisted of a simulated reconnaissance mission with a single UAV controlled from a ground control station (GCS). The mobile GCS of the institute was equipped with the previously described user interface concept. A reduced subset of the MUM-T mission types consisting of area, route and point reconnaissance tasks was used. Each reconnaissance mission consisted of a target, reporting points and a bounding box, created by no-fly zones. Additional obstacles, like enemy surface to air missiles (SAM), increased mission complexity. 5 officers of the German Bundeswehr were tasked to execute simple variations of each mission type. After an introduction and extended training session of around 45 min, the subjects completed three missions, amounting to around 10 min, including the briefing. During the experiments planning time, execution time, reconnaissance accomplishments and errors (e.g. airspace violations, threats to the UAV and low distances to recon targets) were measured. Questionnaires after each mission gathered the subjective rating for the system. In addition, interviews with the subjects provided information about the reasons for their evaluation.

6 Results

All subjects completed the missions successfully in the required time. The total mission time was on average 71 s with a standard deviation of 25 s (min 26 s, max 114 s). Planning time was on average 17 s with a standard deviation of 7 s (max 27 s, min 7 s). Execution time was on average 54 s with standard deviation of 24 s (min 19 s, max 96 s). All subjects reached 100% reconnaissance performance and made on average only 0.74 mistakes per mission, which consisted mainly of violations of no-fly-zones and detection level. Figures 8 and 9 depict the combined questionnaires for the subjective results.

The ratings for the adjustable status feedback were very positive. Even during the training the subjects used the function very often. The warnings system was assessed to be relatively complex, but helpful nonetheless, although its usage varied between subjects. Overall, the concept was fully accepted by the military users.

The usability of the system was

intuitive	x	xx	x		x			not intuitive
understandable	xx	xx	x					not understandable
easy	xxx	x	x					hard
Time saving	xx	xx	x					Time wasting
comfortable	xx	xx	x					uncomfortable

The information presentation was

well arranged	x	xxx	x					badly arranged
appealing	x	xxx	x					not appealing

The information content was

helpful	xx	xx	x					not helpful
understandable	xxx	x	x					not understandable

Zooming in on tasks to access more information was

easy	xxx		x		x			hard
understandable	xx	xxx						not understandable
helpful	xx	xxx						not helpful

Fig. 8. Combined questionnaire results, part 1 of 2

The three stage warning system was

easy	x	x	xx	x				hard
understandable	xx	x	xx					not understandable
helpful	x		xxx	x				not helpful

Task-based guidance for mission execution was

appropriate	xx	xx	x					not appropriate
helpful	xxx		x	x				not helpful

Did the UAV behave unpredictably

No, never	x	xx	x			x		Yes, often

Did the UAV system execute tasks always correctly

Yes, often	x	xx	xx					No, never

I am satisfied with the system

Yes	x	xxx	x					No

Fig. 9. Combined questionnaire results, part 2 of 2

7 Conclusion

This article introduced a concept for a responsive human automation interaction by combining task-based guidance with scalable autonomy for the communication to the machine, and an adjustable status feedback interface, including warnings, for the communication to the pilot. The necessary base for the applicability of this concept is an intelligent agent software, capable of planning and situation assessment. Due to the HTN planning capability plans are calculated as task trees. This offers a simple way to prevent clutter on the interface as well as opacity by linking the amount of information displayed to the length of a task arrow on the map. More information can therefore be obtained by zooming in on the task. Warnings assist in situations were planning conflicts or violations are detected by the system. They are presented in detail if accessed by the user. The evaluation of the system was done with military personnel by using questionnaires. Overall the operators rated the user interface concept very positively. The next intended steps are to evaluate the concept for multi-UAV missions inside a helicopter simulator and to perform actual flight tests with a ground control station commanding a UAV with the presented concept, while measuring the effects on workload.

References

1. Department of Defense, DoD Directive 3000.09 Autonomy in Weapon Systems, no. 3000 (2012)
2. Sheridan, T.B., Verplank, W.L.: Human and computer control of undersea teleoperators. ManMachine Syst. Lab Dep. Mech. Eng. MIT Grant N0001477C0256, p. 343 (1978)
3. Clauß, S., Schulte, A.: Task delegation in an agent supervisory control relationship. Capability awareness in a cognitive agent. In: 2014 IEEE International Conference Systems Man and Cybernatics *(*2014). (in publication)
4. Uhrmann, J., Schulte, A.: Task-based guidance of multiple uav using cognitive automation. In: Cognitive 2011 Third International (2011)
5. Uhrmann, J.: Auftragsbasierte Multi-UAV-Führung aus dem Helikoptercockpit durch kognitive Automation (2013)
6. Clauss, S., Schulte, A.: Implications for operator interactions in an agent supervisory control relationship. In: International Conference on Unmanned Aircraft Systems (ICUAS) (2014)
7. Hbf, P.S., Miller, C.A.: Rules of Etiquette, or How a Mannerly AUI Should Comport Itself to Gain Social Acceptance and be Perceived as Gracious and Well-Behaved in Polite Society (2000)
8. Denk, E., Clauß, S., Borchers, A., Werner, J., Schulte, A.: Experimental evaluation of varying feedback of a cognitive agent system fur UAV mission management. In: 18th International Symposium on Aviation Psychology, pp. 1–6 (2015)
9. Marcotte, E.: Responsive Web Design, no. 306 (2010)
10. Rudnick, G., Schulte, A.: Scalable autonomy concept for reconnaissance UAVs on the basis of an HTN agent architecture. In: 2016 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 40–46 (2016)
11. Drools Business Rules Management System. http://www.drools.org. Accessed 01 Jan 2015
12. Nau, D.S., Au, T.C., Ilghami, O., Kuter, U., Murdock, J.W., Wu, D., Yaman, F.: SHOP2: an HTN planning system. J. Artif. Intell. Res. (JAIR) **20**, 379–404 (2003)
13. Erol, K., Hendler, J., Nau, D.: HTN planning: complexity and expressivity. In: Proceedings of the Twelfth National Conference on Artificial Intelligence (1994)
14. Department of Defense, MIL-STD-2525C: Common Warfighting Symbology, p. 1170, November 2008

Team Situation Awareness: A Review of Definitions and Conceptual Models

Manrong She and Zhizhong Li(✉)

Department of Industrial Engineering, Tsinghua University, Beijing 100084,
People's Republic of China
shemanrong@163.com, zzli@mail.tsinghua.edu.cn

Abstract. Situation awareness (SA) has been a hot topic in the area of human factors and ergonomics. The SA in collaborative socio-technical systems, which is called team situation awareness (TSA), also draws increasing attentions. TSA is considered as a critical influencing factor in task performance. Like SA and many other psychological constructs, TSA receives numerous controversies in its definitions, conceptual models, theoretical underpinnings, etc. Based on a careful review of literature, this paper provides a summary and comparison of different TSA definitions, conceptual models and theoretical underpinnings. Several relevant but confusing terms are distinguished. The major controversies on TSA, including the critiques and responses, are also reviewed. The review is expected to help readers to have a comprehensive and up-to-date understanding of TSA.

Keywords: Team · Situation awareness · Definition · Conceptual model · Controversy

1 Introduction

Situation awareness (SA) has been a hot topic in research areas since the late 1980s [1–3]. Most of the relevant studies are carried out from the perspective of individuals. In complex systems, however, a team of operators often performs tasks collaboratively to achieve an ultimate goal. The SA in collaborative environments, the so-called team situation awareness (TSA), is receiving more concerns [4]. TSA is considered as a critical factor in team decision-making [5–7]. Maintaining TSA helps the team members to know when to work individually and when to work together thus enables fluent shifts in the mixed-focus collaboration [8, 9]. TSA simplifies the verbal communications among team members and improves the collaboration efficiencies [9, 10]. In emergency situations, maintaining a high level of TSA aids the team members to react quickly to abnormalities and avoid unexpected consequences [1]. Due to the potential problems (such as information overload, attention narrowing and over-reliance) caused by system digitalization and automation, TSA becomes more pronounced in collaborative system design, assessment and training [11]. As is the case with SA, TSA receives both attentions and controversies [4]. There have been contentions in its theoretical underpinnings and valid measures [3, 12]. In this paper, we first make clear the concept of SA, because many TSA theories are the extensions of the SA ones.

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 406–415, 2017.
DOI: 10.1007/978-3-319-58472-0_31

The TSA definitions, conceptual models, and their theoretical underpinnings, are then reviewed from literature. Several relevant but confusing terms, including shared situation awareness (shared SA), mutual awareness, task-work situation awareness (task-work SA), teamwork situation awareness (teamwork SA), distributed situation awareness (DSA), compatible situation awareness (compatible SA) and transactive situation awareness (transactive SA) are distinguished. The major controversies on TSA, including the critiques and responses, are also reviewed.

2 Situation Awareness (SA)

As one of the mostly studied and debated topics in human factors and ergonomics, so far SA has not received a universally accepted definition [2, 3]. In simple words, SA can be described as what is going on [5]. Among the attempts to define SA, Endsley's work is mostly cited. Based on the information processing theory and the pattern matching theory, Endsley proposes a three-level SA model [5, p. 35] and defines SA 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 near future' [13]. Consisting of three levels (i.e., perception, comprehension and projection), SA is regarded as a part of the information processing [3]. According to the model, the data in the environment is the basis for acquiring SA. Maintaining SA is the premise of decisions and actions. The environmental data perceived is matched to the mental models held in the individual's mind and the individual forms the understanding of the current situation [5]. The task goals, the mental models and the feedback of actions all direct individual's attention to certain features in the environment [12]. Roughly this model follows an input-process-output-feedback loop. Using the similar theoretical basis, Fracker [14, 15], Sarter and Woods [16], and Dominquez [17] raised their own definitions. Fracker highlights the importance of attention allocation and regards SA as 'the knowledge that results when attention is allocated to a zone of interest at a level of abstraction' [14]. The integration of environmental information with existing knowledge (i.e., mental model, which stores in long-term memory) happens in the working memory [15]. Sarter and Woods consider SA as the temporal product of continuous situation assessment and argue SA is 'the accessibility of a comprehensive and coherent situation representation which is continuously being updated in accordance with the results of situation assessment' [16]. Dominquez emphasizes on the cyclical and dynamic nature of SA and refers to it as 'the continuous extraction of environmental information, the integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in direction future perception and anticipating future events' [17].

Taylor [18] holds the view that SA has three components: the attentional demand, the attentional supply and the understanding of the situation. According to Taylor, the SA is 'the knowledge, cognition and anticipation of events, factors and variables affecting the safe, expedient and effective conduct of a mission' [18]. Based on this view, he raised a simple subjective SA rating technique called SART [18, 19].

Adams et al. [20] and Smith and Hancock [21] base SA on Niesser's perceptual cycle theory [22]. That is, the environment provides the available information to the individual and modifies his or her knowledge. The knowledge possessed by the individual directs him or her to take certain actions. Being acted on, the environment is sampled or altered. The changes in the environment then modify the knowledge in the individual. In this way, the environment, knowledge and action form a circulation [21]. Adams et al. believe that SA can be 'the state of the active schema' or 'the state of the perceptual cycle at any given moment' [20]. Smith and Hancock [21] distinguish SA from competence and performance to clarify what SA is. They argue that competence resides in the individual and is independent of the situation, while performance resides in the world and changes with the situation. SA is the mediation of the two and resides in the interaction between the individual and the world. Smith and Hancock view SA as 'the adaptive, externally directed consciousness'. The 'consciousness' in the definition indicates the environment-directed knowledge creation behavior [21].

Based on the perceptual cycle theory and schema theory, Stanton and his colleagues [23] suggest that the individual's inner held schema (i.e., genotype schema) drives the continuous circulation of perception and action. On the other hand, the tasks and external environments activate and update the schema. This view of perceptual cycle and schema makes no big difference from other researches. What really makes a difference is their view of distributed cognition. Stanton et al. advocate a system thinking of SA and propose that SA is distributed amongst the humans and nonhuman artifacts in the socio-technical system [3, 4, 24]. In their view, SA no longer exists solely in the individuals, but is an emergent property of the system [3]. The possession, usage and exchange of SA-related knowledge depend on the goals and sub-goals, the tasks to be performed and the specific roles of the agents (both human and nonhuman). Thus they define SA as 'the activated knowledge for a specific task, at a specific time within the system' [25].

3 Team Situation Awareness (TSA)

In many organizational environments, especially complicated ones, a team of individuals takes charge of the operational tasks [26]. Team is not a merely summation of individuals, but a 'distinguishable set of two or more people who interact dynamically, interdependently and adaptively toward a common valued goal/object/mission, who have each been assigned specific roles or functions to perform and who have a limited life span of membership' [27, 28]. Within a team, the workloads are distributed and the required knowledge is complemented. Facing unexpected events or emergencies, teams are more robust than individuals [29]. Effective teams engage in both task-work and teamwork [30]. Task-work refers to the performance of tasks undertaken by individual team members with their specific roles [28]. Teamwork refers to the episodic, dynamic and interactive process that shares the attitudes, thoughts and behaviors [30]. Task-work is necessary for the execution of individual goals and teamwork is critical for the coordination to achieve team goals [28, 30]. Generally the team is composed of humans, while from the perspective of system thinking, the team's scope is expanded to both humans and nonhuman artifacts.

3.1 Definitions

The traditional and dominant view emphasizes TSA on a shared understanding of the situation, that is, the team members ought to have a common picture [3]. This is how the concept of “shared SA” is raised. Shared SA is the ‘the degree to which every team member have the same SA on shared SA requirements’ [29]. Team members do not need to share all the environmental elements but those “common SA requirements” with each other. Endsley and Robertson [31] argue that the effective team performance relies on a high level of individual SA and the same SA for the common environmental elements that are required. Team SA, as defined by Endsley, is ‘the degree to which every team member possess the SA needed for his or her job’ [5]. Wellens [32] suggests that to maximize the TSA (in his research TSA is called group SA), the environmental elements monitored by each team member should have enough overlaps. Wellens defines TSA as ‘the sharing of a common perspective between two or more individual regarding current environmental events, their meaning and projected future status’ [32]. Salas et al. [28] figure out the importance of communication and information exchange in successful team performance. They conclude that individual SA and team process constitute TSA, in which the team process is ‘the teamwork behaviors and cognitive processes that facilitate team performance’. This way, Salas et al. point out ‘TSA is at least in part the shared understanding of a situation among team members at one point in time’ [28]. Shu and Furuta [26] extend Bratman’s [33] shared cooperative activity theory (SCA), in which mutual responsiveness, commitment support, and joint activity are identified as key features of SCA. The cooperative activity is simplified to individual activities and mutual beliefs. They believe that TSA is partly shared and partly distributed among team members [26]. Shu and Furuta then define TSA as ‘two or more individuals share the common environment, up-to-the-moment understanding of situation of the environment, and another person’s interaction with the cooperative task’ [26]. This definition implies that TSA consists of individual SA and mutual awareness. Shu and Furuta do not provide a clear definition of mutual awareness. Schmidt subsequently defines mutual awareness as ‘the mutual understanding of each other’s intentions, beliefs and activities’ [34].

Although the above definitions emphasize on different aspects of TSA, the authors all conceive that TSA resides in people’s mind and TSA involves the interactions among team members. To summarize the above definitions, TSA is a multidimensional construct and contains the SA related to the individual’s own roles, the SA of other team members, and the SA of the overall team. The SA of other team members is called task-work SA and the SA of the overall team is named teamwork SA [4].

Several researchers argue that TSA is a characteristic of the socio-technical system, rather than a psychological construct of the individual [3, 4, 24, 25, 35]. Artman and Garbis [35] hold the view that TSA is distributed across the human agents and non-human artifacts in the system. They give the definition of TSA as ‘the active construction of a model of a situation partly shred and partly distributed between two or more agents, from which one can anticipate important future states in the near future’ [35]. The “agents” in this definition can be both human and artifacts. After that, Stanton and his colleagues propose the concept of distributed situation awareness (DSA) [25]. Please note that this is not the first time that DSA is raised. Endsley and Jones [29] used to define DSA as 'the SA in teams which member are separated by distance, time and/or obstacles'. Stanton et al. endow TSA a new meaning. According to [3, 4, 25], SA is not shared, but distributed amongst the whole system. The agents (both human and artifacts) possess unique awareness based on their goals and tasks. Their views can be different even on the same environmental elements. If the different views serve to the overall goal of the system, these views are regarded as compatible. This is what compatible SA means. An agent's deficiency in SA can be enhanced or modified through the transaction with other agents. The transaction implies the interaction among agents, such as the verbal and nonverbal communications in the system. This is where transactive SA comes from. Endsley endorses that the information exists throughout the system containing both human and artifacts. What she finds hard to understand is why the inanimate artifacts have the awareness. In her view, those objects (e.g. displays, technologies, documents) are just repositories of information [12].

3.2 Conceptual Models

In the early years of TSA studies, Endsley raise a TSA model, in which a set of circles overlaps with each other [5, p. 39]. Each circle represents a team member's SA elements related to his or her specific role. The overlaps of the circles represent shared SA and the union of the circles represents TSA.

Salas et al. [28] argue that TSA includes individual SA and team processes. The individual's existing knowledge (mental model) and expectations affects the perception of elements in the environments. The limitation of one's mental model can be complemented and updated through the information exchange with other team members. Team processes, such as planning and assertiveness, facilitate the information exchange among team members. The model raised by Salas et al. describes the cyclical and dynamic nature of TSA [28, p. 130]. In their model, the team members' preexisting knowledge, the team members' characteristics and the team processes interacts with each other. As the product of the team situation assessment, SA also affects the above components in the model.

Incorporating the team processes into the depiction of TSA, Endsley and Jones extend the early TSA model (refer to [5, p. 39]) to reveal the key factors influencing the quality of TSA [12, p. 24, 36, pp. 49–57]. The four factors are TSA requirements, TSA devices, TSA mechanisms and TSA processes. TSA requirements include the elements in thee levels that need to be shared among team members and how well these elements are maintained affects the quality of teamwork. TSA devices include physical ones (e.g. shared displays), nonphysical ones (e.g. verbal and nonverbal communications), and the shared environment. The devices help team members to form shared understanding of the situation. TSA mechanisms refer to the degree to which team members adopt the mechanism (e.g. shared mental models) to understand and project the situation in the same way. Shared mental models benefit the communications [37] and enable the individual to predict other members' behaviors [38]. TSA processes, such as self-checking, coordination, prioritization and questioning, facilitate the ongoing of information exchange [28].

Based on the above TSA models, Salmon et al. provide a summary model in which TSA includes the individual SA, the task-work SA and the teamwork SA [4, p. 318]. However, Salmon et al. argue that this summary model applies only in simple, small-size collaborative systems. As the system becomes more complicated and numerous individuals are distributed far from one another, the DSA model demonstrates its advantages [4].

In reviewing the definitions of SA and TSA raised by Stanton and his colleagues, the DSA theory has been introduced in details. To summarize the key points of the DSA model [3], the perception cycle theory, the schema theory, and the distributed cognition theory together underpin it. From the perspective of system thinking, TSA is distributed across the people and the artifacts in the system. The schema held in the agent drives the perception of elements and directs the action in the world. In turn, the world modifies the schema. This schema-driven process is a continuous circulation. Towards the same external elements, different agents may have different views. These views are exchanged through transactions and are compatible towards the ultimate goal. The compatible SA and transactive SA are crucial for an efficient team performance.

3.3 Controversies

- Whether the three-level model is linear?

Some researchers criticize Endsley's three-level model (perception, comprehension and projection) is linear [39, 40]. According to the model, Level 1 (perception) seems to be necessary to achieve Level 2 (comprehension); Level 1 and Level 2 seem to be inevitable for Level 3 (projection) [41]. This bottom-up, data-driven behavior does exist in real situation. However, a large portion of behaviors follows the top-down, goal-driven pattern [42]. The goal-driven behaviors refer that the task goals, the understanding, and prediction of the situation direct individual's attention to certain elements in the environment. Without the rich information or the detailed model, people may still judge the situation well [3]. Endsley agrees that both data-driven and goal-driven processes of maintaining SA exist [12]. Her view that the comprehension of the situation drives the attention resources to new data in the environment proves this [43]. Endsley denies that the three-level model is a simple 1-2-3 linear one, arguing that the 1-2-3 levels represent the ascending stages of SA [12].

- Is SA a process or product?

There have been debates on whether SA is a product or the process to acquire it. Sarter and Woods regard SA as the comprehensive situation representation as the result of situation assessment [16]. Salmon el al. regards SA as both a product (the mapping of the external information with inner schema) and the process (the continuous perception-action processing) [4]. Endsley used to distinguish SA from situation assessment. The SA in her definition is often regarded as a product [4], which Endsley does not agree with [12]. Ensley argues that the three-level SA model fully describes the process of maintaining SA, such as the matching of mental models on external information and the direction of SA to certain elements in the environment [5].

- Does TSA reside in the mind, in the world or in the interaction of the two?

Whether TSA resides in the mind, in the world, or in the interaction of the two depends on the unit of analysis [24]. Endsley [5, 13], Wellens [32], Salas [28], etc. regard the individual as the unit of analysis and define TSA from the perspective of cognitive psychology. Thus they believe TSA resides in the individual's mind. Ackerman [44], Jenkins [45], etc. regard the objects and artifacts that people interact with as the unit of analysis and define TSA from the engineering perspective. Thus they believe TSA resides in the world. Stanton et al. [3, 25] regard the whole socio-technical system as the unit of analysis and define TSA from the perspective of systems ergonomics. Thus they believe TSA resides in the interaction of human-with-human, human-with-artifact and artifact-with-artifact.

- Whether the TSA is shared or exchanged among team agents?

The researchers who conceive TSA is shared insist that team members ought to have the same understanding on the same information requirements [28, 29, 32]. The common picture facilitates the teamwork [3]. On the contrast, those who conceive TSA is exchanged believe that team agents (both human and artifacts) possess different views on the same information [3, 4, 24, 25]. These different views are exchanged among team agents to form a compatible awareness of the ultimate situation [25]. Stanton et al. [3] hold the view that the shared SA only applies for simple scenarios in small-scale teams, and the DSA is applicable for real complicated situations. Even so, Stanton approves that there are cases in which a common picture (shared awareness) is needed for effective team performance [3].

- What are the cause of Air France 447 crash and many other accidents? Loss of individual SA or loss of DSA?

Many accidents are caused by the loss of SA [11, 12]. Take Air France 447 crash in 2009 for example: the psychological view of TSA [12] attributes the accident to the individual's failure in detecting key information (airspeed) and in accurately judging the situation in time. Attentional narrowing and over-reliance on automation are likely to cause such failures. In the system view of TSA, however, the aircraft crash is not solely due to the failure of individuals. System-view supporters [3] argue that the information is always there; the accident is due to the disconnection between artifacts (the icing tubes that cannot measure the airspeed) and individuals. That is, the information is not transacted successfully within the system. Though the agents possess their own SA, these SAs are not compatible towards the ultimate system goal.

- What TSA model is applicable for what situation?

So far there are no empirical studies on this issue. Stanton et al. [3] propose a three-dimension model to infer what SA approaches are applicable for what situation. The three dimensions in the model are individual/system, stable/dynamic, and normative/formative. Stanton et al. suggest that the individual SA approaches are more suitable for normative, stable and less systemic situations; the team SA approach (e.g. Endsley [29], Salas et al. [28]) for formative, stable and less systemic situations; the systems SA approaches (e.g. DSA model [25]) for formative, dynamic and more systemic situations. Besides the above three dimensions, the dimensions of team size (small-scale/large-scale), the complexity of the environment (low/high), the degree of coupling (loosely-coupled/tightly-coupled), the character of the task (skill-based/rule-based/knowledge-based) might have an impact on TSA model selection. This issue is worth studying in the future.

4 Conclusion

This paper reviews the literature on team situation awareness (TSA) and compares various theories in terms of definitions, conceptual models and theoretical underpinnings. The major controversies on TSA are also provided for a dialectical view on the TSA theories. This review does not attempt to enumerate the TSA definitions and models in an exhaustive manner, but to figure out what is really going on around this fascinating construct.

Acknowledgement. This study was supported by the Open Funding Project of National Key Laboratory of Human Factors Engineering (Grant No. HF2012-K-02) and the National Natural Science Foundation of China (Grant No. 71371104).

References

1. Stanton, N.A., Young, M.S.: A proposed psychological model of driving automation. Theor. Iss. Ergon. Sci. **1**(4), 315–331 (2000)
2. Wickens, C.D.: Situation awareness: review of Mica Endsley's 1995 articles on situation awareness theory and measurement. Hum. Factors **50**(3), 397–403 (2008)
3. Stanton, N.A., Salmon, P.M., Walker, G.H., Salas, E., Hancock, P.A.: State-of-science: situation awareness in individuals, teams and systems. Ergonomics 1–33 (2017, just-accepted)
4. Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., Jenkins, D.P., McMaster, R., Young, M.S.: What really is going on? Review of situation awareness models for individuals and teams. Theor. Iss. Ergon. Sci. **9**(4), 297–323 (2008). https://doi.org/10.1080/14639220701561775
5. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors: J. Hum. Factors Ergon. Soc. **37**(1), 32–64 (1995)
6. Tenney, Y.J., Pew, R.W.: Situation awareness catches on: what? So what? Now what? Rev. Hum. Factors Ergon. **2**(1), 1–34 (2006)
7. Liu, S., Liao, Z., Zhou, Y., Wang, X., Tao, D.: Analyzing and modeling of crew team situation awareness. In: Long, S., Dhillon, B.S. (eds.) Man-Machine-Environment System Engineering. LNEE, vol. 406, pp. 121–131. Springer, Singapore (2016). doi:10.1007/978-981-10-2323-1_15
8. Dourish, P., Bellotti, V.: Awareness and coordination in shared workspaces. In: Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work, December 1992, pp. 107–114. ACM (1992)

9. Gutwin, C., Greenberg, S.: The importance of awareness for team cognition in distributed collaboration (2001)
10. Gao, Q., Yu, W., Jiang, X., Song, F., Pan, J., Li, Z.: An integrated computer-based procedure for teamwork in digital nuclear power plants. Ergonomics **58**(8), 1303–1313 (2015)
11. Endsley, M.R.: Situation awareness: operationally necessary and scientifically grounded. Cogn. Technol. Work **17**(2), 163–167 (2015). https://doi.org/10.1007/s10111-015-0323-5
12. Endsley, M.R.: Situation awareness misconceptions and misunderstandings. J. Cogn. Eng. Decis. Making **9**(1), 4–32 (2015). https://doi.org/10.1177/1555343415572631
13. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, NAECON 1988, pp. 789–795. IEEE (1988)
14. Fracker, M.L.: A theory of situation assessment: implications for measuring situation awareness. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 32, no. 2, pp. 102–106. SAGE Publications (1988)
15. Fracker, M.L.: Measures of situation awareness: review and future directions (No. AL-TR-1991–0128). Logue (George E) Inc., Montoursville, PA (1991)
16. Sarter, N.B., Woods, D.D.: Situation awareness: a critical but ill-defined phenomenon. Int. J. Aviat. Psychol. **1**(1), 45–57 (1991)
17. Dominguez, C.: Can SA be defined. In: Situation Awareness: Papers and Annotated Bibliography, pp. 5–15 (1994)
18. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: AGARD, Situational Awareness in Aerospace Operations, no. 90-28972 23-53, 17 p. (1990)
19. Selcon, S.J., Taylor, R.M.: Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. In: AGARD, Situational Awareness in Aerospace Operations, no. 90-28972 23-53, 8 p. (1990)
20. Adams, M.J., Tenney, Y.J., Pew, R.W.: Situation awareness and the cognitive management of complex systems. Hum. Factors **37**(1), 85–104 (1995)
21. Smith, K., Hancock, P.A.: Situation awareness is adaptive, externally directed consciousness. Hum. Factors: J. Hum. Factors Ergon. Soc. **37**(1), 137–148 (1995)
22. Neisser, U.: Cognition and Reality: Principles and Implications of Cognitive Psychology. WH Freeman/Times Books/Henry Holt & Co., New York (1976)
23. Stanton, N.A., Salmon, P.M., Walker, G.H., Jenkins, D.: Genotype and phenotype schemata as models of situation awareness in dynamic command and control teams. Int. J. Ind. Ergon. **39**(3), 480–489 (2009). https://doi.org/10.1016/j.ergon.2008.10.003
24. Stanton, N.A., Salmon, P.M., Walker, G.H., Jenkins, D.P.: Is situation awareness all in the mind? Theor. Iss. Ergon. Sci. **11**(1–2), 29–40 (2010). https://doi.org/10.1080/14639220903009938
25. Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Linsell, M.: Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics, **49**(12–13), 1288–1311 (2006)
26. Shu, Y., Furuta, K.: An inference method of team situation awareness based on mutual awareness. Cogn. Technol. Work **7**(4), 272–287 (2005). https://doi.org/10.1007/s10111-005-0012-x
27. Salas, E., Dickinson, T.L., Converse, S.A., Tannenbaum, S.I.: Toward an understanding of team performance and training (1992)
28. Salas, E., Prince, C., Baker, D.P., Shrestha, L.: Situation awareness in team performance: implications for measurement and training. Hum. Factors **37**(1), 123–136 (1995)

29. Endsley, M.R., Jones, W.M.: A model of inter-and intra-team situational awareness: implications for design, training and measurement. In: McNeese, M., Salas, E., Endsley, M. (eds.) New trends in Cooperative Activities: Understanding System Dynamics in Complex Environments, pp. 46–68. Human Factors and Ergonomics Society, Santa Monica (2001)
30. Salas, E., Shuffler, M.L., Thayer, A.L., Bedwell, W.L., Lazzara, E.H.: Understanding and improving teamwork in organizations: a scientifically based practical guide. Hum. Resour. Manage. **54**(4), 599–622 (2015). https://doi.org/10.1002/hrm.21628
31. Endsley, M.R., Robertson, M.M.: Situation awareness in aircraft maintenance teams. Int. J. Ind. Ergon. **26**(2), 301–325 (2000)
32. Wellens, A.R.: Group situation awareness and distributed decision making: from military to civilian applications. In: Individual and Group Decision Making: Current Issues, pp. 267–287 (1993)
33. Bratman, M.E.: Shared cooperative activity. Philos. Rev. **101**(2), 327–341 (1992)
34. Schmidt, K.: Some notes on mutual awareness. COTCOS, CTI, DTU, Copenhagen, Denmark (1998)
35. Artman, H., Garbis, C.: Team communication and coordination as distributed cognition. In: 9th Conference of Cognitive Ergonomics, pp. 151–156 (1998)
36. Endsley, M., Jones, W.M.: Situation awareness information dominance & information warfare. Logicon Technical Services Inc., Dayton, OH (1997)
37. Perla, P.P., Markowitz, M., Nofi, A.A., Weuve, C., Loughran, J.: Gaming and shared situation awareness (No. CRM-D0002722. A2). Center for Naval Analyses, Alexandria, VA (2000)
38. Salas, E., Stout, R., Cannon-Bowers, J.: The role of shared mental models in developing shared situational awareness. In: Situational Awareness in Complex Systems, pp. 297–304 (1994)
39. Dekker, S., Lützhöft, M.: Correspondence, cognition and sensemaking: a radical empiricist approach to situation awareness (2004)
40. Salmon, P.M., Stanton, N.A., Young, K.L.: Situation awareness on the road: review, theoretical and methodological issues, and future directions. Theor. Iss. Ergon. Sci. **13**(4), 472–492 (2012)
41. Sorensen, L.J., Stanton, N.A., Banks, A.P.: Back to SA school: contrasting three approaches to situation awareness in the cockpit. Theor. Iss. Ergon. Sci. **12**(6), 451–471 (2011)
42. Plant, K.L., Stanton, N.A.: The explanatory power of schema theory: theoretical foundations and future applications in Ergonomics. Ergonomics **56**(1), 1–15 (2013)
43. Endsley, M.R.: Situation awareness: progress and directions. In: A Cognitive Approach to Situation Awareness: Theory, Measurement and Application, pp. 317–341 (2004)
44. Ackerman, R.K.: New Display Advances Brighten Situational Awareness Picture. Combat Edge (1998)
45. Jenkins, D.P., Stanton, N.A., Salmon, P.M., Walker, G.H., Young, M.S.: Using cognitive work analysis to explore activity allocation within military domains. Ergonomics **51**(6), 798–815 (2008)

Author Index