Advances in Intelligent Systems and Computing 586

Carryl Baldwin *Editor*

Advances in Neuroergonomics and Cognitive Engineering

Proceedings of the AHFE 2017 International Conference on Neuroergonomics and Cognitive Engineering, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Advances in Intelligent Systems and Computing

Volume 586

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/11156

Carryl Baldwin
Editor

Advances in Neuroergonomics and Cognitive Engineering

Proceedings of the AHFE 2017 International Conference on Neuroergonomics and Cognitive Engineering, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Editor
Carryl Baldwin
Department of Psychology,
Human Factors and Applied Cognition
George Mason University
Fairfax, VA
USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-60641-5 ISBN 978-3-319-60642-2 (eBook)
DOI 10.1007/978-3-319-60642-2

Library of Congress Control Number: 2017943022

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

Advances in Human Factors and Ergonomics 2017

AHFE 2017 Series Editors

Tareq Z. Ahram, Florida, USA
Waldemar Karwowski, Florida, USA

8th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences

Proceedings of the AHFE 2017 International Conferences on Cognitive & Neuroergonomics and Cognitive Computing and Internet of Things, July 17−21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Advances in Affective and Pleasurable Design	*WonJoon Chung and Cliff (Sungsoo) Shin*
Advances in Neuroergonomics and Cognitive Engineering	*Carryl Baldwin*
Advances in Design for Inclusion	*Giuseppe Di Bucchianico and Pete Kercher*
Advances in Ergonomics in Design	*Francisco Rebelo and Marcelo Soares*
Advances in Human Error, Reliability, Resilience, and Performance	*Ronald L. Boring*
Advances in Human Factors and Ergonomics in Healthcare and Medical Devices	*Vincent G. Duffy and Nancy Lightner*
Advances in Human Factors in Simulation and Modeling	*Daniel N. Cassenti*
Advances in Human Factors and System Interactions	*Isabel L. Nunes*
Advances in Human Factors in Cybersecurity	*Denise Nicholson*
Advances in Human Factors, Business Management and Leadership	*Jussi Kantola, Tibor Barath and Salman Nazir*
Advances in Human Factors in Robots and Unmanned Systems	*Jessie Chen*
Advances in Human Factors in Training, Education, and Learning Sciences	*Terence Andre*
Advances in Human Aspects of Transportation	*Neville A. Stanton*

(continued)

(continued)

Advances in Human Factors, Software, and Systems Engineering	*Tareq Z. Ahram and Waldemar Karwowski*
Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries	*Paul Fechtelkotter and Michael Legatt*
Advances in Human Factors, Sustainable Urban Planning and Infrastructure	*Jerzy Charytonowicz*
Advances in the Human Side of Service Engineering	*Louis E. Freund and Wojciech Cellary*
Advances in Physical Ergonomics and Human Factors	*Ravindra Goonetilleke and Waldemar Karwowski*
Advances in Human Factors in Sports, Injury Prevention and Outdoor Recreation	*Tareq Z. Ahram*
Advances in Safety Management and Human Factors	*Pedro Arezes*
Advances in Social & Occupational Ergonomics	*Richard Goossens*
Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future	*Stefan Trzcielinski*
Advances in Usability and User Experience	*Tareq Ahram and Christianne Falcão*
Advances in Human Factors in Wearable Technologies and Game Design	*Tareq Ahram and Christianne Falcão*
Advances in Communication of Design	*Amic G. Ho*
Advances in Cross-Cultural Decision Making	*Mark Hoffman*

Preface

This book brings together a wide-ranging set of contributed articles that address emerging practices and future trends in cognitive engineering and neuroergonomics—both aim to harmoniously integrate human operator and computational system, the former through a tighter cognitive fit and the latter a more effective neural fit with the system. The chapters in this book uncover novel discoveries and communicate new understanding and the most recent advances in the areas of workload and stress, activity theory, human error and risk, and neuroergonomic measures, cognitive computing as well as associated applications.

This book is organized into seven main sections:

Section 1: Human-Autonomy Teaming
Section 2: Audition and Workload
Section 3: Spatial Perception
Section 4: Vision and Memory
Section 5: Neuroergonomics Theory and Design
Section 6: General and Systemic Structural Activity Theory
Section 7: Cognitive Computing and Internet of Things: Techniques and Applications

Collectively, the chapters in this book have an overall goal of developing a deeper understanding of the couplings between external behavioral and internal mental actions, which can be used to design harmonious work and play environments that seamlessly integrate human, technical, and social systems.

Each chapter of this book was either reviewed or contributed by members of the Cognitive & Neuroergonomics Board. For this, our sincere thanks and appreciation go to the Board members listed below:

S. Ramakrishnan, USA
Duncan Speight, UK
Martin Stenkilde, Sweden
Ari Visa, Finland
H. Adeli, USA
Gregory Bedny, USA
Winston "Wink" Bennett, USA
Alexander Burov, Ukraine
P. Choe, Qatar
M. Cummings, USA
Madga Fafrowicz, Poland
Cali Fidopiastis, USA
Chris Forsythe, USA
X. Fang, USA
Qin Gao, China
Y. Guo, USA
Peter Hancock, USA
David Kaber, USA
Kentaro Kotani, Japan
Ben Lawson, USA
S.-Y. Lee, Korea
Harry Liao, USA
Y. Liu, USA
Tadeusz Marek, Poland
John Murray, USA
Denise Nicholson, USA
A. Ozok, USA
O. Parlangeli, Italy
Robert Proctor, USA
David Rodrick, USA
A. Savoy, USA
Dylan Schmorrow, USA
Kay Stanney, USA
Neville Stanton, UK
K. Vu, USA
Thomas Waldmann, Ireland
Brent Winslow, USA
G. Zacharias, USA
L. Zeng, USA

It is our hope that professionals, researchers, and students alike find this book to be an informative and valuable resource; one that helps them to better understand important concepts, theories, and applications in the areas of cognitive engineering and neuroergonomics. Beyond basic understanding, the contributions are meant to inspire critical insights and thought-provoking lines of follow on research that further establish the fledgling field of neuroergonomics and sharpen the more seasoned practice of cognitive engineering. While we don't know where the confluence of these two fields will lead, they are certain to transform the very nature of human–systems interaction, resulting in yet to be envisioned designs that improve form, function, efficiency, and the overall user experience for all.

July 2017 Carryl Baldwin

Contents

Audition and Workload

Spatial Perception

Vision and Memory

Human-Autonomy Teaming

Why Human-Autonomy Teaming?

R. Jay Shively[1], Joel Lachter[1](✉), Summer L. Brandt[2], Michael Matessa[3], Vernol Battiste[2], and Walter W. Johnson[1]

[1] NASA Ames Research Center, Moffett Field, CA 94035, USA
{Robert.J.Shively,Joel.Lachter}@nasa.gov, Walter.W.Johnson@gmail.com
[2] NASA Ames Research Center, San José State University, Moffett Field, CA 94035, USA
{Summer.L.Brandt,Vernol.Battiste}@nasa.gov
[3] Rockwell Collins, 350 Collins Rd NE, Cedar Rapids, IA 52498, USA
Michael.Matessa@rockwellcollins.com

Abstract. Automation has entered nearly every aspect of our lives, but it often remains hard to understand. Why is this? Automation is often brittle, requiring constant human oversight to assure it operates as intended. This oversight has become harder as automation has become more complicated. To resolve this problem, Human-Autonomy Teaming (HAT) has been proposed. HAT is based on advances in providing automation transparency, a method for giving insight into the reasoning behind automated recommendations and actions, along with advances in human automation communications (e.g., voice). These, in turn, permit more trust in the automation when appropriate, and less when not, allowing a more targeted supervision of automated functions. This paper proposes a framework for HAT, incorporating three key tenets: transparency, bi-directional communication, and operator directed authority. These tenets, along with more capable automation, represent a shift in human-automation relations.

Keywords: Human-Autonomy Teaming · Automation · Human factors

1 Introduction

Where is my flying car? For years, we have been told that automation will change our lives dramatically with both positive and negative consequences. On the positive side, we would have flying cars, self-driving cars, personal robots and much more. On the negative side, some jobs might be lost to automation, and visionaries such as Bill Gates, Elon Musk and Stephen Hawking [1] caution us as we approach singularity[1].

However, for better or worse, the promise of automation has not been realized. Great strides have been made in some areas. Self-driving cars are not just on the horizon, but on our streets (with mixed results) [2]; however, most of the promises remain the domain of science fiction. The Roomba leaves a great deal to be desired when compared to Rosie, the Jetson's robotic maid. Why is this the case? Technology

[1] Singularity is the concept that artificial intelligence will eventually think beyond human capacity, which according to some could negatively affect civilization.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_1

development has proceeded at a rapid pace and our desire for increasingly automated systems remains robust. One thing looms large among the many things probably responsible for our dissatisfaction and slow acceptance of automation - that is how automation interacts with humans. In a word, badly. Why is this? Hints and outright answers are there in the literature.

First and foremost, there is the notion that automation replaces and/or relieves the human, making their role less critical. This is not the case. For example, Lee [3] in his review of Parasuraman's and Riley's [4] seminal paper on human-automation interaction, pointed out that the amount of training needed by the humans goes up, not down, when automation is introduced, and the design of the automation interfaces becomes more challenging and important. Also, and maybe more critically, automation usually does not replace the human; rather, it changes the nature of the human's work [5]. For example, with increasing automation on the flight deck, commercial pilots have moved away from their traditional role directly managing the aircraft to a more supervisory role managing the automation.

Of course there are other factors as well. Parasuraman and Riley [4] identified a number of these. Prominent among them were trust and reliance. People need to know when and how they can rely on automation, and often they do not. The result, initial over-reliance on the automation until the inevitable failure, followed by under-reliance on automation are well documented. Not to mention, misunderstanding of the automation leads to abuse/misuse, with potentially disastrous consequences.

So despite the fact that human-automation interaction has been recognized as a critical part of automation design for some time, and has seen development of guidelines for specific domains (e.g., Aerospace [6], Robots and Unmanned Systems [7]), many issues that have plagued this area have remained. Why? Perhaps first and foremost among the unresolved problems is that, in order for us to realize the promise of automation, it must be able work with us, not just for us. It must be a teammate, able to do things that do not require us to supervise it all the time and in all contexts. It needs a greater degree of autonomy. Along with this it must have our trust - earned and appropriate trust.

The concepts behind a relatively new area, called Human-Autonomy Teaming (HAT), are increasingly being seen as one important way in which to realize the capabilities of new powerful automation while gaining its acceptance. This role for HAT is due to many factors such as the recent increases in the speed and quality of automation technology (e.g., Watson [8]), the advances in voice interaction including increasingly natural language understanding (e.g. Siri, Alexa, Google Home), and the ability of automation to work in more collaborative manners (e.g. advanced recommender systems). These have all contributed to automation being thought of and used as a team member. Understanding and effectively designing for HAT may be the key to finally realizing the promise of automation.

2 Human-Autonomy Teaming: Critical Factors

As automation increases to the point of being able to exhibit greater autonomy and team with humans, so does the promise of automation. But, there are still many hurdles to overcome. Full system autonomy continues to be quite difficult (some may claim impossible) to achieve, so, for at least the foreseeable future, most systems will exist in a semi-autonomous state. Given this projection, humans and semi-autonomous systems will continue to need to interact in teams, and the development of autonomous systems that can support this teaming should be based on a foundation of research on human–automation interaction. Here we discuss several commonly cited issues with current automation.

2.1 Brittle Automation

Most automation to date suffers from *brittleness*, operating well for the range of situations it is designed for but needing human intervention to manage situations outside of those environments [9], a situation that will continue to exist in degrees in the future. Automation is designed for a certain environment. True, that environment might be fairly general. An automated car, for instance, might be designed for a wide variety of driving environments. However, there are boundaries. When placed in an environment which the designers did not anticipate, maybe a white panel truck the same color as the sky behind it, the results can be catastrophic [2]. Similarly, software bugs may appear in unusual (and thus poorly tested) situations.

2.2 Lack of Transparency

Many semi-autonomous systems badly lack understandability and predictability. This is referred to as a problem with transparency. This lack of transparency often makes it unclear why the automation is taking the actions that it is taking or not taking an expected action. Maybe the classic example of this was given by the purported perplexity of pilots when faced with the behavior of the flight management system, "What is it doing now, why is it doing that, and what will it do next?" [10]. This opaqueness often makes it impossible for the operator to analyze whether the actions taken by the automation are appropriate.

2.3 Lack of Shared Awareness

Related to the lack of transparency is a lack of shared awareness: operators often do not know what information the automation is using to perform the task. What are the consequences of an operator not knowing the accuracy or reliability of information used by any automated tool (e.g., weather information)? Errors, over- and under-trust and reduced usage are all possible consequences. What are the consequences of the automation not knowing the accuracy or reliability of information used by the operator (e.g., the weather seen out the window)? Poor or confusing recommendations at the very least. Onken [11] suggests a number of factors which will be required for shared situation awareness. He suggests that the key specifications for developing the next generation of cockpit automation should include comprehensive machine knowledge of the actual flight situation and efficient communication between crew and machine. In his model of the "Knowledge Required for Situation Assessment" he identified factors related to the aircraft (e.g., performance data, system status, etc.), crew (resources, standard behavior, individual behavior, etc.), mission (goals and constraints), air traffic control (clearances, instructions, etc.) and environmental factors (navigation data, weather, and traffic). He suggested that the machine cannot assist in an efficient way without a clear understanding of the situation.

2.4 Miscalibrated Trust

One important implication of poor transparency and a lack of shared awareness is miscalibrated trust. Lyons and colleagues [12, 13] demonstrated the importance of transparency in a recent study that assessed different levels of transparency while interacting with an automated route planner. Their study showed that as transparency increased, user ability to understand and *appropriately* trust automation also increased. Here trust was defined using the definition of Lee and See [14, p. 54] as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability.” Inappropriate trust can lead to both underuse of automation due to mistrust, or overuse due to over-trust. To truly and effectively team with automation, humans must be able to trust those systems, or else they cannot rely on them. However, that trust cannot be blind or operators will use the automation under conditions it was not designed for. *Miscalibrated trust* is the enemy here. If the operators do not understand why automation is taking the actions it is taking, they may not be able to determine when they can rely on the automation. They have no basis upon which to calibrate their trust. If operators do not trust the automation when it is supplying correct information, they may not use it, undermining the reason for having the automation. Conversely, if operators trust the automation even under conditions where the automation lacks information necessary to perform adequately, severe problems may result. Lees and Lee [15] suggested that there are at least three dimensions of trust (utility, predictability, and intent) that need to be considered when designing autonomous systems. Their research suggests that operators reliance on automation was more appropriate when transparent information was present. Additionally, research suggests that with semi-autonomous systems, the degree to which people monitor automation decreases with increased *trust* in the automation [16], making the importance of appropriately calibrated trust paramount. Lee and See [14] suggested that to improve reliance, the automation should not only present options but possible solutions. Calibrated trust, an integral part of HAT, is fundamental to making semi-autonomous systems robust and acceptable.

2.5 The Challenge of Monitoring

Brittleness, lack of transparency and miscalibrated trust are not independent issues of course, and as Endsley [17] points out, they combine in ways that are particularly dangerous. Because systems are brittle, they appear to be operating well, until suddenly things go wrong. It often falls to the human operator to monitor for such failures. But monitoring a system that appears to be operating correctly is a job humans are particularly poor at. Unless the operator can see potential problems on the horizon (which they cannot because of poor transparency), there is a strong pull to over-trust the automation. When the system does fail, the operator is in a state Endsley refers to as "reduced situation awareness after being out-of-the-loop." The automation has been performing the task, so they are unaware of the system state. The system has poor transparency, so they cannot discover the system state quickly, an issue exacerbated by the fact that, if the situation were simple, the automation would probably have been able to handle it.

3 A Conceptual Model for HAT

In our work at the Human-Autonomy Teaming Laboratory at NASA Ames Research Center, we have been developing a conceptual model for HAT. There are three major tenets of the HAT model as presented here: Bi-directional Communication, Transparency, and Operator Directed Interface.

3.1 Bi-directional Communication

Teammates often discuss options, brainstorm on solutions and openly discuss courses of action. For automation to be on a team, this bi-directional communication needs to exist. We see bi-directional communication as key to solving a number of the issues typically found in highly automated systems. Bi-directional communication can make systems less brittle. History has shown that humans, generally, are better able to adapt to unfamiliar situations, poorly structured problems, or situations with incomplete information. With HAT, the human can provide the missing information and context, apply judgment from experience in similar situations that would not be recognized as relevant by the automation (Christofferson and Woods [18] have shown how automation does not generalize well from one domain to another) and if necessary, override the decisions made by automation in these situations.

Bi-directional communication can also help to build shared awareness. Alberts and colleagues [19] define shared awareness as a cognitive capability that builds on shared information. Shared awareness of an event can be developed through four mechanisms: (1) directly, (2) though independent sensors, (3) through information that is passed from one agent to another, and (4) through information that is shared and the fused results presented to two (or more) agents. The last two, sharing and fusing of information require a communication channel through which information can pass in either direction. And, while the same information could be gathered independently (mechanisms 1 and 2), to assure shared awareness, this information needs to be cross-validated or eventually the information will be out of sync.

Bi-directional communication requires some ability to have a common cognitive understanding and for both the human and automation to communicate their intentions [20]. To have true bi-directional communication, a shared language will be required. A shared communication channel allowing both implicit and explicit communication between the human and agent, is required. Understanding will be facilitated by explicit discussion of goals (as opposed to intent inferencing), as well as confidence, and rationale.

3.2 Transparency

Lyons [21] argues that the system and its operators should be able to recognize what the human/automated teammate is doing and why. He defines automation transparency as the enabler of such recognition. Transparency is more than simply making the information available to the human operator. Often the calculations done by automation do not correspond directly to those a human would do when performing the same task. To be transparent, the automation must bridge this gap by presenting information in ways that fit the operator's mental model.

Transparency is important because it enables the operator to determine whether the automation is likely to have the correct answer in a particular situation (trust calibration). Without this information operators are likely to trust (and thus rely on) the automation under conditions where it is likely to fail (over-trust) or ignore the automation in conditions where it could be helpful (under-trust).

3.3 Operator Directed Interface

Previous work has focused on task allocation between the humans and automation. However, this static relationship can lead to non-optimal performance. At the same time, if the automation reallocates tasks, the operator can become disoriented. A well planned and understood allocation strategy coupled with an operator directed dynamic allocation of tasks based on workload, time pressure and other important factors allows a much more agile, flexible system. One potential enabler of such a dynamic system of task allocation is the concept of a play [22]. A play encapsulates goals, tasks, and a task allocation strategy appropriate for a particular situation. Operators can call a play to quickly adapt to a new situation.

3.4 A HAT Agent

One approach to creating systems that instantiate these HAT tenets, is to develop a HAT agent that intercedes between the automation and the human operator. Figure 1 provides a conceptual illustration for such an agent. The interface between the HAT agent and the operator instantiates a bi-directional communication channel. Through it the operator can input requests and queries that set the goals for the system. The agent tracks these goals and formulates requests to the automation to meet them. These requests might take the form of status checks to make sure that the system is on track to meet the goals (at least as far as the automation knows), or, if it is not, requests for options to meet the goals. The automation returns possible courses of action, along with their predicted results, and information about the rationale for their selection, and confidence in their success. These results are filtered based on the current context to avoid having too much information overwhelm the contextually relevant results.

An example may illustrate these concepts. The example steps through a pilot teaming with aircraft automation including a recommender system advising which airport to divert to, if needed. In this example an automated alerting system detects an anomaly, such as a wheel well fire. This information is then communicated to the agent,

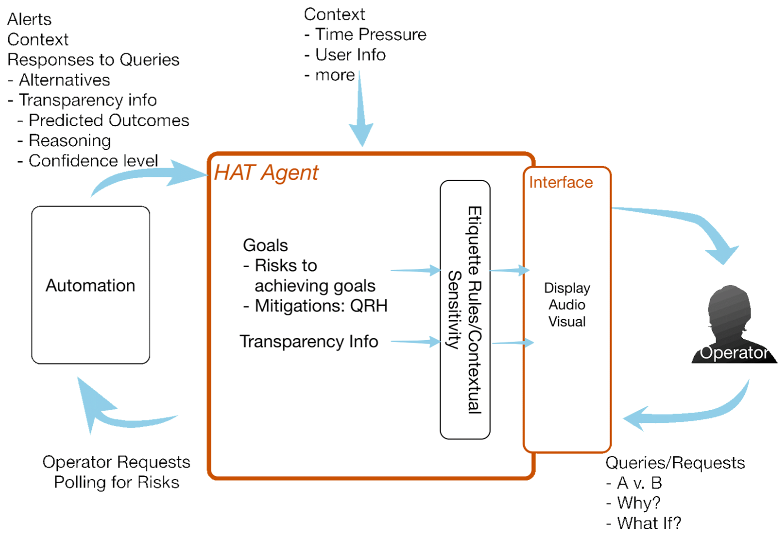


Fig. 1. HAT agent architecture.

which forwards it to the pilot while the automation simultaneously initiates a search for candidate airports. Depending on context, the agent determines which candidates are presented on the dynamic interface, taking into consideration such things as time pressure. If the pilot is under extreme time pressure, it is not appropriate to present multiple options with a great deal of transparency detail on how the automation arrived at its recommendation. This dynamic interface would allow flexible, agile allocation of tasks to humans and pilots, rather than the planned static allocation. In our example, if the pilot does have adequate time, the agent would present multiple options generated by the automation with ratings as to the adequacy of the alternative. This provides transparency to the pilot. This is very important in building the pilot's trust in the system. But, what if the pilot has questions or additional information that the automation does not know about? This highlights a key tenet; bi-directional communication. The pilot needs to be able to ask questions of the automation or provide additional information. S/he should be able to ask: How confident are you in the recommendations? How did you determine the scores? And where appropriate provide additional information: There is a trauma center close to airport X (for a medical emergency). In addition, the pilot should be able to propose solutions and have the automation critique and evaluate them, with the same question asking, perhaps from automation as well as from the pilot. HAT enabled collaborative problem solving is a back-and-forth process. This kind of bi-directional communication transforms automation from a tool to a teammate and has potential to help automation achieve its promise.

4 Conclusion

HAT holds great potential to help automation truly become a partner and for us to realize the promise of automation. Many researchers are hard at work exploring and investigating HAT principles and techniques. We hope that the initial design principles and model for HAT outlined in this brief paper can help add to a foundation on which to base research. Other papers in this session will look at empirically evaluating these concepts and generalizing the results to other domains.

Acknowledgments. We would like to acknowledge NASA's Safe and Autonomous System Operations Project, which funded this research.

References

1. The Observer. http://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence
2. PBS. http://www.pbs.org/newshour/rundown/driver-killed-in-self-driving-car-accident-for-first-time
3. Lee, D.D.: Review of a pivotal human factors article: "humans and automation: use, misuse, disuse, abuse". Hum. Factors **50**, 404–410 (2008)
4. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Factors **39**, 230–253 (1997)
5. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: an attentional integration. Hum. Factors **52**, 381–410 (2010)
6. Billings, C.E.: Human-Centered Aviation Automation: Principles and Guidelines. NASA, Washington, DC (1996). NASA-TM-110381
7. Chen, J.Y.C., Barnes, M.J.: Human-Agent Teaming for Multi-Robot Control: A Literature Review. Army Research Lab Technical report, ARL-TR-6328 (2013)
8. The Atlantic. https://www.theatlantic.com/magazine/archive/2013/03/the-robot-will-see-you-now/309216/
9. Christoffersen, K., Woods, D.D.: How to make automated systems team players. In: Advances in Human Performance and Cognitive Engineering Research, vol. 2, pp. 1–12. Emerald Group Publishing Limited (2002)
10. Wiener, E.L.: Cockpit automation. In: Wiener, E.L., Nagel, D.C. (eds.) Human Factors in Aviation, pp. 433–461. Academic Press Inc., New York (1989)
11. Onken, R.:. The cockpit assistant system CASSY as an on-board player in the ATM environment. In: Proceedings of First Air Traffic Management Research and Development Seminar, pp. 1–26 (1997)
12. Lyons, J.B., Saddler, G.G., Koltai, K., Battiste, H., Ho, N.T., Hoffmann, L.C., Smith, D., Johnson, W., Shively, R.: Shaping trust through transparent design: theoretical and experimental guidelines. Adv. Hum. Factors Robot. Unmanned Syst. **499**, 127–136 (2017)
13. Sadler, G., Battiste, H., Ho, N., Hoffmann, L., Johnson, W., Shively, R., Lyons, J., Smith, D.: Effects of transparency on pilot trust and agreement in the autonomous constrained flight planner. In: Digital Avionics Systems Conference (DASC) IEEE/AIAA 35th, pp. 1–9. IEEE (2016)
14. Lee, D.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**, 50–80 (2004)

15. Lees, M.N., Lee, J.D.: The influence of distraction and driving context on driver response to imperfect collision warning systems. Ergonomics **50**, 1264–1286 (2007)
16. Hergeth, S., Lorenz, L., Vilimek, R., Krems, J.F.: keep your scanners peeled: gace behavior as a measure of automation trust during highly automated driving. Hum. Factors **58**, 5–27 (2016)
17. Endsley, M.R.: From here to autonomy: lessons learned from human-automation research. Hum. Factors **59**, 5–27 (2017)
18. Christoffersen, K., Woods, D.D.: How to make automated systems team players. In: Advances in Human Performance and Cognitive Engineering Research, pp. 1–12. Emerald Group Publishing Limited (2002)
19. Alberts, D.S., Garstka, J.J., Hayes, R.E., Signori, D.A.: Understanding Information Age Warfare. Command Control Research Program, Washington, DC (2001)
20. Chen, J.Y.C., Barnes, M.J., Harper-Sciarini, M.: Supervisory control of multiple robots: human performance issues and user interface design. IEEE Trans. Syst. Man Cybern.–Part C: Appl. Rev. **41**, 435–454 (2011)
21. Lyons, J.B.: Being transparent about transparency: a model for human robot interaction. In: AIAA Spring Symposium Series (2013)
22. Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Hum. Factors **49**, 57–75 (2007)

A Human-Autonomy Teaming Approach for a Flight-Following Task

Summer L. Brandt[1(✉)], Joel Lachter[2], Ricky Russell[1], and Robert Jay Shively[2]

[1] NASA Ames Research Center, San José State University, Moffett Field, CA 94035, USA
{Summer.L.Brandt,Ricky.Russell}@nasa.gov
[2] NASA Ames Research Center, Moffett Field, CA 94035, USA
{Joel.Lachter,Robert.J.Shively}@nasa.gov

Abstract. Human involvement with increasingly autonomous systems must adjust to allow for a more dynamic relationship involving cooperation and teamwork. As part of an ongoing project to develop a framework for human-autonomy teaming (HAT) in aviation, a study was conducted to evaluate proposed tenets of HAT. Participants performed a flight-following task at a ground station both with and without HAT features enabled. Overall, participants preferred the ground station with HAT features enabled over the station without the HAT features. Participants reported that the HAT displays and automation were preferred for keeping up with operationally important issues. Additionally, participants reported that the HAT displays and automation provided enough situation awareness to complete the task, reduced the necessary workload and were efficient. Overall, there was general agreement that HAT features supported teaming with the automation. These results will be used to refine and expand our proposed framework for human-autonomy teaming.

Keywords: Human-autonomy teaming · Ground station · Recommender system · ACFP

1 Introduction

Managing aircraft is becoming more complex with increasingly sophisticated automation responsible for more flight tasks. With this increased complexity, it is becoming more difficult for operators to understand what the automation is doing and why. Human involvement with increasingly autonomous systems must adjust to allow for a more dynamic relationship involving cooperation and teamwork.

As part of an ongoing project to develop a framework for human-autonomy teaming (HAT) in aviation [1], a part-task study was conducted to demonstrate, evaluate and refine proposed tenets of HAT. The HAT features were derived from three tenets and were built into an automated recommender system on a ground station. These HAT tenets include:

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_2

- *Bi-directional Communication*: For automation to act as a teammate, there needs to be bi-directional communication about mission goals and rationale. This requires a clear communication channel with a shared, understandable language [2].
- *Transparency*: Automated systems often do not facilitate understanding or tracking of a system [3]. Providing the automation's rationale for selecting particular actions helps the human understand what the automation is doing and why. Again, to be truly transparent, communication should use a shared language that matches the operators' mental model.
- *Operator Directed Interface*: While increased automation can help with manual performance and workload, recovering from automation failure is often worse [4, 5]. A dynamic allocation of tasks based on operator direction and context allows a much more agile, flexible system and a greater opportunity to keep the operator in the loop.

This study focused primarily on interactions with one piece of automation, the Autonomous Constrained Flight Planner (ACFP). The ACFP is an automated recommender system designed to support rapid diversion decisions for commercial pilots in off-nominal situations [6]. The ACFP was designed to generate a list of diversion options in a ranked order. It compiles information from several sources such as ATIS broadcasts, METAR weather reports, GPS location and terrain, aircraft condition, and airport/runway characteristics. Evaluations are made for various factors (e.g., risk associated with the enroute, approach, and landing phases of flight, fuel usage, weather, terrain, distance, facilities). These evaluations are then aggregated to produce an overall score. The initial implementation of the ACFP provided little transparency regarding the evaluation and weighting of these factors [7]. Much effort has gone into enhancing this tool not only in capability but also in transparency [8–10]. For the purpose of this study, participants used the ACFP at a ground station to reroute aircraft in situations such as inclement weather, system failures and medical emergencies. Participants performed this task both with HAT features enabled and without and provided feedback.

2 Method

2.1 Participants

Four dispatchers (median dispatch experience was 11 years) and two pilots (both active duty with over 10,000 h flown as a line pilot) participated in this simulation.

2.2 Simulation Environment

Our simulation ground station has been developed through a series of human-in-the-loop simulations to examine issues associated with collaboration between an onboard pilot and ground support from a dispatcher or ground pilot [11–13]. Each successive simulation advanced the ground station and evaluated a more fully evolved operational concept. Our current framework envisions a role for ground support in monitoring and assisting aircraft in an advanced flight following mode, which requires increasingly sophisticated automation and an opportunity to enhance collaboration between the operator and the automation.

The following sections describe the components of the ground station for multi-aircraft monitoring and support. The HAT features included for this simulation are called out separately.

Aircraft Control List. The center of the station hosts an Aircraft Control List (ACL), the primary tool for managing multiple aircraft and switching the focus between aircraft (see Fig. 1A). The ACL provides information crucial for situation awareness such as callsigns, departure and destination city pairs, estimated time of arrivals, flight plans, souls on aboard, and pilot details.

Fig. 1. Simulation ground station. A: Aircraft Control List (ACL), augmented with timeline, alerting information and HAT features. B: Traffic Situation Display (TSD). C: Flight controls and displays for the selected aircraft in read-only mode. D: CONUS map and charts.

This version of the ground station was designed to monitor, with the help of automation, a large number of aircraft (up to 30). Automated alerts are provided in three priority levels and two burdening levels. Priority refers to whether the situation must be addressed immediately (high shown in red), can wait before being addressed (medium shown in amber), or is merely advisory (low shown in green), corresponding respectively to flight deck warning, caution, and advisory alerts. Burdening refers to whether the operator is expected to take an action or if another agent or onboard pilot is responsible. In the current simulation, alerts were issued for failure to adhere to a clearance, failure to stay on path, environmental threats (weather on flight path or at the destination, and airport closures), system issues, and failure of the pilot to acknowledge a flight deck alert.

Creating an Operator Directed Interface with Plays. We have adopted the playbook approach to set system goals and manage roles and responsibilities between the operator and the automation [14]. Our first implementation provides 13 different plays the operator could call to address various scripted events. When the operator selects a play, the ACFP is triggered with preset weights and the corresponding checklist appears on the display identifying operator tasks in white and automation tasks in blue (see Fig. 2).

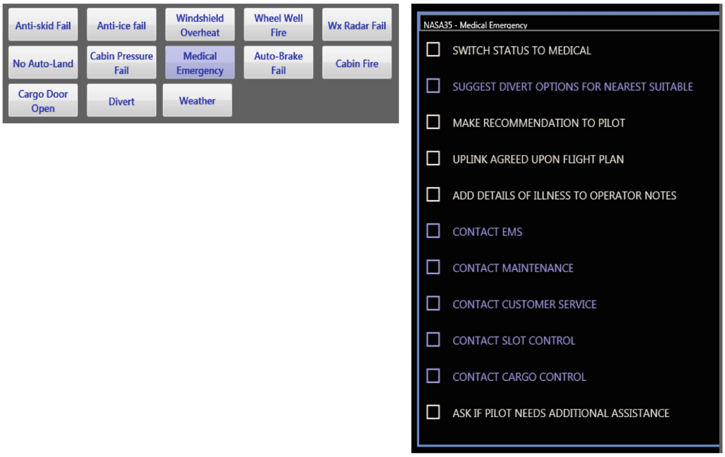


Fig. 2. Operators could call plays for various events and in the HAT condition were provided a checklist of roles and responsibilities. This is an example of a Medical Emergency play.

Building in Transparency. The ACFP takes into account more than just risk; it looks at fuel, distance, services available, etc. It also has the capability of weighting these factors differently based on the situation. We increased transparency by explicitly showing the factors and weights of the recommended divert airports when the ACFP is enabled (see Fig. 3). Additionally, we translated the scores for the ACFP factors to more meaningful numbers and descriptors for the operator (e.g., presenting nautical miles (nm) instead of a score). In the example above, a Medical Emergency play was called which resulted in the distance to medical facilities (Medical row) and time to destination (ETA row) given more weight than other factors. As a result, Cheyenne (KCYS) was the top recommendation showing a trauma care facility 3 nm from the airport. Although Denver (KDEN) was closer, the trauma care facility is further from the airport.

Building in Bi-directional Communication. We preset weights for each play and presented the weight settings (top of Fig. 3). The operator is able to negotiate with the system by changing these weights to better fit the situation. The operator can adjust the weights and see how the divert recommendation is affected. Again, using the example

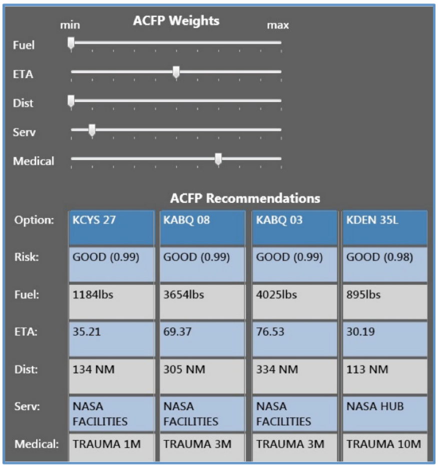


Fig. 3. In the HAT condition, operators were provided ACFP factors (on bottom) and weights (on top) to increase transparency and bi-directional communication. This is an example of a Medical Emergency play.

in Fig. 3, if the operator decided that distance to the airport or estimated time of arrival were a higher priority than available medical facilities, the operator could adjust the ACFP weights and re-run the query.

Traffic Situation Display. The Traffic Situation Display (TSD) is a 3D map display of company aircraft (see Fig. 1B). Information such as flight plans, trajectories and data tags are selectable. Color-coding from the ACL is maintained allowing the operator to, at a glance, identify the priority level of each aircraft. NextRad weather and turbulence boxes are graphically displayed.

Building in Transparency. The ACFP was augmented to display ATIS at the destination airport as well as indicate which of a number of risk factors are present in any potential divert location (see Fig. 4) [9, 10]. Operators could also request such ratings for airports that are not recommended by the tool. The recommended route to a given airport is displayed in grey on the TSD.

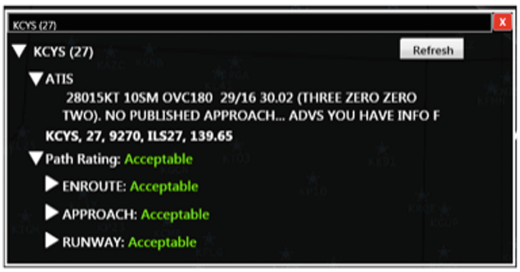


Fig. 4. When using the ACFP, regardless of condition, a transparency window appeared on the TSD for the recommended airport. An ATIS report, runway information, path rating and reasoning statements were also included.

Additional Displays. The left-side display contains aircraft flight controls and instrumentation for the selected aircraft (see Fig. 1C). A graphical display of the Flight Management System (FMS) through a GUI Control Display Unit (CDU) was carried over from previous builds of the ground station in part because dispatchers previously reported that seeing the flight controls improved their situation awareness. The controls are view-only and do not allow for any manipulation of the aircraft. Instrumentation and information displays include a Primary Flight Display (PFD) and some Engine Indication and Crew Alerting System (EICAS) functionality. The right-side display contains a CONUS map with an overlay of company aircraft and weather (see Fig. 1D). Airport charts are provided below the map.

Voice Interaction. Operators are able to perform some functions by voice, such as selecting specific aircraft and invoking the ACFP. Alerts and certain system changes (e.g., aircraft landing) are also announced vocally. All stations (simulated ground station and flight decks) are equipped with push-to-talk communication.

Airspace. Participants managed the ground station while confederates supported the simulation by piloting and initiating various off-nominal situations. Aircraft approximated 737–800s and were located in the western half of the U.S. The Multi Aircraft Control Systems [15] was used to simulate the airspace and aircraft.

2.3 Experimental Design

The experimental design consisted of a single fixed factor, HAT, and a random factor, Subject. There were two levels of HAT: HAT (ground station with HAT features enabled) and No HAT (ground station without HAT features). Participants performed the flight-following task once in each condition with the order of trials counterbalanced across participants. We collected behavioral and subjective data. Subjective data results are presented below; behavioral data are reported separately [16].

One participant was tested per day. Each participant received approximately 3.5 h of training before running two 50-minute experimental scenarios. Questionnaires were administered post-scenario and post-simulation. A debrief session was conducted post-simulation to gather additional feedback.

Participants were provided with a concept of operation where automation and ground personnel provide “another set of eyes” monitoring aircraft. The role of ground in this flight following task was to support aircraft in high workload and off-nominal situations. Our primary interest was in participant feedback of the HAT features.

Scenarios were developed to test HAT features in making diversion decisions under different weather conditions and emergency landing situations. Each scenario required participant ground operators to make approximately six diversions using the ACFP. Confederates supported the simulation by piloting and initiating various scripted off-nominal situations for the ground station operator. Example reasons for diversions include deteriorating weather conditions at the destination airport, a mechanical emergency such as an aft cargo door open, or a critically ill passenger needing immediate medical care.

At the start of a scenario, the operators had up to 30 aircraft to flight-follow after takeoff until landing. Shortly into the scenario, the ground station began alerting the operator to various situations. An event started either with an alert on the ground station, elevating the aircraft priority and queuing the operator to contact the pilot, or with a radio call from the pilot. In either case, if it was determined that the aircraft needed to divert to a new destination, the operator would invoke the ACFP by selecting the appropriate play. In the HAT condition, once the play was selected, a checklist of procedures appeared with the automation responsible for a certain set of identified tasks. In the No HAT condition, operators had a paper checklist available for procedure items. In both conditions, the ACFP provided multiple recommendations in rank order and the transparency window was displayed on the TSD for the selected airport (see Fig. 4). The ACFP would select the highest rated airport based on the event and related factors, though the operator could explore additional airports and view both the suggested route and transparency window. In the HAT condition, the ACFP factors and weights were displayed on the ACL providing additional transparency and allowing for manipulation of factor weights. When there was consensus on the new airport, the operator would datalink the route to the pilot who would, according to the concept of operations, contact air traffic control for approval. Operators could then determine to what extent they needed to follow that aircraft and either leave the aircraft priority elevated (amber or red) or reduce the aircraft priority (green).

3 Results

3.1 Post-scenario Comparisons

After each scenario, participants completed a post-scenario questionnaire rating agreement with (1 = strongly disagree, 5 = neither agree nor disagree, 9 = strongly agree) and confidence in (1 = no confidence, 5 = confident, 9 = extremely confident) several aspects of the ACFP. Paired-samples t-tests were run between the HAT and No HAT conditions. Given the small sample size and range of responses, few significant differences were found.

Little differences were found in participant ratings of reliance on the ACFP. In both the HAT ($M = 5.67$, $SD = 1.37$) and No HAT ($M = 5.17$, $SD = 2.40$) conditions, pilots neither agreed nor disagreed that they would rely on the ACFP recommendation without hesitation, $p = .41$. However, if faced with a very hard and time constrained task in the future, participants agreed they would rely on the ACFP in both the HAT ($M = 8.00$, $SD = 1.55$) and No HAT ($M = 7.33$, $SD = 1.63$) conditions, $p = .24$.

Although mean scores were not significantly different, four of the six participants agreed (ratings 7–9) that overall, the diversion decisions recommended by the ACFP were acceptable in the HAT condition ($M = 6.67$, $SD = 2.16$), compared to two of the six participants (ratings 8–9) in the No HAT condition ($M = 5.33$, $SD = 2.58$), $p = .12$. There was a significant difference in confidence where participants reported greater confidence that the diversions they chose were appropriate in the HAT condition ($M = 7.83$, $SD = 1.47$, all participants rating 6 and above) compared to the No HAT condition ($M = 6.33$, $SD = 2.07$, four participants rating 6 and above), $t(5) = 4.39$, *p* = .01. Five of the six participants reported confidence that the diversions recommended by the ACFP were appropriate in the HAT condition (*M* = 6.67, *SD* = 2.25) compared to three of the six participants in the No HAT condition (*M* = 5.33, *SD* = 2.42; *p* = .17).

3.2 Simulation Ratings

In addition to the post-scenario questionnaires, a final, post-simulation, questionnaire was administered after both trials were completed.

Display Preference. Participants were asked to rate their preferred displays and automation on a 1 = No HAT to 9 = HAT scale. Participants unanimously preferred the HAT displays. Specifically, HAT displays were preferred with regard to:

- keeping up with operationally important issues (*M* = 8.67, *SD* = 0.52);
- ensuring the necessary situation awareness for the task (*M* = 8.67, *SD* = 0.52);
- integrating information from a variety of sources (*M* = 8.67, *SD* = 0.52);
- reducing workload necessary for the task (*M* = 8.33, SD = 0.82); and
- efficiency (*M* = 8.33, *SD* = 0.82).

Participants were in agreement that overall they preferred interacting with the automation in the HAT condition (*M* = 8.50, *SD* = 0.55).

ACFP Recommendations. In additional to ratings of agreement, participants rated several items in terms of usefulness (1 = not useful, 5 = somewhat useful, 9 = very useful). Post-simulation, the ACFP was rated a useful tool (*M* = 7.33, *SD* = 1.37). Four of the six participants agreed that the recommendations were generally consistent with what s/he would have recommended (*M* = 5.83, *SD* = 2.56). The ACFP seems to be particularly helpful during emergency situations, as explained by one participant, *"Everything is easy and accessible in emergency situations. No need to consult many other programs to get various info."*

HAT Features. Items specific to the HAT tenets were rated both post-scenario and post-simulation.

Supporting Bi-directional Communication. Participants agreed that the ACFP weights improved the automation's ability to handle unusual situations (*M* = 7.83, *SD* = 1.60) and were useful in making divert decisions (*M* = 8.33, *SD* = 0.82). Participants liked having the weights (*M* = 8.33, *SD* = 1.21) and one participant commented that, *"[the display] gave me the ability to see why, gave me control to change weights in variable(s)."*

Building in Transparency. Participants agreed that the ACFP table was helpful in making divert decisions (*M* = 7.67, *SD* = 1.51) and they liked having the table (*M* = 8.33, *SD* = 1.03). One participant commented that, "*This [table] is wonderful… You would not find a dispatcher who would just be comfortable with making a decision without knowing why.*"

Creating an Operator Directed Interface with Plays. Participants liked having the electronic checklist for each play (*M* = 8.67, *SD* = 0.52) and did not prefer the paper to the electronic checklist (*M* = 2.67, *SD* = 1.97). One participant claimed that, *"The electronic list was easier because it was right there on the screen and it eliminated a couple of the steps."* Another participant was hesitant to rely solely on the electronic checklist and explained that s/he, *"found it necessary to have both on hand."*

Overall. Participants all agreed that they would like to have a tool like the ACFP with HAT features to use with real flights (*M* = 8.00, *SD* = 0.89).

Additional Displays. Participants reported consulting ATIS information while making a diversion decision in both the No HAT and HAT conditions, and half of the participants consulted airport charts while making a diversion decision (the other half did not consult the charts because they were already familiar with the diversion airport).

Voice Input. Participants did not prefer using voice input over mouse input (*M* = 4.17, *SD* = 3.13), with one participant commenting that the preference for mouse was *"due to reliability."* Participants slightly agreed that voice input would be more useful if it worked better (*M* = 6.33, *SD* = 2.07).

Voice Annunciations. Participants found the voice annunciations for alerts useful (*M* = 7.17, *SD* = 2.56) with one participant commenting that it, *"gave another dimension to alerting."* However, another participant cautioned that, "*the voice annunciations became somewhat mundane when announcing the new routes when planning for a diversion."* Participants found the voice annunciations for arrivals somewhat useful (*M* = 5.50, *SD* = 2.66), with one participant commenting that they, *"helped with flight following."* Overall, participants rated voice annunciations for aircraft for the ACFP as useful (*M* = 7.00, *SD* = 1.67).

4 Discussion and Conclusion

HAT is a promising solution for increasingly complex systems. Guidelines for improving HAT are growing, e.g., [4], and this project is our early effort at designing for human-autonomy teaming in an actual system. Our approach emphasizes three tenets: bi-directional communication between the human and automation, transparency for automation, and an operator directed interface.

We acknowledge this was an initial study with a small sample size, as our primary purpose was to demonstrate HAT features and gather feedback for further refinement. Overall, participant feedback was positive, supporting our implementation of HAT features. Participants liked having a recommender system with factors and weights and expressed interest in having similar automation for real flights. They valued the integration of the displays, commenting that this level of integration is not currently available. Participants found the electronic checklist useful as were a number of the alerts.

Suggestions for improvement were also provided, which we are working toward. In regards to the ACFP, participants expressed reluctance to allow the ACFP to make diversion decisions for them at this point. As one person put it, it still *"needs some TLC* *to be trusted."* We are continuing our work on how transparency affects trust [9, 10] and planning to incorporate the results into the ACFP. Furthermore, in our current implementation, plays included simple checklists. We are working towards making these more flexible with varying levels of automation and branch points. While the concept of voice input and annunciations was received well, our implementation is in its infancy. Voice input was unreliable, working well for some participants but not all, and voice annunciations lacked etiquette, speaking over the operator and pilot. Improvements are being made to increase the vocabulary and grammar and better manage the output to not interfere with operator tasks.

As a next step, we have ported some of these ground station tools to a tablet for use on a simulated flight deck and will again evaluate the tools with and without HAT features. Our goal is to develop a framework for HAT, consisting of tenets and guidelines for implementing them. We eventually hope to create software libraries that make this implementation easier.

Acknowledgments. We would like to acknowledge NASA's Safe and Autonomous System Operations Project, which funded this research.

References

1. Shively, R.J., Lachter, J., Brandt, S.L., Matessa, M., Battiste, V., Johnson, W.: Why human-autonomy teaming? In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (2017, this edition)
2. Goldman, C.V., Degani, A.A.: Team-oriented framework for human-automation interaction: implication for the design of an advanced cruise control system. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 2354–2358. SAGE Publications (2012)
3. Lyons, J.B.: Being transparent about transparency: a model for human robot interaction. In: AIAA Spring Symposium Series (2013)
4. Endsley, M.R.: From here to autonomy: lessons learned from human-automation research. Hum. Factors **59**, 5–27 (2017)
5. Onnasch, L., Wickens, C.D., Li, H., Manzey, D.: Human performance consequences of stages and levels of automation: an integrated meta-analysis. Hum. Factors **56**, 476–488 (2013)
6. Meuleau, N., Plaunt, C., Smith, D.E., Smith T.B.: An emergency landing planner for damaged aircraft. In: Proceedings of the Twenty-First Innovative Application of Artificial Intelligence Conference, pp. 114–121 (2009)
7. Dao, A.-Q., Koltai, K., Cals, S.D., Brandt, S.L., Lachter, J., Matessa, M., Smith, D., Battiste, V., Johnson, W.W.: Evaluation of a recommender system for single-pilot operations. Procedia Manuf. **3**, 3070–3077 (2015)
8. Lyons, J.B., Koltai, K.S., Ho, N.T., Johnson, W.W., Smith, D.E., Shively, R.J.: Engineering trust in complex automated systems. Ergon. Des. **24**, 13–17 (2016)
9. Lyons, J.B., Saddler, G.G., Koltai, K., Battiste, H., Ho, N.T., Hoffmann, L.C., Smith, D., Johnson, W., Shively, R.: Shaping trust through transparent design: theoretical and experimental guidelines. Adv. Hum. Factors Robot. Unmanned Syst. **499**, 127–136 (2017)

10. Sadler, G., Battiste, H., Ho, N., Hoffmann, L., Johnson, W., Shively, R., Lyons, J., Smith, D.: Effects of transparency on pilot trust and agreement in the autonomous constrained flight planner. In: IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), pp. 1–9. IEEE (2016)
11. Lachter, J., Brandt, S.L., Battiste, V., Ligda, S.V., Matessa, M., Johnson, W.W.: Toward single pilot operations: developing a ground station. In: Proceedings of the International Conference on Human-Computer Interaction in Aerospace, Santa Clara, CA (2014)
12. Brandt, S.L., Lachter, J., Battiste, V., Johnson, W.W.: Pilot situation awareness and its implications for single pilot operations: analysis of a human-in-the-loop study. Procedia Manuf. **3**, 3017–3024 (2015)
13. Lachter, J., Brandt, S.L., Battiste, V., Matessa, M., Johnson, W.W.: Enhancing Ground Support: Lessons from Work on Reduced Crew Operations. Cognition Technology & Work (in press)
14. Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Hum. Factors **49**, 57–75 (2007)
15. Prevot, T.: Exploring the many perspectives of distributed air traffic management: the multi aircraft control system: MACS. In: International Conference on Human-Computer Interaction in Aeronautics, HCI-Aero, pp. 23–25 (2002)
16. Strybel, T., Keeler, J., Mattoon, N., Alvarez, A., Barakezyan, V., Barraza, E., Park, J., Vu, K.-P., Battiste, V.: Measuring the effectiveness of human automation teaming in reduced crew operations. In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (2017, this edition)

Measuring the Effectiveness of Human Autonomy Teaming

Thomas Z. Strybel[1(✉)], Jillian Keeler[1], Natassia Mattoon[1], Armando Alvarez[1], Vanui Barakezyan[1], Edward Barraza[1], James Park[1], Kim-Phuong L. Vu[1], and Vernol Battiste[2]

[1] California State University Long Beach, Long Beach, CA, USA
{Thomas.Strybel,Kim.Vu}@csulb.edu,
Jillian.Keeler@student.csulb.edu,
nat.mattoon@gmail.com, mandoalva9@gmail.com,
barraza.edward@gmail.com, jyp.718@gmail.com,
vanuibarakezyan@yahoo.com

[2] San Jose State University Foundation and NASA Ames Research Center, Moffett Field, CA, USA
vernol.battiste@nasa.gov

Abstract. We examined two metrics for assessing Human-Autonomy-Teaming (HAT) performance, subjective workload and eye-gaze durations, in a simulation of dispatcher flight following while interacting with the HAT features of an Autonomous Constrained Flight Planner. Operator workload was lower in the HAT condition and decreased with time in the scenario. However, participants took more time to uplink flight plan changes in this condition.

Keywords: Human-Autonomy-Teaming · Autonomous Constrained Flight Planner

1 Introduction

According to a recent report from the National Research Council's (NRC) Committee on Autonomy Research for Civil Aviation Aeronautics [1], increasingly autonomous aircraft systems are being developed at an accelerating pace, with the expectation that these systems will significantly benefit air travel in terms of safety, reliability, efficiency and affordability. These autonomous systems are expected to perform more complex, mission-related tasks with substantially less human intervention for longer durations and at remote distances.

These increasingly autonomous aircraft systems are made possible by recent advances in automation capabilities in sensing, decision making and planning. They have transformed thinking about the role of automation from the view of automation as replacing or assisting human operators to new concepts of human-autonomy interaction—the view that an autonomous system can be a valuable team member and have a well-defined role on a team. This new concept is called Human Autonomy Teaming (HAT).

HAT is currently recognized as a promising solution to the problems of human operators managing increasingly complex work systems. A human-autonomy team can

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_3

be defined as the interdependent coupling between one or more human operators and one or more autonomous systems requiring collaboration and coordination to accomplish system and task goals [2]. As such, it is being developed and pursued in many operational areas such as robotics [3], commercial aviation [4], and UAS operations [5]. Additional information on the tenets of HAT can be found in other papers in this volume (see e.g., [6, 7]).

The development of effective HAT agents and optimal interfaces for HAT interactions will be necessary to overcome some of the barriers to the success of increasingly autonomous systems [1]. The design and development of these agents is in its infancy stages; nevertheless, effective agents must consider the factors known to influence HAT effectiveness.

1.1 Factors Determining HAT Effectiveness

In aviation, many factors will determine the extent to which HAT will achieve the desired level of effectiveness when employed in the National Airspace System (NAS). These factors are based on previous research on human-automation interaction, human-machine interface design, and human-human team performance.

Automation Transparency. The operator must understand the automation agent's intent and reasoning, and be able to elicit additional factors on which the agent has based a recommendation for a course of action. Operators must be able to see what the agent is doing and, ideally, predict what the agent will do next. Transparency requires operator knowledge of the general logic of how the agent works and accurate mental models of its functioning [8]. At the same time, the autonomous agent must understand (to some extent) the current preferences, attitudes and states of human operators on the team. Detecting and understanding the status of human team members may be one of the more challenging aspects of autonomous-agent design in HAT.

Bi-directional Communications. Communications between humans and autonomous agents are necessary for establishing and maintaining a shared knowledge of task goals, exchanging information, reporting errors (human- or autonomy-based) and reporting status. Natural-language processing capabilities that match human capabilities would appear to be important for effective human-autonomy collaborations in pursuit of a common goal. In aviation, a logical starting point for human-autonomy communication and collaboration is current Crew Resource Management (CRM) principles, as these principles regulate communications between human team members [9]. Good CRM between humans requires team members to understand what the others are doing and why. Effective communication also means that the human operator can easily and accurately direct the automation, and override autonomous-system decisions, if necessary. The method of communication will also affect an operator's trust in automation, situation awareness and workload [10].

Trust in the Autonomous Agent. Trust in automation is a complex psychological state that involves beliefs, attitudes and expectations about the reliability and other characteristics of an automation tool to produce a favorable outcome. Trust in automation involves the willingness to place oneself in a position of uncertainty and vulnerability in the expectation that the agent will do what it is supposed to do, or communicate to the human why it is unable to do so. Human trust must be appropriately calibrated: over trust and under trust can lead to issues of complacency and increased workload [11].

Operator Workload. If poorly designed, human interactions with autonomous agents can produce excessive operator workload because of difficulty understanding the autonomous agent's current thinking or current tasks goals, awkward communication methods, and lack of trust in the autonomous agent [12, 13].

Situation Awareness of Humans and Autonomous Agent. Human operators must have awareness of the tasks, systems and environments, as well as adequate awareness of the status of the autonomous agent. Similarly, the agent must have adequate awareness of the system state, environment, and status of human team members [3, 14].

Individual Differences. Differences between human operators with regard to skills in multitasking, attentional control, and spatial processing, communication styles, personalities and trust in automation will affect HAT effectiveness. It is therefore important to design HAT taking into account these individual differences [3].

1.2 Measuring HAT Effectiveness

The desired incremental gains in safety and efficiency produced by HAT could be less than expected unless they are designed properly, taking into account the factors noted above, as well as, design principles for promoting effective HAT interactions [14]. A poorly designed HAT might even create new risks and hazards that threaten the safety and efficiency of the NAS; therefore, it is important to measure the impact of HAT on operator and system outcomes. Evaluations of HAT designs necessitate methods and metrics for assessing effectiveness and efficiency, yet such metrics are currently lacking. The effectiveness of HAT must be measured across several dimensions, for example, mission effectiveness, automation behavior, human operator performance and human-autonomy collaboration. Proposed methods for measuring HAT outcomes can be classified in terms of system outcomes, automation behavior efficiency, human behavior precursors, and collaborations [15, 16].

At the highest level, the HAT agent must be evaluated for its effect on NAS operations. NAS performance is typically assessed with measures of safety (e.g., LOS, average separation between aircraft) and efficiency (e.g., system complexity, traffic count). Aircraft system performance is measured with individual aircraft parameters (e.g., flight plan deviations, fuel consumption). Methods and metrics for evaluating the impact of HAT agents on NAS performance must be developed and tested with simulations involving all major NAS operators (e.g., pilots, dispatchers, air traffic controllers) and these simulations must be sensitive to changes brought about by the introduction of HAT.

Focusing on system outcomes, however, fails to provide insights into the processes that lead to these outcomes. NAS safety and efficiency are affected by operator and crew performance, meaning that demonstrations of the HAT agent's effectiveness will require measures of performance at these levels. Traditional metrics for operator and crew performance include operational errors and adherence to CRM principles. Moreover, human behavior precursors such as workload and situation awareness, known to affect operator performance also must be assessed. Although many methods for assessing workload [17] and situation awareness [18] have been validated, these must be reassessed for their sensitivity in eliciting changes brought about by HAT.

Finally, collaborative metrics for evaluating the quality of the human-autonomy interactions must be determined. These should be related to the unique coordination efforts between human and autonomous agents, and the extent to which they assess the desired characteristics noted above. There are subjective methods for assessing human trust in automation [19] and subjective methods for measuring human-human team collaborations [20] but measures based on operator behavior and performance are lacking. These measures are complicated because they involve designing scenarios and tasks that are sensitive to differences in behaviors and performance, as well as generalizing to the operational context. For example, Higham et al. [21] measured air traffic controllers' trust in an advanced conflict detection and resolution tool by developing scenarios that contained near-miss events in which aircraft pairs approached but never lost separation. Trust in the automation tool was shown when the controllers took no action. Additional metrics for assessing the effectiveness of human-autonomy interactions include attention allocation, timesharing efficiency and decision-making, as well as measures of communication and collaboration between humans and HAT agents [15].

Behavioral and performance measures of HAT efficiency have been suggested but not tested. Hancock et al. [22], for example, proposed "neglect tolerance," the degree to which the autonomous agent is ignored, as a measure of trust in automation. Muir and Moray [23] suggest that operators engage in "defensive monitoring," which is the operator's rate of sampling of the autonomous agent: operators can employ sampling rates that are in either excess of, well below, or optimal with respect to the agent's characteristics.

This paper reports on a preliminary investigation of possible metrics for assessing HAT effectiveness as part of a demonstration of a HAT agent used by ground station operators in a Reduced Crew Operations project that was conducted by NASA Ames' Human Autonomy Teaming Laboratory. In this paper, we examined two metrics for assessing HAT effectiveness and subjective workload to determine if HAT interactions changed operator workload, and eye-gaze durations while interacting with the HAT features of an Autonomous Constrained Flight Planner (ACFP). For additional details on the purpose, methods, and results of this demonstration, see [6, 7].

2 Method

2.1 Participants

Six participants were tested. Four were experienced dispatchers and two were air transport pilots. None had previous experience with the advanced ground station used in the project.

2.2 Ground Station Configuration

The baseline autonomous agent was developed for an advanced dispatch station in which the dispatcher follows flights, as is done today, but can also provide first-officer assistance to simulated aircraft flown by single pilots or reduced crews. As such, the ground station has some capabilities currently used by dispatchers. The dispatch operator, with automation-agent assistance, is responsible for additional duties (e.g., flight planning, system diagnoses, and communications) that are traditionally performed by the flight crew. The advanced ground station was a component of a simulation environment comprising external systems that provide traffic simulation, communications, air traffic control, and multi-pilot interactions.

Figure 1 (left) shows an example of the layout of the ground station. Displays on the right (CONUS Map and Jepson Charts) are current-day dispatcher displays. The Traffic Situation Display (TSD) provides an untethered view of all of the traffic and weather within a range defined by the operator. The Aircraft Control List (ACL) contains a list of callsigns, city pairs and alert information for each aircraft being followed. The list is dynamically updated as aircraft enter and exit the dispatcher's responsibility. Autonomous Constrained Flight Planner (ACFP) is a flight-planning recommender system for assisting operators in generating and evaluating flight routes. The Aircraft Monitoring and Management System (AMMS; not shown) monitors all aircraft currently assigned to the dispatch station for clearances and path adherence, EICAS messages and route quality. The flight controls display shows the cockpit instrumentation of the aircraft currently selected by the dispatcher.

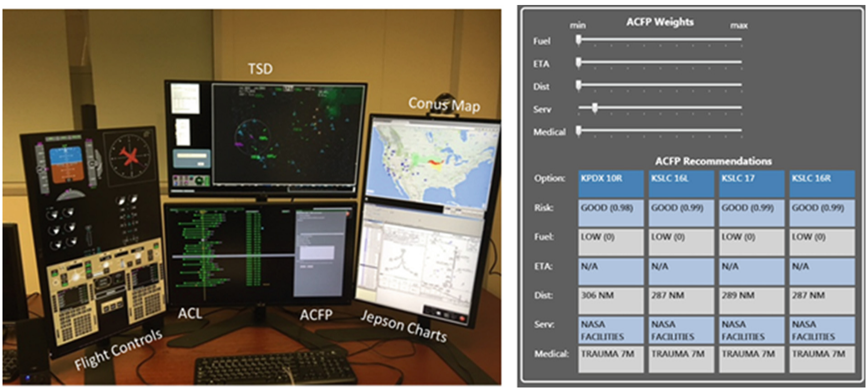


Fig. 1. *Left*. Advanced ground station used in the simulation showing Aircraft Control List (ACL), and ACFP (bottom center), Traffic Situation Display (TSD); (top center), flight controls and displays (left), CONUS map (top right) and charts (bottom right). Cameras were mounted on the center, left and right displays to record participant gaze while responding to an event. *Right.* ACFP display in HAT condition showing the recommended flight plans, criteria used in generating the recommendations and the sliders for adjusting the weights associated with each factor.

2.3 Procedure

Participants assumed the role of an ‘advanced dispatcher.” They performed a flight following task and supported aircraft that were assumed to have only one onboard pilot operating in high workload and off-nominal situations. Aircraft in the scenarios flew across the western United States, and in each scenario, produced events that might require dispatcher assistance (see Table 1). Each participant completed two scenarios, containing approximately 40 aircraft, and 6 off-nominal events. When an event occurred, the participant would select a “play” for the event, which brought up a checklist showing dispatcher, autonomous agent and pilot responsibilities. The participant determined if a flight deviation was needed, and when necessary, used the ACFP to generate recommended flight plan deviations that could be uplinked to the aircraft. Two conditions were tested based on whether the ACFP was equipped with HAT interaction features. In both conditions, the ACFP provided up to four flight plan recommendations in ranked order, and showed partial reasoning for the suggestions consisting of the risk associated with each recommendation and the factors (Fuel, ETA, Services, Distance, Medical Services) used to arrive at the recommendation (see Fig. 1, right). In the HAT condition, the ACFP also provided the dispatcher with the ability to modify the weighting of these factors by showing sliders that indicated graphically the current weight for each factor (Fig. 1, right). The participant could adjust the weights used in generating a recommendation by moving the sliders and requesting updated recommendations.

Table 1. Scenario events requiring dispatcher assistance

Aircraft event	Description
Fire in lavatory	Fire ignited in the lavatory; immediate landing is likely
Airport weather	Weather at destination airport is near or below minimums. If below, pilot must divert to a suitable nearby airport
Wheel well fire	Fire detected in the main wheel well shortly after takeoff
Medical emergency	Passenger onboard an aircraft requires medical attention and possibly immediate landing based on severity of condition
Anti-skid inoperative	Locked wheel protection is unavailable. Antiskid prevents wheels from skidding during braking by minimizing speed difference between wheel speed and aircraft speed
Windshield overheat	Windshield heating system has malfunctioned and may cause damage to the windshield
Aft cargo door open	One or more cargo doors are not closed and secure; detrimental if aircraft is above 8,000 ft
Weather radar fail	Failure prevents cockpit crew from viewing weather near the aircraft

2.4 HAT Metrics

This preliminary demonstration of a HAT agent provided limited opportunities for assessing its effectiveness. Nevertheless, we assessed subjective workload at regular intervals during the scenario and analyzed the participant's eye gaze to determine whether the duration eye gazes between the HAT and No HAT conditions differed. Subjective workload ratings were measured every 8 min beginning at 6 min in the scenario.

Eye gaze data was obtained from four cameras mounted on advanced ground station monitors. Data was collected and analyzed in both conditions only for events resulting in a flight plan change based on the ACFP. The duration of the event was the time between activation of a play and uplink of a flight plan. This event time was subsequently broken down by the amount of time spent gazing at each of the five displays (Flight Controls, TSD, AMMS+ACFP, CONUS, and Jepson Charts) during the event, by examining the position of the participant' sclera relative to each camera.

3 Results

HAT effectiveness was measured by workload ratings obtained at regular intervals in the scenario. For the HAT condition, we counted the number of times the participant adjusted the weight factors prior to uploading a flight plan. Finally we examined the amount of time spent gazing at various displays while resolving an event.

Overall, workload ratings were low. One subject did not respond to workload queries throughout the simulation; therefore, a repeated measures ANOVA was performed on workload ratings for the remaining five participants with the factors HAT condition (HAT vs. No HAT) and time in scenario. As shown in Fig. 2, workload ratings in the HAT condition (M = 1.65, SEM = 0.93) were lower overall than ratings in the No- HAT condition (M = 2.33, SEM = 1.24), although this difference only approached significance ($p = .13$). The interaction between HAT condition and time interval was significant, however, $F(5,11) = 3.42$; $p = 0.04$ (see Fig. 2). For the HAT

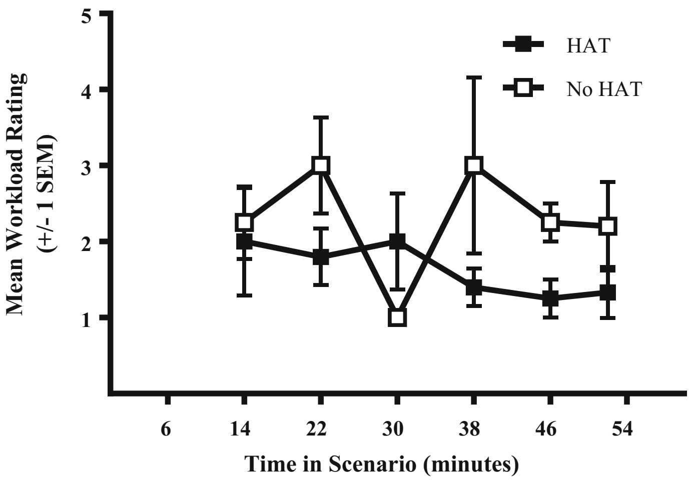


Fig. 2. Workload ratings for HAT and No HAT conditions at regular intervals in the scenarios.

condition, workload ratings were generally lower, and decreased slightly with time in the scenario. Workload ratings for the No-Hat condition were more variable but were on average unchanged.

Table 2 summarizes differences between conditions in terms of event and eye-gaze durations. From this table, it can be seen that participants did not always uplink a flight plan change for each event. On average only three flight-plan changes were made per scenario under each condition. However, in the HAT condition, the amount of time to select and uplink a flight change was significantly longer by 40 s in the HAT condition, compared with the No HAT condition. The increase in event times was due to participants gazing significantly longer at the ACFP (where the recommendations and weights were displayed) and the TSD (see Table 2) in the HAT condition. Some of the increased time must have been caused by slider movement; however, when restricting gaze times to events in which no slider movement occurred the difference between the mean gaze time for HAT was still higher (and marginally significant) compared with the No HAT condition. Very little time was spend gazing at the remaining displays in both conditions (Table 2).

Table 2. Summary of flight plan changes in response to off-nominal events and time spend on ground station displays for HAT and No HAT conditions

Summary statistic	HAT	No HAT	p*
Mean (SEM) number flight plan changes	3.3 (0.8)	3.0 (0.6)	.78
Mean (SEM) time per flight plan change (s)	101.2 (19.4)	63.7 (14.9)	.026
Mean (SEM) gaze time on ACFP (s)	53.6 (9.5)	27.5 (3.5)	.009
Mean (SEM) gaze time on ACFP - sliders not moved in HAT condition (s)	41.9 (8.7)	27.5 (3.5)	.08
Mean (SEM) gaze time on TSD (s)	38 (14.0)	20.7 (11.0)	.10
Mean (SEM) gaze time on other displays (s)	11.9 (13.0)	9.05 (12.3)	.28

*probability of obtained difference based on repeated measures t test ($df = 5$)

Finally, we examined the HAT collaborations by counting the number of times each participant adjusted the weights in the HAT condition. As shown in Table 3, these weights were rarely adjusted, and these adjustments were restricted mostly to one off-nominal event. Five (of 6) participants adjusted weights for the "Fire in Lavatory" event. Two participants adjusted weights for "Airport Weather" and one participant for "Wheel Well Fire" and "Medical Emergency." Table 3 also shows that in the majority of cases, participants increased the weight of the Distance factor. One participant increased the weight for ETA, and two participants increased both Distance and ETA for "Medical Emergency" and "Airport Weather." For the remaining events in the HAT condition, participants simply accepted and uplinked the initial recommendation of the ACFP. Note also in Table 3 that one participant adjusted the weights for all events listed, and one participant never used the weights in the HAT condition.

Table 3. Number of participants who adjusted factor weights by event in the HAT Condition.

Event	N participants	Factor(s) adjusted
Fire in lavatory	5	Distance
Airport weather	2	Distance or ETA
Wheel well Fire	1	ETA
Medical emergency	1	Distance and ETA

4 Discussion

We examined subjective and behavioral indicators of HAT collaborations using a preliminary proof-of-concept demonstration of HAT principles for Reduced Crew Operations [6, 7]. This demonstration simulation was not designed specifically for experimental comparisons of HAT vs. No HAT conditions. Nevertheless, we obtained preliminary data that suggest potential advantages and disadvantages of HAT. More important, we identified recommendations for future simulations in which HAT can be experimentally evaluated.

First, we showed that operator workload was lower in the HAT condition and decreased with time in the scenario. Seeing that participants completed only one scenario with the HAT agent, this may point to the need for additional training. In the No-HAT condition, workload was higher and roughly constant throughout. These results on workload obtained during a scenario agree with workload data obtained at the end of the scenario [6].

Second, although based on a limited sample, we showed that participants took more time to uplink flight-plan changes in the HAT condition, even when no weighting adjustments were made. On the surface, this contradicts our findings regarding workload; longer task times typically indicate higher workload. However, given that workload overall was very low, and that very few flight plan changes were made in each scenario, it is possible that the increases in event times were insufficient to change subjective workload. If this is the case, we speculate that in the HAT condition, providing an opportunity to alter the recommendations of the HAT agent led participants to consider the initial recommendations more carefully before either accepting or adjusting them. In the No-HAT condition, if participants rejected recommendations of the ACFP, he or she would have to create a flight deviation with no automation assistance, and this may have led participants to accept recommendations with less prior evaluation. Of course, these speculations must be verified with further research.

Our findings point out the difficulty in establishing behavioral and performance metrics for evaluating the effectiveness of HAT collaborations. On the one hand, longer event times might indicate higher workload and less efficient operations. Alternatively, longer event times might indicate more thorough processing, indicating that the HAT agent is improving operator and system performance. To understand the effect of HAT agents on operator performance and the quality of HAT collaboration, requires methods and measures be carefully developed and tested. We offer the following recommendations in this regard:

Evaluations of HAT must be performed at all levels of the system, and use a battery of assessment techniques. Evaluations must include system, autonomous-agent, and operator performance as well as behavioral precursors to operator performance [12]. The latter measures include workload, situation awareness and trust in automation. HAT evaluations should also utilize subjective, behavioral and performance measures of effectiveness. Subjective measures are simplest; they can be administered at different times during a scenario and capture the operator's perspective on the construct of interest. Note that, in our preliminary evaluation of the HAT agent for RCO, weight adjustments were made by only one of our six participants. Possibly, this participant trusted HAT recommendations more than the remaining participants. On the other hand, it is possible that the HAT recommendations were consistent with what most participants would recommend. Subjective measures of HAT effectiveness are currently unavailable; however, validated subjective measures of human-human collaborations do exist [20], and they represent a reasonable starting point in this regard.

Behavior outcomes of HAT effectiveness will require a priori knowledge of how experts currently perform tasks because, as found in the current investigation, one must determine the effect of behaviors on overall performance outcomes. This will require task analysis, verbal protocols, and other methods to determine behaviors showing trust in automation, communication, and collaboration. Finally, performance measures of HAT effectiveness will necessitate carefully designed tasks and scenarios that are sensitive to differences brought about by a HAT agent.

Acknowledgements. This research was supported by the NASA cooperative agreement #NNA14AB39C "Single Pilot Understanding through Distributed Simulation (SPUDS)," R.J. Shively, Technical Monitor.

References

1. National Research Council: Autonomy Research for Civil Aviation: Toward a New Era of Flight. National Academies Press, Washington D.C. (2014)
2. Langan-Fox, J., Canty, J.M., Sankey, M.J.: Human-automation teams and adaptable control for future air traffic management. Int. J. Ind. Ergon. **39**, 894–903 (2009)
3. Chen, J.Y.C., Barnes, M.J.: Human-Agent Teaming for Multi-Robot Control: A Literature Review (ARL-TR-6328). Human Research and Engineering Directorate, Aberdeen Proving Grounds (2013)
4. Shively, R.J, Lachter, J., Brandt, S., Matessa, M., Battiste, V., Johnson, W.W.: Why human-autonomy teaming? In: Proceedings of the 8th Annual International Conference on Applied Human Factors and Ergonomics (AHFE 2017), Los Angeles, CA (this volume)
5. Fern, L., Shively, R.J.: A comparison of varying levels of automation on the supervisory control of multiple UASs. In: Proceedings of AUVSI's Unmanned Systems North America, Washington, DC (2009)
6. Brandt, S., Lachter, J.B., Russel, R., Shively, R.J.: A human-autonomy teaming approach for a flight-following task. In: Proceedings of the 8th Annual International Conference on Applied Human Factors and Ergonomics (AHFE 2017), Los Angeles, CA (this volume)

7. Lachter, J.B., Brandt, S., Sadler, G. Shively, R.: Beyond point design: general pattern to specific implementations. In: Proceedings of the 8th Annual International Conference on Applied Human Factors and Ergonomics (AHFE 2017), Los Angeles, CA (this volume)
8. Endsley, M.R., Jones, D.G.: Designing for situation awareness: an approach to human-centered design, 2nd edn. CRC Press, New York (2012)
9. Matessa, M.: Using a crew resource management framework to develop human-autonomy teaming measures. In: Proceedings of the 8th Annual International Conference on Applied Human Factors and Ergonomics (AHFE 2017), Los Angeles, CA (this volume)
10. Joe, J.C., O'Hara, J., Medema, H.D., Oxstrand, J.H.: Identifying requirements for effective human-automation teamwork. In: Probabilistic Safety Assessment and Management PSAM 2012, Honolulu, HI (2014)
11. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: Humans: still vital after all these years of automation. Hum. Factors Gold. Anniv. Spec. Issue **50**, 511–520 (2008)
12. Cummings, M.L., Stimpson, A., Clamann, M.: Functional Requirements for Onboard Intelligent Automation in Single Pilot Operations. AIAA 2016-1652. AIAA Infotech @ Aerospace, San Diego (2016)
13. Chen, J.Y.C., Barnes, M.J., Harper-Sciarini, M.: Humans: still vital after all these years of automation. IEEE Trans. Syst. Man Cybern.—Part C: Appl. Rev. **41**, 435–454 (2011)
14. Endsley, M.R.: From here to autonomy: lessons learned from human–automation research. Hum. Factors **59**, 5–27 (2016)
15. Donmez, B., Pina, P.E., Cummings, M.L.: Evaluation criteria for human-automation performance metrics. In: Madhavan, R., Tunstel, E., Messina, E. (eds.) Performance Evaluation and Benchmarking of Intelligent Systems, pp. 21–40. Springer, Heidelberg (2009)
16. Birsen, D., Pina, P.E., Cummings, M.L.: Evaluation criteria for human-automation performance metrics. In: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems. Gaithersburg, Maryland, pp. 77–82 (2008)
17. Battiste, V., Bortolussi, M.: Transport pilot workload: a comparison of two subjective techniques. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **32**(2), 150–154 (1988)
18. Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D., Ladva, D., Rafferty, L., Young, M.: Measuring Situation Awareness in complex systems: comparison of measures study. Int. J. Ind. Ergon. **39**(3), 490–500 (2009)
19. Jian, J., Bisantz, A., Drury, C.: Foundations for an empirically determined scale of trust in automated systems. Int. J. Cogn. Ergon. **4**, 53–71 (2000)
20. Dickinson T.L., McIntyre R.M.: A conceptual framework for teamwork measurement. In: Team Performance Assessment and Measurement: Theory, Methods, and Applications, pp. 19–43. Lawrence Erlbaum (1997)
21. Higham, T., Vu, K.-P., Miles, J., Strybel, T.Z., Battiste, V.: Training air traffic controller trust in automation within a NextGen environment. In: Yamamoto, S. (ed.) HIMI/HCII 2013, Part II. LNCS, pp. 76–84. Springer, Heidelberg (2013)
22. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in automation in human-robot interaction. Hum. Factors **53**, 518–527 (2011)
23. Muir, B.M., Moray, N.: Trust in automation: Part II. Experimental studies of trust and human intervention on a process control simulation. Ergonomics **39**, 429–460 (1996)

Beyond Point Design: General Pattern to Specific Implementations

Joel Lachter[1(✉)], Summer L. Brandt[2], Garrett Sadler[3], and R. Jay Shively[1]

[1] NASA Ames Research Center, Moffett Field, CA 94035, USA
{Joel.Lachter,Robert.J.Shively}@nasa.gov
[2] NASA Ames Research Center, San José State University, Moffett Field, CA 94035, USA
Summer.L.Brandt@nasa.gov
[3] HATS Inc., 20944 Sherman Way, Suite 211, Canoga Park, CA 91303, USA
Garrett.Sadler@hats.solutions

Abstract. Elsewhere we have discussed a number of problems typical of highly automated systems and proposed tenets for addressing these problems based on Human-Autonomy Teaming (HAT) [1]. We have examined these principles in the context of aviation [2, 3]. Here we discuss the generality of these tenets by examining how they might be applied to photography and automotive navigation. While these domains are very different, we find application of our HAT tenets provides a number of opportunities for improving interaction between human operators and automation. We then illustrate how the generalities found across aviation, photography and navigation can be captured in a design pattern.

Keywords: Human-Autonomy Teaming · Automation · Human factors

1 Introduction

1.1 Problems with Highly Automated Systems

Elsewhere we have discussed a number of problems typical of highly automated systems [1]. Such systems are brittle, working properly within some bounded space for which they have been programmed, then failing when parameters fall outside that space. They are opaque, lacking transparency; human operators often do not know what the automation is doing or why. Operators often do not know when to trust automation, relying on it to handle conditions it cannot, or not taking advantage of it to handle conditions it can. As automation does more of the work, operators become less practiced. When the automation performs a task, the operator is often less aware of the system state.

While each of these issues is troubling by itself, they often manifest together. An operator, over-trusting the system, does not realize that some parameter has gone out of bounds. Because it is out of bounds, the automation either quits or is no longer reliable. The out of practice operator must then try to regain situation awareness using opaquely

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_4

presented information. This situation has been responsible for a number of accidents (e.g., Air France 477, Korean Air 801). Similar concerns have been identified by others [4–6].

1.2 HAT Solutions

A number of authors have suggested that these concerns are ameliorated by developing interfaces and procedures analogous to guidelines for improving teamwork among humans [5, 7]. This area of research has been termed "Human-Autonomy Teaming" or "HAT." Our current approach to HAT emphasizes the following tenets:

Bi-directional Communication: Following the common aviation crew resource management (CRM) practices that encourage input from all relevant parties into decisions, communication should be bi-directional concerning all levels of planning and execution. At the highest level (the mission), this means that operators should be explicitly informing the automation of mission goals, and automation should be able to recognize when those goals are not being met and inform the operator (preferably with an alternate course of action). At the lowest level (implementation), the operator should be able to "see" what the automation is doing and understand why, propose adjustments and have the automation report the predicted consequences of those adjustments before execution. Similar bi-directional communication should occur at intermediate levels. To facilitate this dialog, automation should be able to present a rationale for recommendations and warnings, for example, indicating why a route is rated as unacceptable or what event triggered a warning light. In addition, under conditions of uncertainty, automation should indicate its confidence in the data or analysis being presented. This allows operators to better integrate information that may be significant but unreliable (e.g., when I need to leave for work might be significantly different if I want an 80% chance of being on time versus a 99% chance).

Transparency: Providing rationale and confidence levels also fall under a more general tenet, that automation should be transparent. Lyons [8] defines automation transparency as a shared awareness and shared intent between the system and its human operator(s). That is, the system and its operators should be able to recognize what the human/automated teammate is doing and why. To be truly transparent, communication should use a shared language: Automation should present information in ways that fit the operator's mental model.

Operator Directed Interface: Interfaces should allow for dynamic task allocation directed by the operator. In particular, intent inferencing about the operator's state or goals should be minimized.

1.3 The Value of a Generalizable Solution

Elsewhere we discuss our efforts to demonstrate and test these principles implemented in an airline dispatcher ground station designed for flight following [2, 3]. But to what degree can the success of that implementation be attributed to the HAT principles as opposed to other design considerations? One way to think about this problem is "How easy would it be to apply the HAT system we developed for our ground station to a novel situation?" This is the "turnaround test" discussed by Woods [9]. We are attempting to address this generalizability issue by first developing use cases that specify what HAT would look like in a broad variety of domains, and second extracting patterns from those use cases that capture generalities about HAT. In the following sections, we examine two such use cases and one such pattern.

2 HAT in Other Domains

2.1 Photography

Photography provides a very different set of use cases for automation and HAT from those associated with aviation. In photography the operator has much more latitude in setting goals and in many cases may have to change these goals quickly to capture fleeting images. Automation can help. It can focus and adjust exposure (the brightness of the picture), much more quickly and accurately than any human operator. However, first it must know the goal. Automation can also detect and react to items and events of interest such as faces, blinks, and camera shake. Thus the automation is capable of doing much of the work to realize the operator's goal for a shot, if that goal can be accurately conveyed. In this section we briefly discuss some of the choices facing a photographer and then discuss how these choices might be more quickly and accurately conveyed to the automation.

Perhaps the most obvious choice the photographer makes is what to focus on. Ignoring the issue of pointing the camera, there are typically many objects within the field of view that the operator could choose to be in focus. Objects closer or further than the chosen object will appear increasingly blurry in the final image. The issue here is that, while it is usually obvious to the operator what the picture is of, the autofocus system does not know. On many cameras, the autofocus system can be told to look for the closest object or faces, but those abilities are quite brittle, braking when there are multiple faces or the subject is not the closest object in the scene. Cameras come with these priorities because portraits and subjects in the foreground are common goals. But where does that leave someone taking pictures of flowers or waterfalls? It would be nice to have a more generalized solution.

Choosing what to focus on is, of course, not the only choice that must be made when taking a photograph. How bright (or dark) should the photograph be? How do we control the brightness? Without turning this paper into a Photography 101 textbook, let us just say that these are complex questions. Controlling the brightness of a photograph has side effects for how sharp the image appears; side effects that may be desirable, or not, depending on the goals of the photographer. While automation can determine the appropriate level of brightness quickly, different settings affect how it controls the brightness with major effects on the appearance of the final product.

We would like to mention one last type of automation that is creeping into some cameras. Cameras can read information off of the sensor and process it far more quickly than a human operator. As a result, it is possible to have the camera trigger (or briefly suppress triggering) based on certain events. Cameras can take a picture when lightening flashes, or, under flickering lights, at the brightest point of cycle. Cameras can momentarily delay firing when the subject blinks or when your hand is moving. This raises the possibility of having a camera fire at specific moments that the operator wants to capture, say, when a bat hits a ball.

Now imagine taking a hike with your son. You want different settings when you take a picture of your son, a flower, a waterfall, a fox that crosses your path. How can you easily move from one to another?

One solution would be to specify your goal, the kind of picture you want to take, at a high, "mission" level. Dialing in a Portrait, or Waterfall play would choose the appropriate settings. We envision the ability to create packages of settings offline that can then be loaded onto the camera. Following a similar proposal by Miller and Parasuraman [10], we refer to these packages as "plays." Plays could be "called" quickly by voice or using a scroll wheel. In some ways, plays may seem similar to the "scene modes" appropriate for many types of photography (sports, portrait, night, fireworks, etc.) offered on many cameras. However scene modes generally lack transparency, customizability (strategic, offline), and flexibility (tactical, real time). By allowing operators to create plays offline, it is possible to imagine building in crazy levels of specificity. Imagine a baseball play where you could specify the field location. If you then tell it you are taking pictures of the pitcher, and it senses your location (GPS) it can determine the approximate distance greatly improving autofocus speed and accuracy. You can specify the uniform color so it can select the appropriate people out of a scene. The operator could say "play at second" and the autofocus would focus at the appropriate distance and on the appropriate player, making it faster and less likely to focus on the wrong subject entirely. (Yes, one of us spends way too much time taking pictures at Little League games.)

Plays do more than simply allow the operator to change many more specific settings at once, because plays can contain information about the desired end product; the goal for the shot. This enables, the automation to perform intent based actions based on real-time information about the goal and environment where the shot is taken. For example, the automation could be set to take the picture when the ball hits the bat, or the bee lifts off the flower. These events happen too quickly to be reliably captured manually, but can be captured by automation if it "knows" what to look for.

Once the automation knows the goal, it can also determine whether the goal is being met. For example, if there is not enough light, the shutter speeds necessary to stop motion may result in underexposure. Similarly, it may not be possible to expose a scene in a way that captures both shadow and highlight detail. In these cases the camera might warn the operator, suggesting the use of a flash or asking whether it should sacrifice the shadows or the highlights.

2.2 Navigating by Car

Today nearly everyone drives with a navigation system that would have been unimaginable just ten years ago. Enter your destination, and these systems plot a route for you that, to the degree possible, avoids traffic delays. They then provide you with turn-by-turn directions and offer to reroute you as the traffic situation changes.

These systems already provide some HAT-like features. They allow some specification of high-level goals (e.g., mode of transportation, fastest time, shortest route, avoid tolls, avoid highways) and the navigation system will generate route options with estimated driving times. On a computer, you can even create a route manually, and Google Maps, will give you an estimate of how long it will take to drive it. However, in current systems, the list of options is relatively limited. To some extent, this maybe unavoidable. Today's automation may not have a good sense of what makes a road scenic or fun to drive. Thus, if finding a scenic route is a goal, greater input from the operator may be required. However, in choosing a scenic route, the operator presumably would appreciate feedback from the automation about things it does know, for example, time to destination and road closures.

Current navigation systems also fail to give operators crucial information related to what is often the primary question on their minds: Am I going to make it there on time? They *do* give you an estimate of your arrival time based on current driving conditions, and often a fairly useless reason for any delay (e.g., "Traffic is heavy"). However, driving in traffic is generally not so simple. Apple, Google, and Waze have access to large databases from which they could generate statistical profiles that would allow them to answer questions like: How early do I have to leave in order to have a 95% chance of being on time? What is the probability that I hit traffic on the Bayshore Freeway if I leave at 3:00? What alternatives do I have if traffic gets worse?

The answers to these questions would allow a user to develop a plan for a trip that goes beyond the routing currently provided, to develop alternatives in case problems develop in transit. Current navigation systems offer to reroute you if traffic patterns change and another route becomes faster. Unfortunately, while you are driving is not the time for "bi-directional communication" with your navigation system; carefully vetting the proffered route is difficult (and possibly illegal) while driving. On the other hand, simply accepting this offer, can be a bit of a crapshoot. You might end up zipping along a highway, but you might find yourself in a warren of little streets in a questionable neighborhood. A solution to this problem might be to move the bi-directional communication to before departure. If you live in a large city with traffic problems, you probably find yourself periodically discussing your commute with co-workers. You probably discuss your strategies for getting home. Leave by 4:00, take one freeway unless it is unusually slow, in which case switch to surface streets. Maybe a co-worker has suggested a new better route. We suggest that navigation systems could become like very knowledgeable co-workers (at least knowledgeable when it comes to traffic) sitting in the right seat looking at the bigger picture. Using the navigation system, you could develop a strategy for your commute. This strategy does not have to be static. Maybe you have a preferred route and are only willing to change if you can save ten minutes. Once en route, maybe you are willing to be rerouted from one freeway to another to save five minutes, but only want to switch to surface streets if it will save ten minutes. Maybe you only want to switch if there is a 90% chance that the new route will actually end up being faster. Maybe you want the automation to ask before switching you to surface streets but simply to reroute you when it finds a faster highway. These strategies look a lot like the plays discussed above. They can be very complicated. However, because they can be formulated offline, they allow you to direct how to adapt to changing driving conditions without requiring you to negotiate with the navigation system while you are driving. Further, because these plays can be reused every time you drive to and from work, investing some time in developing a good one can save significant amounts of time down the road.

3 HAT Design Patterns

Across aviation, photography, and automobile navigation, we see very capable automation that does not achieve its full potential because it is not aware of the goals and expertise of its human operators. In each case, default parameters are set by designers and engineers that assume a set of generic goals on the part of the user. These defaults are difficult or impossible for the operators to modify. It is often unclear what they even are. While the human could, in many cases, ignore the automation, this would sacrifice important abilities the automation has that could improve outcomes. In each case we propose an interface that allows the operator better access to modify these parameters by specifying goals in a more nuanced way, by providing transparency into how the automation will meet those goals, and by allowing for negotiation with the automation when those goals cannot be met. Here we attempt to capture what generalizes across these domains.

3.1 Plays

One solution that appears to be useful across all the domains we have looked at is plays. Plays encapsulate goals, procedures, and division of responsibility into a package that can be specified offline and instantiated quickly in real-time situations. Plays help to realize our tenet that the operator is in control by allowing the operator to explicitly request a course of action quickly, reducing the need for automation to guess at the operator's intent. Plays do this by shifting much of the communication about context and authority (see Structure in Sect. 3.5 below), offline. We see this in the development of the play for photographing baseball where team colors and field position are entered before the game, and in the navigation example where various route options and their priority are entered before departure. Plays can also help with transparency, for instance, in the automobile navigation case, the plays make the priorities used by the automation explicit.

3.2 Timing

One interesting generalization between these examples is the effect of timing. In both the photography and the navigation example, there is a planning phase, where specification of the set of relevant plays occurs, and an execution phase. The execution phase itself consists of discrete action events (taking of pictures; path changes), with pauses between them. Changes to the play can occur between these actions, but would be disruptive during execution.

3.3 Bi-directional Communication

Another solution that seems to generalize across domains is bi-directional communication. Much work has gone into the proper allocation of functions between automation and human operators (e.g., [4, 7]). However, in human-human teams, team-members often perform similar if not identical functions; just imagine a brainstorming session. More formally, with traditional "Pilot Flying/Pilot Monitoring" procedures in aviation, a second person is used to generate ideas and catch errors more than to add new functionality. Interestingly, there is a similar style of programming, called pair programming, where two programmers sit together at one monitor, one typing code and the other monitoring for errors. We see something similar in both the photography and navigation discussions above. Both the automation and the human operator may have a role in performing a particular function, potentially the same role. For example, in photography, the operator can focus at roughly the correct distance and let the automation fine tune, but sometimes the automation may focus and the operator may need to fine tune. Similarly in navigation, the operator and automation may go back and forth fine tuning a cross country trip to go along scenic routes and visit particular locations while also meeting a timetable and reaching a camp ground each evening. While the human sets the mission level goals, even at that level automation may have input as to whether the goals are achievable. We see this sort of back and forth, bi-directional communication, as a critical part of making human-computer interaction into teaming. Thus, we believe the development of interfaces that support bi-directional communication is crucial for HAT.

3.4 What Is a Design Pattern?

We have been discussing two HAT-inspired solutions to common problems with automation: plays and bi-directional communication. In other fields, such generalized solutions to common problems are often captured as "design patterns." Design patterns were introduced in architecture by Alexander, et al. in the influential book A Pattern Language: Towns, Buildings, Construction [11]. For example, the pattern Raised Walk, is offered as a solution to the problem "Where fast moving cars and pedestrians meet in cities, the cars overwhelm the pedestrians. The car is king, and people are made to feel small." Design patterns have been particularly influential in computer programming. The book Design Patterns: Elements of Reusable Object-Oriented Software [12], introduces 23 patterns, following a more elaborate format than A Pattern Language. Each pattern is broken down into sections specifying (among other things) the intent, motivation, when to use it, consequences, related patterns, and advice on implementation.

In conjunction with the NATO working group on Human Autonomy Teaming (HFM-247), we have been working to develop similar design patterns for HAT. These patterns are evolving as members of the working group attempt to define them in ways that will be useful to their current projects, support generalization to new projects, and interact well with each other. Below we give a preliminary sketch of a bi-directional communication pattern, based on the observations above. This sketch follows an abbreviated version of the format used by Gamma et al. [12].

3.5 A Bi-directional Communication Pattern

Intent. First, our pattern lays out a brief description of what the pattern attempts to do. Our bi-directional communication pattern supports generation of input from all relevant parties and its integration into decisions.

Motivation. Next comes a description of the problem and how the pattern solves it. From the above examples, it is apparent that for many problems humans and automation bring differing strengths and weaknesses to a problem. Looking at the task of focusing the camera from the photography example, automation is generally faster and more accurate than human operators. But automation can focus on the wrong object or fail to find an object to focus on entirely. The operator can supply information that improves the autofocus's performance and add information when the autofocus none-the-less gets it wrong. The result is a system that is less error prone than either operator or automation by itself. The situation with navigation is very different, yet, in many ways, very similar. Again the automation has important strengths. It can pull together great stores of information and make them available to the user. However, because navigation systems work in a domain filled with uncertainties (often behaving in a complex and non-linear way), they cannot provide certain answers. Above we imagine a future system that provides a more detailed statistical description of the various options available to the operator. Still, it falls on the user to decide what types of risk to take.

Further, the examples given here show the advantages of making communication bi-directional with information going back and forth between the parties (as opposed to a simpler system in which, for example, the automation simply alerts the operator to some information).

Applicability. The examples cited above show the potential utility of implementing the bi-directional communication pattern in cameras and navigation systems. Our studies in aviation have shown that a back and forth between humans and automation results in solutions that are more acceptable to the human operators [2, 13]. Bi-directional communication facilitates sharing of information regarding problem definition, potential solutions and authority to act, information that has been shown to be critical in a wide variety of situations [5, 8, 14–16]. Conversely, it is important that automation be designed to take advantage of human knowledge and expertise. Automation can only react appropriately within the range of situations it was developed for. Outside this range, it may lack access to relevant information or the ability to generate appropriate plans, making it brittle. Allowing human operators to input information improves the system's flexibility. Thus, we believe this pattern is broadly applicable to complex automation.

There are some exceptions, however. Bi-directional communication may not work with all types of automation such as genetic algorithms and neural networks, because these systems lack the structure necessary to provide a rationale for their ratings and recommendations. Also, there are situations where it may be necessary to limit communications. This is particularly true in urgent situations where time is not available for comprehensive communications.

Structure. Next we describe, abstractly, what the solution looks like. For our bi-directional communication pattern, we are concerned primarily with what types of information need to be shared between automation and the human operator. We have divided the information that needs to be shared between automation and the human operator into three: Authority (what level of automation should the automation be working at?), Context (what problem/what goals are the human and automation attempting to work on?), and Options (how could we achieve our goal or solve our problem?)

Authority. One crucial piece of information that needs to be communicated between automation and the human operator regards authority. This could be as simple as whether the operator or automation is performing certain tasks. At a slightly more complex level, we can imagine assigning the automation intermediary levels of automation (LOA) [17]. For example, the automation might propose a course of action (e.g., the navigation system proposing a re-route) which must be accepted by the operator. As automation gets more complex, however, we envision a more complex authority structure. For instance, automation may be assigned more complex "working agreements" like contingent LOAs (e.g., reroute me to surface streets if such a reroute is predicted to save more than 10 min). Using such automation, human operators will need access to the current working agreement and will need the ability to dynamically change task allocations and levels of automation.

Context. Before developing a plan, automation and human operators must communicate about the context in which the plan is being created:

- What is the goal? Ordinarily we would expect the operator to set the goal (Where do I want to go?), but even here the automation may play a role (e.g., if a system failure is detected on an aircraft, the automation might propose diverting [2]).
- What are the constraints? A given situation typically comes with some constraints that rule out certain solutions (or, at least, make them categorically worse than others). For example, an aircraft cannot divert to an airport that is outside its fuel range, and a car cannot drive the wrong way down a one-way street. Both the operator and the automation may be aware of constraints on how the goal might be achieved and need to be able to convey this information to the other party. There may also be temporal constraints, time limits by which actions must be taken.
- What are the priorities? In addition to constraints, other factors can be more naturally traded-off against one another. If a passenger has a heart attack, you want to transport him to a good medical facility as quickly as possible. That is great if the closest airport also has the best medical facilities, but if it does not, how much time are you willing to give up in order to get better facilities? Weights allow you to define a function across the different factors that go into your decision, communicating to the automation how much the operator cares about these different factors.

Options. Of course the reason you have the automation is to calculate options. This could be at a very low implementation level (e.g., auto-throttles of an aircraft adding thrust to maintain airspeed), or at a mission level (e.g., a route planning tool suggesting where to divert to). The automation may be authorized to implement options without operator input (aside from the initial authorization). However, the operator should have access to these options, the rationale for selecting them, the projected consequences of their implementation, and the automation's confidence in these outcomes. The operator should also be able to generate options and have the automation evaluate those as well.

Implementation. What guidelines are available for implementing this pattern? Bi-directional communication takes time. If one had to negotiate with the autofocus system before capturing an action shot, or a route-planning tool while driving in traffic, it would not happen. The examples given above suggest, however, that some or all of the communication can be done offline. Plays offer a means of encapsulating and abbreviating this communication allowing the operators to specify when and how decisions are to be made before the urgency of real-time operations sets in. Even when the play is in progress, there are points of greater and lesser urgency. Interaction should be scheduled between shots in the photography example or turns in the case of navigation. In implementing this pattern for other domains, designers should be conscious of similar rhythms.

4 Next Steps: Toward a Framework

HAT has been a goal since the dawn of the computer age [4]. However, today, with self-driving cars on our streets and self-flying aircraft in our skies, we have a much clearer picture of where automation is heading. While things that were only dreams a short while ago have quickly become indispensible, our interactions with the automation are often frustrating. Human factors engineers, unfortunately, are playing catch-up in trying to shape a more satisfying relationship with these automated systems. In this paper we present a snapshot of our strategy for developing a framework for HAT. Our strategy begins with tenets derived from CRM. CRM aims to maintain clear lines of command and authority while fostering free and open exchange of relevant information. We started from a position that the human operator should remain in control, and that goals, plans, and information relevant to accomplishing those goals and plans should be freely shared. We then asked, what does the operator remaining in control and sharing this information look like in practice? Our goal is to iterate this process, updating our tenets based on our exploration of their implications for the design of real systems. In doing so, our goal is to develop a framework for HAT, consisting of our tenets, guidelines for implementing the tenets, and software libraries that make this implementation easier. To achieve these goals we must be able to find generalizations in how human operators effectively use automation. We see the development of HAT interfaces as being parallel to the development of graphical user interfaces in the 1980s or touch interfaces earlier this century. In both cases there was an early period of experimentation, which eventually settled into a familiar set of design elements (e.g., desktop, windows, and menus). Software frameworks developed that mirrored these design elements allowing for easy reuse, and accelerating adoption. The kind of intelligent automation for which HAT would be useful is still in its infancy.

We expect fluidity in HAT interface design, until the underlying automation matures. However, we expect the kind of interaction discussed here to become increasingly prevalent in the years to come.

Acknowledgments. We would like to acknowledge NASA's Safe and Autonomous System Operations Project, which funded this research. Thanks to Walter Johnson, Vernol Battiste, Mike Matessa, and Gilles Coppin, for their thoughts and comments.

References

1. Shively, R.J., Lachter, J., Brandt, S.L., Matessa, M., Battiste, V., Johnson, W.: Why human-autonomy teaming? In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (2017, this edition)
2. Brandt, S.L., Russell, R., Lachter, J., Shively, R.J.,: A Human-autonomy teaming approach for a flight-following task. In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (2017, this edition)
3. Strybel, T., Keeler, J., Mattoon, N., Alvarez, A., Barakezyan, V., Barraza, E., Park, J., Vu, K.-P., Battiste, V.: Measuring the effectiveness of human automation teaming in reduced crew operations. In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (2017, this edition)
4. Licklider, J.C.R.: Man-computer symbiosis. IRE Trans. Hum. Factors Electron. **1**, 4–11 (1960)
5. Endsley, M.R.: From here to autonomy: lessons learned from human-automation research. Hum. Factors **59**, 5–27 (2017)
6. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multi-robot control: a literature review. Army Research Lab Technical report, ARL-TR-6328 (2013)
7. Feigh, K.M., Pritchett, A.R.: Requirements for effective function allocation: a critical review. J. Cogn. Eng. Decis. Mak. **8**, 23–32 (2014)
8. Lyons, J.B.: Being transparent about transparency: a model for human robot interaction. In: AIAA Spring Symposium Series (2013)
9. Woods, D.D.: The risks of autonomy: Doyle's catch. J. Cog. Eng. Decis. Mak. **10**, 131–133 (2016)
10. Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Hum. Factors **49**, 57–75 (2007)
11. Alexander, C., Ishikawa, S., Silverstein, M., Ramió, J.R., Jacobson, M., Fiksdahl-King, I., Angel, S.: A Pattern Language: Towns Buildings Construction. Oxford University Press, New York (1977)
12. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, Boston (1994)
13. Battiste, V., Johnson, W.W., Dao, Q., Brandt, S.L, Johnson, N.H., Granada, S.: Assessment of flight crew acceptance of automated resolution suggestions and manual resolution tools. In: 26th International Congress of the Aeronautical Sciences, Anchorage, AK (2008)
14. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**, 50–80 (2004)
15. Hoff, K.A., Bashir, M.: Trust in automation: integrating empirical evidence on factors that influence trust. Hum. Factors **5**, 407–434 (2015)

16. Wiener, E.L., Kanki, B.G., Helmreich, R.L. (eds.): Crew Resource Management. Academic Press, Cambridge (2010)
17. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **30**, 286–297 (2000)

Using a Crew Resource Management Framework to Develop Human-Autonomy Teaming Measures

Michael Matessa(✉)

Rockwell Collins, 350 Collins Rd NE, Cedar Rapids, IA 52498, USA
michael.matessa@rockwellcollins.com

Abstract. Recent developments in technology have permitted an increased use of autonomy. To work best with humans, autonomy should have the qualities of a good team member. But how can these qualities be measured? One way is to use similar measures to those used to measure good teams with human members. For example, the Non-Technical Skills (NOTECHS) framework measures Crew Resource Management (CRM) skills that allow pilots to work together as a team. The framework consists of skill categories, elements of those categories, and behavioral markers that demonstrate good or poor performance in the elements. This paper introduces CMSD (Cooperation/Management/Situation Awareness/Decision Making), a measurement system based on NOTECHS and other widely-used skill level systems, which provides quantitative measures of Human-Autonomy Teaming (HAT).

Keywords: Human-Autonomy Teaming · Crew Resource Management · Measures

1 Introduction

Based on progress in Artificial Intelligence and Human Performance Modeling, Human-Autonomy Teamwork systems are being developed and tested in which humans and increasingly autonomous systems dynamically adjust and cooperate to accomplish a joint objective. In these systems, team members' responsibilities and commitments are managed such that the human and automation jointly enhance performance and manage contingencies. This paper introduces CMSD (Cooperation/Management/Situation Awareness/Decision Making), a measurement system which provides quantitative measures of Human-Autonomy Teaming (HAT). In the following sections, HAT systems are designed using concepts of operation, design patterns, and use cases. CMSD is used to assist the design and evaluation of these systems.

2 Concept of Operation for Reduced Crew Operations

One concept that takes advantage of increasingly autonomous systems is reduced crew operations (RCO) for transport category aircraft. RCO envisions having one pilot on board domestic flights, and two pilots on board long-haul operations, where one of the

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_5

two pilots is often off of the flight deck resting in the bunk. An important element of RCO research seeks to develop a concept of operation (ConOp) that covers the roles and responsibly of the principal human operators, the automation tools used by them, and the operating procedures for human-human and human-automation. The human-automation function allocation, in particular, has been an ongoing NASA focus, drawing upon insights gathered from subject matter experts in industry, academia and government during technical interchange meetings and from empirical human-in-the-loop research [1–3].

The proposed NASA RCO ConOp [4] includes three basic human roles: the pilot on board (POB), the dispatcher, and when necessary, a ground pilot. The POB (unless incapacitated) would serve as the captain and pilot-in-command. As such, s/he would determine when to call on automation and ground support. The POB's main tasks would be to manage risk and resources (both human and automation).

Onboard automation would assist the POB with many tasks currently performed by the pilot monitoring, such as flight management system (FMS) input, assisting with checklists and validating inputs.

Ground-based automation would assist the dispatcher in a variety of tasks. Dispatch tasks would be similar to current operations (e.g., preflight planning, monitoring aircraft positions, and enroute reroutes), but for those tasks currently performed jointly with the pilot, the dispatchers (aided by automation) would absorb some of the POB's workload (e.g., creating new flight plans, vetting them with air traffic control (ATC) and uplinking them to the aircraft). In this ConOp, automation would assist the dispatcher with creation of pre-flight briefings, flight path monitoring, selection of divert airports, and optimizing reroutes. Automation would also be responsible for monitoring many flights and alerting the dispatcher to aircraft needing assistance. In addition, the POB could call the dispatcher for consultation on important decisions where s/he might previously have consulted the first officer (e.g., diagnosing an aircraft system caution light or determining the fuel consequences of a holding instruction).

Commonly occurring HAT situations in the RCO ConOps can be characterized with design patterns.

3 Building a Design Pattern

To put this in context, it is helpful to see how these patterns have been constructed. To that end, I have excerpted this section from our previous work [5]. Schulte [6], in conjunction with Neerincx and Lange [7], proposed using a set of primitives to build HAT patterns. They proposed three types of agents: (1) human operators, (2) intelligent/cognitive agents, and (3) automated tools. The agents, which can either be co-located or distributed, can be connected by a cooperative, supervisory, or communications link.

As described by Schulte [6], an important preliminary task in the construction of HAT patterns is the identification of the Work Objective. The Work Objective identifies the aspects that initiate and characterize the mission or purpose of the work. The Work Objective provides a black box description of the Work Process, which includes informational inputs (e.g., ATC clearance), environmental inputs (e.g., airspace) and supply inputs (e.g., fuel). The Work Process, utilizing all of these inputs and the Work Objective, produces a Work Process Output (e.g., reducing target speed) on the Work Object (e.g., speed of aircraft) that distributes meaningful physical and conceptual actions to human-automation team members within the overall Work Environment.

Using these primitives, a pattern was built that describes HAT in an RCO context. An initial use case was identified, and by walking through each step, the agents and links required to depict one such pattern were identified [5]. Figure 1 provides a legend for the elements that were used in the design pattern.

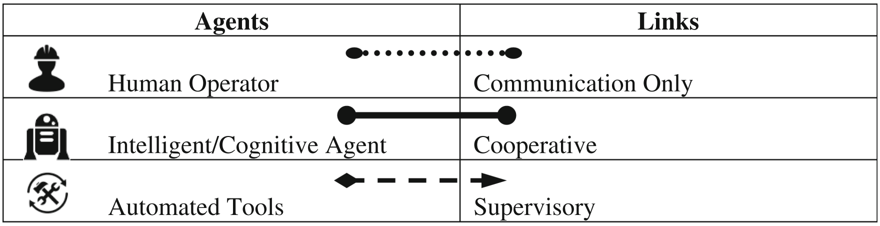


Fig. 1. Legend for design pattern elements; cooperative and supervisory links imply communication.

3.1 Use Case: Designing Thunderstorm Alerting

Initial Conditions. FLYSKY12 is en route from SFO to ORD. There is one POB and a dispatcher flight following. The *Work Objective* is to avoid a thunderstorm. The *Work Process* is the set of steps necessary to resolve the situation with the output to divert to an alternate airport.

Step 1. Detection and Alerting of Thunderstorm. Dispatch automation informs dispatcher of convective cell growing on flight path of FLYSKY12. *This requires a communication link between dispatch automation and the dispatcher (covered by supervisory link in the pattern).*

Step 2. Dispatcher Informs POB of Cell. *This requires a link between the dispatcher and the POB. This link is as a cooperative link (as in the pattern) because, by regulation, the dispatcher and POB share responsibility for safe operation of the flight (including detecting and responding to thunderstorms).*

Step 3. Modification of Flight Plan. Seeing a need to re-route, the dispatcher requests modified flight plan from dispatch automation. Dispatch automation returns modified flight plan. *The delegation of flight path planning to the automation requires a supervisory link. This planning requires consideration of multiple strategies making this automation an agent.*

Step 4. Dispatch Uplinks Modified Flight Plan. *Uses the link between dispatch and POB from Step 2.*

Step 5. POB requests clearance for flight plan from ATC. *POB and ATC are both responsible for safety of flight and thus this is a cooperative link.*

Step 6. ATC Rejects Clearance. ATC tells POB that aircraft must take additional six-minute delay for new arrival slot coming into ORD. *Uses cooperative link from Step 5.*

Step 7. Planning for Delay. POB asks automation for alternatives to take six-minute delay. Automation provides two alternatives: (a) Slow down, saves fuel but risks further movement/growth of cell (b) Hold past cell, more fuel burn but lower risk of further deviations. *POB is delegating this task to the automation, requiring a supervisory link. The automation is developing multiple strategies for taking the delay, making it an agent.*

Step 8. POB requests clearance from ATC, modified with holding after passing cell; ATC approves request. *Uses the same cooperative link from Steps 2 and 5.*

Step 9. POB tells Agent to implement the new clearance. Agent sets autopilot in accord with clearance. *The POB delegates tasks to the agent, and the agent uses tools to perform the task.*

The elements of this use case are captured in the pattern shown in Fig. 2. The POB and the assigned dispatcher are jointly responsible for the flight, assisting each other in a cooperative relationship. Similarly, the POB and the ATC have complementary roles in assuring safety of flight, and thus must also cooperate. Further, the ConOp specifies that both dispatch and the POB acquire significantly enhanced automation. Thus, in most situations the operators, tools, agents, and their underlying relationships are fixed by the ConOp. Further, at a high level, the *Work Objective* remains constant for RCO: getting the aircraft to the best airport possible for the airline (usually its destination and ideally on time) while maintaining safety of flight. The relevant informational, environmental, and supply inputs also remain constant (although possibly with different weightings) across operations.

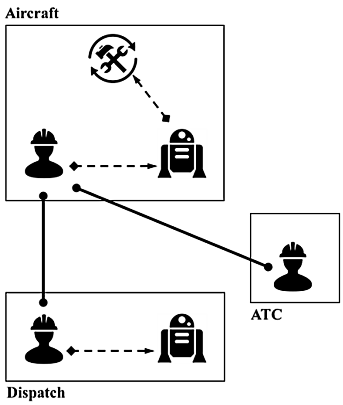


Fig. 2. Use case design pattern

4 The CMSD Measurement System

The development of the HAT design pattern required several subjective decisions. For example, considerable time was spent determining which piece of automation was a cognitive agent versus a tool and the type of connection between the various agents. To aid in making these decisions, the CMSD measurement system was developed based on a framework that Line Check Airmen use to evaluate Crew Resource Management performance by human pilots. The Non-Technical Skills framework (NOTECHS) has four Categories: Cooperation, Management and Leadership Skills, Situation Awareness, and Decision Making. Each are subdivided into Elements and behavioral markers (Fig. 3).

Category	**Element**	**Behavior**
Cooperation	Considering others	Consider condition of other
	Supporting others	Offer assistance
Management/ Leadership	Authority/Assertiveness	Take Initiative
	Maintain standards	Enforce SOP
	Planning/Co-ordinating	State plan
	Workload management	Distribute tasks
Situation Awareness	System awareness	Monitor/report system (incl. other crew)
	Env awareness	Monitor/report env
	Awareness of time (anticipation)	Monitor/report time constraints
Decision Making	Problem diagnosis	ID problem
	Option generation	Generate/elicit options
	Option selection	Select option
	Outcome review	Review outcome

Fig. 3. NOTECHS CRM measurement framework

The categories are used by CMSD to determine individual and relationship skills in the design pattern. For individual skills, the degree to which agents exhibit situation awareness is indexed by the levels of situation awareness described by Endsley [8], the management capabilities of agents is indexed by Sheridan's levels of automation [9], and decision-making ability is assessed using the Non-Technical Skills framework (NOTECHS) categories [10]. See Fig. 4 for details. In the following, CMSD is used to give a more quantitative assessment of the automation in the use case. Each step is re-examined, giving the reasoning behind these assessments.

Situation Awareness (Endsley)
1) Perceive
2) Comprehend
3) Project

Decision Making (NOTECHS)
1) ID problem
2) Generate options
3) Select option
4) Review outcome

Management (Sheridan)
1) The computer offers no assistance: human must take all decision and actions.
2) The computer offers a complete set of decision/action alternatives, or
3) narrows the selection down to a few, or
4) suggests one alternative, and
5) executes that suggestion if the human approves, or
6) allows the human a restricted time to veto before automatic execution, or
7) executes automatically, then necessarily informs humans, and
8) informs the human only if asked, or
9) informs the human only if it, the computer, decides to.
10) The computer decides everything and acts autonomously, ignoring the human.

Fig. 4. Quantitative HAT measures for SA, decision making, and management

4.1 Use Case 1: Designing Thunderstorm Alerting

Step 1. Detection and Alerting of Thunderstorm. Dispatch automation informs dispatcher of convective cell growing on flight path of FLYSKY12. *No CRM skill indicated.*

Step 2. Dispatcher Informs POB of Cell. *Requires a cooperative relationship between POB and dispatcher with Collaboration ability (awareness of needs of other) and SA ability (monitoring other) labeled C/S.*

Step 3. Modification of Flight Plan. Seeing a need to re-route, the dispatcher requests modified flight plan from dispatch automation. Dispatch automation returns modified flight plan. *Requires management ability in a supervisory relationship between dispatcher and agent (labeled M), with the agent having a Decision Making ability to select an option (NOTECHS level three, labeled D3).*

Step 4. Dispatch Uplinks Modified Flight Plan. *Requires a cooperative relationship between POB and dispatcher with Collaboration ability (awareness of needs of other), Decision Making ability (eliciting divert options from dispatcher), and Management ability (delegating divert location task to the dispatcher), labeled C/D/M.*

Step 5. POB requests clearance for flight plan from ATC. *Requires a cooperative relationship between POB and ATC with Collaboration ability (awareness of needs of other), Management ability (maintaining standard of requesting clearance from ATC), SA ability (monitoring of other), and Decision Making ability (eliciting divert options from POB), labeled C/M/S/D.*

Step 6. ATC Rejects Clearance. ATC tells POB that aircraft must take additional six-minute delay for new arrival slot coming into ORD. *Uses cooperative relationship from Step 5.*

Step 7. Planning for Delay. POB asks automation for alternatives to take six-minute delay. Automation provides two alternatives: (a) Slow down, saves fuel but risks further movement/growth of cell (b) Hold past cell, more fuel burn but lower risk of further deviations. *Requires management ability in a supervisory relationship between POB and agent (labeled M), with the agent having a Situation Awareness ability to project consequences of actions (Endsley level 3, labeled S3) and a Decision Making ability to generate options (NOTECHS level two, labeled D2).*

Step 8. POB requests clearance from ATC, modified with holding after passing cell; ATC approves request. *Uses the same cooperative relationship from Step 5.*

Step 9. POB tells Agent to implement the new clearance. Agent sets autopilot in accord with clearance. *Uses supervisory relationship between POB and agent from Step 7. Requires Management ability in a supervisory relationship between agent and aircraft (labeled M).*

Figure 5 shows how the RCO pattern is labeled using the CMSD system. Note that the final labeling shows the maximum capabilities of relationships and agents. The measure labels permit the reader to quickly determine the capabilities required of the agents and the relationships.

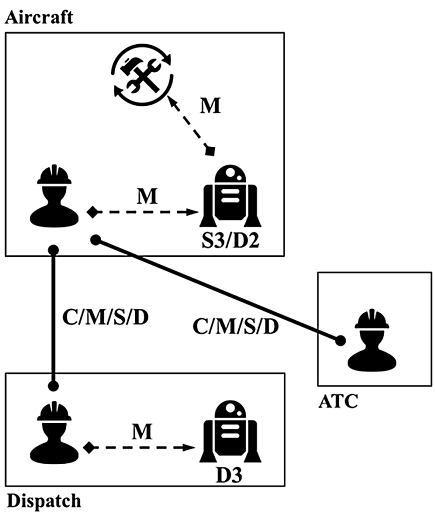


Fig. 5. RCO design pattern with additional CMSD measures

4.2 Use Case 2: Designing Traffic Avoidance

To further exercise the CMSD measurement system, a use case was developed that included an agent with the ability to monitor traffic, predict Loss of Separation, and take evasive action. The following examines each step of the use case and specifies the HAT measure reasoning.

Step 1. **Engaged.** On board agent with Traffic Collision Avoidance capability is engaged. *No CRM skill indicated.*

Step 2. Detection. Agent predicts future LOS with traffic, provides avoidance option, and waits for evaluation of option from POB. *Requires a supervisory relationship between agent and traffic where the agent has Situation Awareness ability to monitor traffic (labeled S). Requires a cooperative relationship between POB and agent with Collaboration ability (offering assistance) and Decision Making ability (eliciting options), labeled C/D. The agent needs a Situation Awareness ability to project a future Loss of Separation (Endsley level 3, labeled S3), a Decision Making ability to select an avoidance maneuver option (NOTECHS level 3, labeled D3), and a Management ability to suggest that option (Sheridan level 4, labeled M4).*

Step 3. Lack of Response. POB does not react in time. *Lack of Cooperation by POB for not considering the condition of the agent.*

Step 4. Execution. Agent executes maneuver option to avoid traffic. *The agent has a Management relationship with the POB (labeled M) with an ability that allows the human a restricted time to veto before automatic execution (Sheridan level 6, labeled M6).*

Figure 6 shows the RCO pattern updated using the CMSD system for an agent with TCAS capabilities. A comparison with Fig. 5 shows a need for an onboard agent with increased Decision Making and Management abilities and increased cooperative Collaborative and Decision Making abilities.

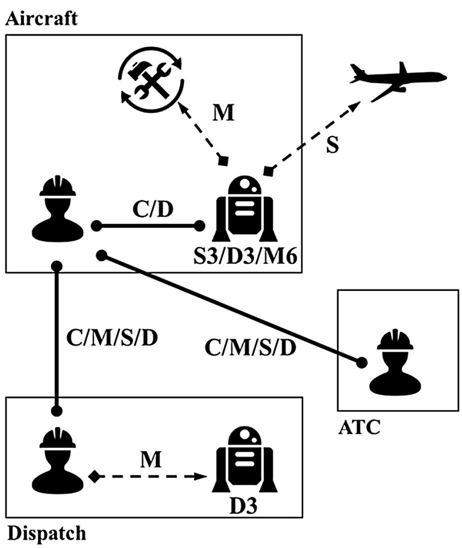


Fig. 6. RCO pattern and CMSD measures for an agent with TCAS capabilities

4.3 Use Case 3: Evaluating Traffic Avoidance

In addition to design, the CMSD system was developed to assist in the evaluation of existing systems. NASA Ames Research Center is developing the J-HATT ground control station to investigate human factors issues involved with the control of Unmanned Aerial Vehicles. Figure 7 shows a screenshot of the J-HATT and highlights Detect And Avoid (DAA) features. When Loss Of Separation is predicted, a colored arc is presented showing evasive turning options that are safe (green), questionable (yellow), and dangerous (red).

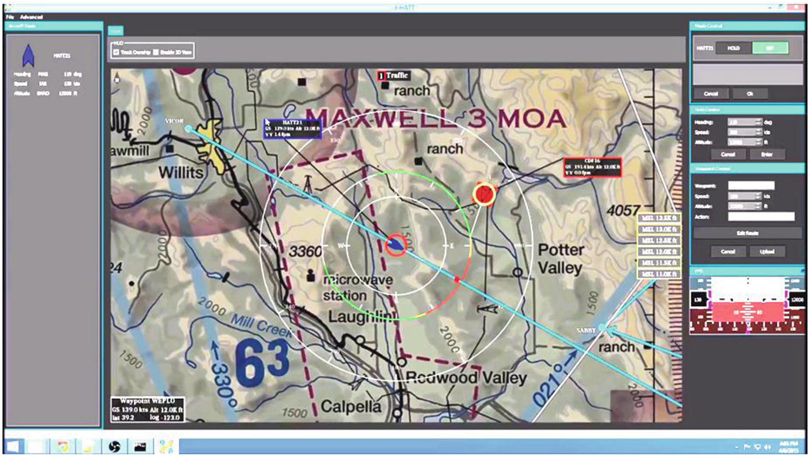


Fig. 7. J-HATT UAV ground control station

Figure 8 shows an evaluation of the system using the CMSD system. The ability of the agent to monitor traffic demonstrates a supervisory relationship between agent and traffic where the agent has a monitoring Situation Awareness skill (labeled S). The agent demonstrates a Situation Awareness ability to project a future Loss of Separation (Endsley level 3, labeled S3), a Decision Making ability to generate avoidance maneuver options (NOTECHS level 2, labeled D2), and a Management ability to narrow the selection of options down to a few (Sheridan level 3, labeled M3).

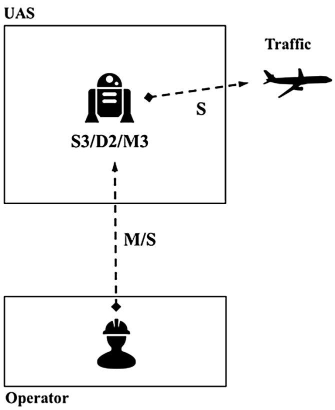


Fig. 8. Pattern and measures of system with J-HATT UAV ground control station

The CMSD system can be used to suggest improvements to J-HATT. For individual agent skills, decision making ability could be improved to select one of the generated options (raising the level to D3) and management ability could be improved to automatically execute the option after giving the operator time to veto (raising the level to M6). For the last option to be seen as cooperative by the operator, the agent would need to indicate that it was able to offer assistance and execute the option, and that it was waiting for the operator to accept or reject the option (C and D labels on a cooperation relationship).

As a summary of how the labels of the CMSD system are related to NOTECHS skills, Fig. 9 shows the NOTECHS framework with relationship and individual skill levels.

Relationship skill	**Category**	**Element**	**Behavior**	**Individual skill**
C	Cooperation	Considering others	Consider condition of other	
		Supporting others	Offer assistance	
M	Management/ Leadership	Authority/Assertiveness	Take Initiative	**M1-10**
		Maintain standards	Enforce SOP	
		Planning/Co-ordinating	State plan	
		Workload management	Distribute tasks	
S	Situation Awareness	System awareness	Monitor/report system (incl. other crew)	**S1-3**
		Env awareness	Monitor/report env	
		Awareness of time (anticipation)	Monitor/report time constraints	
D	Decision Making	Problem diagnosis	ID problem	**D1-4**
		Option generation	Generate/elicit options	
		Option selection	Select option	
		Outcome review	Review outcome	

Fig. 9. NOTECHS framework with additional labels for relationship and individual skills

5 Discussion

This paper suggests that defining design patterns and measures can help prescribe and describe human-autonomy relationships and abilities. Based on the same framework that is used to evaluate human teams, the CMSD system can help to design and evaluate human-autonomy teams. Widely-used skill level systems [8–10] incorporated into CMSD allows it to give quantitative descriptions of Situation Awareness, Decision Making, and Management abilities.

How does CMSD relate to other methodologies that use quantitative measures of autonomous systems? One such methodology is being used by the National Highway

Traffic Safety Administration (NHTSA) to define levels of automation for autonomous vehicles [11]. The NHTSA adopts the SAE International (SAE) definitions for levels of automation:

- At SAE Level 0, the human driver does everything;
- At SAE Level 1, an automated system on the vehicle can sometimes assist the human driver conduct some parts of the driving task;
- At SAE Level 2, an automated system on the vehicle can actually conduct some parts of the driving task, while the human continues to monitor the driving environment and performs the rest of the driving task;
- At SAE Level 3, an automated system can both actually conduct some parts of the driving task and monitor the driving environment in some instances, but the human driver must be ready to take back control when the automated system requests;
- At SAE Level 4, an automated system can conduct the driving task and monitor the driving environment, and the human need not take back control, but the automated system can operate only in certain environments and under certain conditions; and
- At SAE Level 5, the automated system can perform all driving tasks, under all conditions that a human driver could perform them.

Using the SAE levels, a distinction is drawn between Levels 0–2 and 3–5 based on whether the human operator or the automated system is primarily responsible for monitoring the driving environment. For purposes of State traffic laws that apply to drivers of vehicles (e.g., speed limits, traffic signs), the NHTSA suggests that States may wish to deem a Level 3–5 system that conducts the driving task and monitors the driving environment to be the “driver” of the vehicle.

Comparing SAE levels and CMSD, both contain elements of monitoring and control. SAE levels uniquely mention parts of an overall task but do not go into detail on what those parts might be. CMSD explicitly mentions the ability of state projection for situation awareness and option selection for decision making, both of which may be useful in further defining task parts in driving.

Further improvements in CMSD would need to balance the generalizability of more abstract representation of skills with the diagnostic ability of specific representations of skills and tasks. For example, it may be sufficient to say that for all RCO use cases a design would need advanced onboard autonomy to assist the pilot. But to evaluate if a driver is monitoring the automated system and environment sufficiently in order to be able to override the decision of the system, details of subtasks and timing would need to be added to the CMSD system.

References

1. Comerford, D., Brandt, S.L., Wu, S.-C., Mogford, R., Battiste, V., Johnson, W.W.: NASA’s Single Pilot Operations Technical Interchange Meeting: Proceedings and Findings. Report No. NASA-CP-2013-216513, NASA Ames Research Center, CA (2012)
2. Lachter, J., Brandt, S.L., Battiste, V., Ligda, S.V., Matessa, M., Johnson, W.W.: Toward single pilot operations: developing a ground station. In: Proceedings of the International Conference on Human-Computer Interaction in Aerospace, Santa Clara, CA (2014)

3. Brandt, S.L., Lachter, J., Battiste, V., Johnson, W.W.: Pilot situation awareness and its implications for single pilot operations: analysis of a human-in-the-loop study. In: Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics. Las Vegas, NE (2015)
4. Bilimoria, K.D., Johnson, W.W., Shutte, P.C.: Conceptual framework for single pilot operations. In: Proceedings of the International Conference on Human-Computer Interaction in Aerospace, Santa Clara, CA (2014)
5. Shively, R.J., Brandt, S., Lachter, J., Matessa, M., Sadler, G., Battiste, H. Application of Human-Autonomy Teaming (HAT) Patterns to Reduced Crew Operations (RCO). In: Proceedings of the 18th International Conference on Human-Computer Interaction (2016)
6. Schulte, A.: Human systems Engineering (HSE) Approach to Human-Autonomy Teaming, NATO STO HFM-247 Task Group Presentation (2015)
7. Neerincx, M., Lange, D.: HAT Patterns, NATO STO HFM-247 Task Group Presentation (2015)
8. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**, 32–64 (1995)
9. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. **30**(3), 286–297 (2000)
10. Flin, R., Martin, L., Goeters, K.-M., Hörmann, H.-J., Amalberti, R., Valot, C., Nijhuis, H.: Development of the NOTECHS (non-technical skils) system for assessing pilots' CRM skills. Hum. Factors Aero. Saf. **3**, 95–117 (2003)
11. National Highway Traffic Safety Administration: Federal Automated Vehicles Policy. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/federal_automated_vehicles_policy.pdf. Accessed 9 Mar 2017

The Impact of Neurocognitive Temporal Training on Reaction Time and Running Memory of U.S. Active Duty Personnel

Cory R. Overby[1], Valerie Rice[2(✉)], Gary Boykin[2], Leah Enders[1], and Jessica Villarreal[1]

[1] DCS Corporation, 6909 Metro Park Drive, Suite 500, Alexandria, VA 22310, USA
coverby@dcscorp.com

[2] HRED-ATSD, Army Medical Department Field Element, U.S. Army Research Laboratory, Fort Sam Houston, San Antonio, TX 78234, USA
{valerie.j.rice.civ, gary.l.boykin.civ}@mail.mil

Abstract. Battlefields are dynamic environments that require robust cognitive skills and rapid responses. Temporal (timing and rhythm) training may improve coordination and cognition. This study examined Soldiers' reaction times and working memory following neurocognitive temporal training. Forty-two soldier volunteers were assigned to either a temporal training intervention (TTI) group (n = 19) or a control (n = 23) group. Outcome measures were simple reaction time (SRT) and running memory continuous performance (RMCP). No significant difference were found in group, time, or group x time ($p < .05$). While not reaching statistical significance, the following were considered of practical relevance for soldiers in combat. Those receiving TTI reduced their mean correct SRT, while the control groups' reaction time increased. For RMCP, the number of correct responses per unit of time increased for the TTI group and decreased for controls. A trend was shown for temporal training improving service members' reaction times and running memory.

Keywords: Timing · Rhythm · Working memory

1 Introduction

Approximately 1.5 million U.S. servicemen and women have deployed to hazardous regions of the world [1, 2]. Physical, cognitive, and emotional fitness are essential for service members to perform executive functions such as formulating strategy and executing operational plans, prior to and during deployment [3]. Recognizing the overall fitness needs of soldiers, the military has increased efforts to boost soldier's cognitive skills and abilities [4]. There is ample evidence suggesting such goals are achievable as the brain exhibits plasticity [5] and strengthening neuro-pathways promotes optimal brain performance [6]. Executive functions are associated with the prefrontal cortex of the brain [7–9]. The prefrontal cortex is also involved in the innate sense of timing, rhythm, attention, and working memory [10–12]. Therefore, training

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_6

that activates prefrontal brain regions could enhance soldiers performance in stressful battlefield conditions [13].

The Interactive Metronome® (IM) is a computer-based software program that trains a person's timing and rhythm by having them tap a trigger in time with a reference tone, by using their hands and feet [14]. Organizing and sequencing are dependent on an internal sense of timing and rhythm [15] which involves the prefrontal cortex, and pre- and post-central gyrus [16]. Thus, it is possible that increasing one behavior (timing) may yield increases in the other (organizing and sequencing). Also, 'exercising' brain regions appears to strengthen synaptic connections and information flow [17]. These findings suggest that the use of repetitive task training may influence one's internal sense of timing, rhythm, and cognitive functions.

The published literature on the IM has shown it to be useful for enhancing attention/concentration in children [23, 24], auditory processing [24, 35], academic achievement [23, 24, 36], processing speed and motor planning [35], motor timing, and perceptual motor performance [23, 28, 35, 36]. The IM has also been utilized by professional athletes to improve and enhance performance in golf, football, and hockey on both the amateur and professional levels [26–28]. The performance tasks in professional sports are akin to those performed on a battlefield. Battlefield cognitive processes require object identification, object discrimination, situational awareness, and quick reaction [37].

Although the IM is widely used as an intervention for clinical disorders, no published articles have examined the use of the IM to enhance cognitive performance of healthy, active duty military service members. Enhancing soldiers' cognitive ability would be of great interest to the military in many facets of operations. The purpose of this study was to examine the impact of IM training on reaction time and working memory of active duty personnel.

2 Methods

2.1 Participants

A total of 104 active duty volunteers were recruited. 42 (34 men and 8 women) completing post-testing. Volunteers were recruited from the Fort Sam Houston, Texas area. All volunteers completed an informed consent form prior to participation. Volunteers did not receive compensation for their participation. The protocol received IRB approval before subject recruitment began.

2.2 Procedure

Once volunteers completed the informed consent process, they were randomly assigned to either a temporal training intervention group (TTI) (n = 19) or a control group (n = 23). Both groups completed baseline testing using the Automated Neuropsychological Assessment Metric (ANAM) and an initial assessment with the IM. The independent variable was the temporal training intervention and the dependent measures were the ANAM simple reaction time (SRT) and running memory continuous performance (RMCP) tests, described below. Post-training testing was conducted after the completion of the TTI training or after a comparable period of time for the control group.

2.3 Instrumentation

Automated Neuropsychological Assessment Metric (ANAM). The ANAM software tool hosts customizable batteries of neuropsychological tests to assess cognitive performance among service members before and after deployment [25]. Dependent measures were Simple Reaction Time (SRT) and Running Memory Continuous Performance (RMCP). SRT measures attention (reaction time & vigilance) and visuo-motor response time. RMCP measures attention, concentration and working memory [30]. The SRT has have a reliability of 0.38 [32] and the RMCP has a reliability of 0.64 [33]. The specific outputs from the two assessments were mean reaction time of the items with correct responses and the number of correct responses per unit of available response time referred to as “throughput”.

After viewing the instructions, users proceed to a practice test. After the practice test, the assessment begins. For SRT, a series of asterisks (Fig. 1) are presented on the display at random intervals of time. The user is instructed to respond as quickly as possible by pressing a button each time the stimulus appears. For the RMCP Test, single characters are presented on the display in rapid sequence (Fig. 2). The user presses designated buttons to indicate if the displayed character matches or does not match the preceding character.

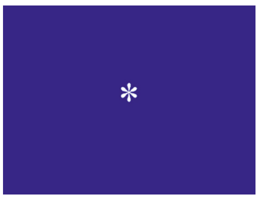


Fig. 1. An example of the stimuli from the simple reaction time test on the ANAM

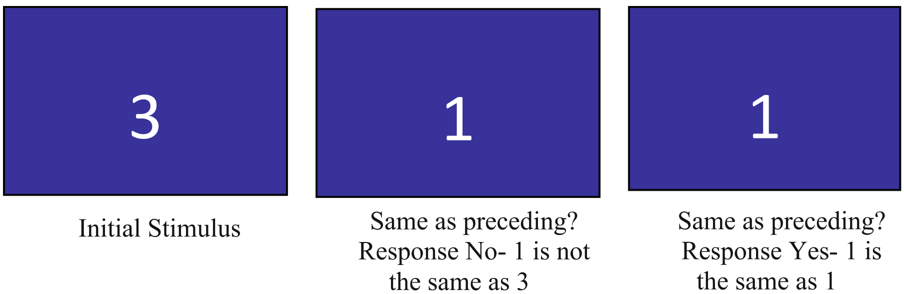


Fig. 2. An example of the stimuli from the ANAM RMCP.

Temporal Training Intervention Using the Interactive Metronome. The Interactive Metronome is a temporal training software which makes use of peripheral hardware known as 'triggers' for the hands and feet to record responses to the software training prompts. Volunteers attempt to match a reference tone by performing a prescribed exercise such as clapping in a circular motion or toe tapping. Visual references known as 'guides' were shown as colored circles indicate how close or far away in timing an individual's response is from the reference tone. Volunteers also received auditory guides of performance (beeps and buzzes) to indicate their proximity to the reference tone on select exercises.

Prior to training, both volunteer groups participated in a baseline IM training and assessment. Volunteers were taught the tasks by watching an instructional video, seeing a demonstration by research staff, and then by performing the task themselves. While this assessment was administered again post training, those data are not presented in this paper. The pre and post IM assessments consisted of 14 hand and foot exercises that lasted for approximately 30 min. These assessments were completed as part of a larger study.

The TTI consisted of 12 IM training sessions of approximately 45 to 60 min each, with each session occurring at least 24 h subsequent to the preceding session. During training, the volunteer wore a headset and heard a rhythm-based reference sound which they attempted to match as closely as possible by tapping a trigger with their hands, feet, or a combination of both. After each iteration of hearing and responding to the reference sound, the volunteer received biofeedback as a guide tone and a visual representation informing them of whether their response occurred before or after the reference tone. The visual representation showed several levels of feedback so volunteers would know if they were close to the reference tone or far from the tone in terms of milliseconds (ms). Volunteers would also hear a different tone and see a visual representation that let them know if they were 'right on' the reference tone (within ± 15 ms of the beat). Training sessions are scored and volunteers were aware of their scores on the 12 training sessions. Training time increases over the 12 sessions, first lasting a few minutes per task and increasing over subsequent sessions to approximately 45 min for a single task.

2.4 Data Analysis

Repeated Measure Analyses of variance (ANOVAs) were conducted using IBM SPSS Statistics for Windows version 22. The study used a mixed 2×2 factorial design comparing outcomes of the training groups (temporal training vs no training) over time (pre- and post-training). A significance level of 0.05 was set for all statistical analyses and were performed.

Prior to analysis, all data were inspected for outliers utilizing standardized scores (z-scores) set at ± 3 standard deviations from the mean and Tukey's boxplots. Identified extreme scores were further scrutinized to determine if they were, in fact, a valid score or some other factor affected the outcome such as testing interruption (e.g., participant falling asleep) or a participants' misunderstanding the test directions. The scores that were not a true capture of participant performance were removed from the analysis.

3 Results

3.1 Demographics

Demographics can be seen in Table 1.

Table 1. Demographics of study population.

Group	Men (n)	Women (n)	Mean age (Range)
TTI	16	3	18–44
Control	18	5	18–42

3.2 Simple Reaction Time (SRT)

Reaction time for correct responses did not differ by time ($F(1,40) < 0.001$, $p = 0.995$) or group ($F(1,40) = 0.085$, $p = 0.772$). An interaction between group and time was observed but was not statistically significant ($F(1,40) = 2.67$, $p = 0.110$) (Fig. 3a). The training group displayed a decrease in mean RT, while the control showed an increase from pre- to post-training.

Throughput scores did not differ for time ($F(1,40) = 0.165$, $p = 0.686$) or group ($F(1,40) = 0.001$, $p = 0.979$). An interaction between group and time was observed but was not statistically significant ($F(1,40) = 1.024$, $p = 0.318$) (Fig. 3b).

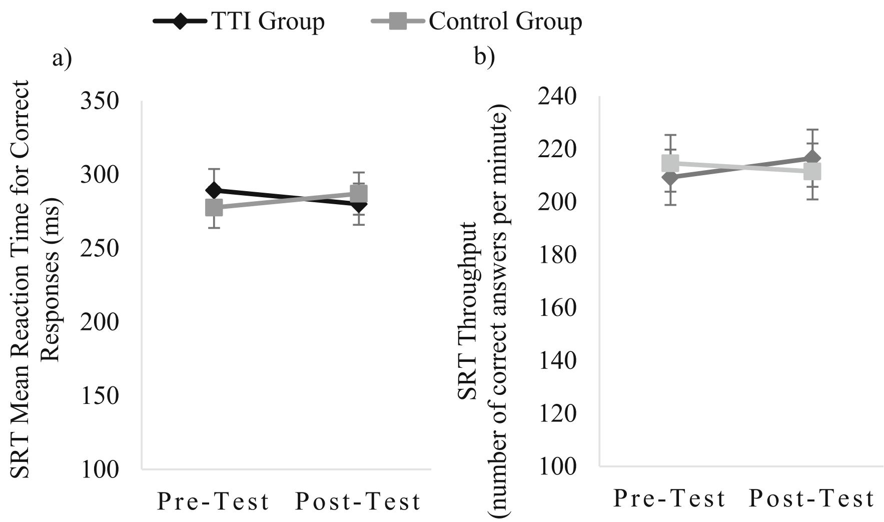


Fig. 3. ANAM scores for (a) mean RT for the correct answers and (b) throughput.

3.3 Running Memory Continuous Performance Test (RMCP)

RMCP mean reaction time for correct responses did not differ by time ($F(1,33) = 0.064$, $p = 0.801$) or group ($F(1,33) = 0.771$, $p = 0.386$) (Fig. 4a). The temporal training group showed a slight, non-significant ($p > .05$) decrease in reaction time compared to the control group.

RMCP throughput scores did not differ by time ($F(1,33) = 0.212$, $p = 0.648$) or group, ($F(1,33) = 0.590$, $p = 0.448$). The temporal training group displayed an increase in throughput, compared to the control group, but the difference was not statistically significant (Fig. 4b).

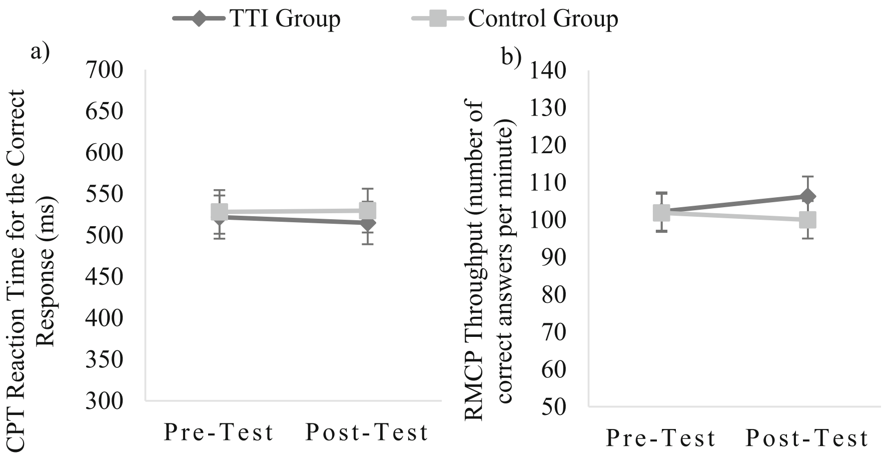


Fig. 4. Mean reaction time of correct responses (a) and throughput (b) display differences, yet not statistically significant, between groups from pre- to post-training.

4 Discussion

Although this study failed to find any statistically significant difference between the groups over time on the selected ANAM measures, the non-significant differences observed for both the SRT and the RMCP tasks followed the same trend of those receiving temporal training decreasing their reaction times for correct responses, while the control group times increased or remained the same. Likewise, throughput scores increased (improved) for those receiving temporal training, but decreased (worsened) for the control groups.

Mean RT for correct responses on the SRT for the temporal training group were faster relative to the control group. The findings suggest the training had an effect on simple reaction time performance by decreasing the TTI groups' latency in response to the target. A similar result was found for the SRT throughput scores as well. The same pattern was seen for the RMCP scores, the reaction times for correct responses were faster for the intervention group, while the control group scores remained the same.

Similarly, the RMCP throughput measure improved for those receiving training, while the control group worsened.

As noted previously, simple reaction time is a measure of attention (visuo-motor reaction time & vigilance) and small improvements were seen in SRT among those receiving training in timing and rhythm. In addition, RMCP measures attention, concentration and working memory. Barrouillet et al. [39] showed that attention plays an important part in the working memory retrieval process by attenuating the cognitive load. Kane and Engle [40] postulated that working memory oversees attentional control and the ability to ignore distractors was just as important as active maintenance of recalled items. Tang and Posner [41] showed that attention training in adults improved visual information processing and fluid intelligence (i.e., reasoning and problem-solving). Thus, attention and working memory are intrinsically linked to one another.

Stated another way, a training task that requires consistent auditory and visual attention, along with visuo-motor performance, for long periods of time, may be beneficial to the refinement of cognitive processes, due to the sharing of neural networks [42]. More traditional cognitive training for working memory has had limited success and computerized attention training results have been mixed [43].

The trends in this study suggest there may be benefits to incorporating temporal training (within cognitive or weapon firing training) for military service members, however further research with greater number of subjects is needed. Individuals with poor coordination or low attention spans may derive more benefit from temporal training than those with high levels of coordination and focused attention. It is suggested that future research focus on these questions.

The research results from this study demonstrate consistent trends supporting improvements in of reaction time and working memory following IM training. While our findings did not reach significance, improvements in sustained attention and reaction time, although small, could mean the difference between life and death on a battle field.

5 Limitations

The primary limitations of this study were the small sample size, the conditions surrounding the sample, and the sample population itself. A larger sample size is more representative of the population and limits the influence of outliers or extreme observations. Second, the sample population were military trainees involved in demanding training schedules. Several participants displayed impairments in performance attributed largely to lack of sleep (they fell asleep during testing). Finally, this research was conducted with active duty military personnel and may not be generalizable to the overall public. Future research to control these factors is suggested.

Acknowledgements. The views expressed in this article are those of the author(s) and do not reflect the official policy or position of the Department of the Army, Department of Defense, U.S. Army Research Laboratory, or the U.S. Government.

References

1. Baiocchi, D.: Measuring Army Deployments to Iraq and Afghanistan. Rand Corporation, Santa Monica (2013)
2. Cox, M., Military.com: US Army Announces 1,800 Soldiers to deploy to Iraq, Afghanistan. http://www.military.com/daily-news/2016/01/08/us-army-announces-1800-soldiers-to-deploy-to-iraq-afghanistan.html
3. United States Army: The U.S. Army concept for the human dimension in full spectrum operations (TRADOC Pamphlet 525–3-7). United States Army, Washinton, DC (2008)
4. Casey, G.W.: Comprehensive soldier fitness: a vision for psychological resilience in the U.S. army. Am. Psychol. **66**(1), 1–3 (2015)
5. Pascual-Leone, A., Amedi, A., Fregni, F., Merabet, L.B.: The plastic human brain cortex. Annu. Rev. Neurosci. **28**, 377–401 (2005)
6. Rabipour, S., Raz, A.: Training the brain: fact and fad in cognitive and behavioral remediation. Brain Cogn. **79**(2), 159–179 (2012)
7. Kane, M.J., Engle, R.W.: The role of prefrontal cortex in working memory capacity, executive attention and general fluid intelligence: an individual - differences perspective. Psychon. Bull. Rev. **9**(4), 637–671 (2002)
8. Elhalal, A., Davelaar, E.J., Usher, M.: The role of the frontal cortex in memory: an investigation of the Von Restorff effect. Front. Hum. Neurosci. **8**, 410 (2014)
9. Stuss, D.T., Knight, R.T.: Principles of Frontal Lobe Function. Oxford University Press, New York (2002)
10. Meck, W.H., Benson, A.M.: Dissecting the brain's internal clock: how frontal-striatal circuitry keeps time and shifts attention. Brain Cogn. **48**(1), 195–211 (2002)
11. Thaut, M.H., Trimarchi, P.D., Parsons, L.M.: Human brain basis of musical rhythm perception: common and distinct neural substrates for meter, tempo, and pattern. Brain Sci. **4** (2), 428–452 (2014)
12. Benchenane, K., Tiesinga, P.H., Battaglia, F.P.: Oscillations in the prefrontal cortex: a gateway to memory and attention. Curr. Opin. Neurobiol. **21**(3), 475–485 (2011)
13. Morgan, C.A., Doran, A., Steffian, G., Hazlett, G., Southwick, S.M.: Stress-induced deficits in working memory and visuo-constructive abilities in special operations soldiers. Biol. Psychiatry **60**(7), 722–729 (2006)
14. Interactive Metronome. http://www.interactivemetronome.com/
15. Karmarkar, U.R., Buonomano, D.V.: Timing in the absence of clocks: encoding time in neural network states. Neuron **53**(3), 427–438 (2007)
16. Leisman, G., Melillo, R.: Effects of motor sequence training on attentional performance in ADHD children. Int. J. Disabil. Hum. Dev. **9**(4), 275–282 (2010)
17. Jacobs, K.M., Donoghue, J.P.: Reshaping the cortical motor map by unmasking latent intracortical connections. Science **251**(4996), 944–947 (1991)
18. Dahlin, E., Neely, A.S., Larsson, A., Backman, L., Nyberg, L.: Transfer of learning after updating training mediated by the striatum. Science **320**(5882), 1510–1512 (2008)
19. Taub, G.E., McGrew, K.S., Keith, T.Z.: Effects of improvements in interval timing on the mathematics achievement of elementary school students. J. Res. Child. Educ. **29**(3), 352–366 (2015)
20. Carr, K.W., Fitzroy, A.B., Tierney, A., White-Schwoch, T., Kraus, N.: Incorporation of feeback during beat synchronization is an index of neural maturation and reading skills. Brain Lang. **164**, 43–52 (2017)

21. Rice, V.J., Marra, D., Butler, J.: Neuro-cognitive assessment, symptoms of attention deficit and hyperactivity disorder, and soldier performance during 68W advanced individual training. (ARL Publication No. ARL-TN-469). Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground (2007)
22. Trujillo, L., Painter-Patton, J.: The Validation of the Interactive Metronome: a pilot study prior to implementation for post deployment service members. J. Neurol. Neurosci. **6**, 2–7 (2015)
23. Cosper, S.M., Lee, G.P., Peters, S.B., Bishop, E.: Interactive metronome training in children with attention deficit developmental coordination disorders. Int. J. Rehabil. Res. **32**(4), 331–336 (2009)
24. Shaffer, R.J., Jacokes, L.E., Cassily, J.F., Greenspan, S.I., Tuchman, R.F., Stemmer, P.J.: Effect of interactive metronome training on children with ADHD. Am. J. Occup. Ther. **55**(2), 155–162 (2001)
25. Taub, G.E., Lazarus, P.J.: The effects of training in timing and rhythm on reading achievement. Contemp. Issues Educ. Res. **5**(4), 343 (2012)
26. Silverman, S.J.: Effecting peak athletic performance with neurofeedback, interactive metronome, and EMDR: a case study. Biofeedback **39**(1), 40–42 (2011)
27. Golf Digest: The beat is on: some PGA tour pros believe timing is everything, vol. 54, p. 60 (2003)
28. Sommer, M., Ronnqvist, L.: Improved motor-timing: effects of synchronized metronome training on golf shot accuracy. J. Sports Sci. Med. **8**(4), 648–656 (2009)
29. Dickey, N.: Defense health board recommendation memorandum pertaining to automated neuropsychological assessment metrics 2011-10. Defense Health Board, Leesburg Pike Falls Church (2011)
30. Reeves, D.L., Bleiberg, J., Roebuck-Spencer, T., Cernich, A.N., Schwab, K., Ivins, B., Salazar, A.M., Harvey, S.C., Brown, F.H., Warden, D.: Reference values for performance on the automated neuropsychological assessment metric V3.0 in an active duty military sample. Mil. Med. **171**(10), 982–994 (2006)
31. Cepeda, N.J., Blackwell, K.A., Munakata, Y.: Speed isn't everything: complex processing speed measures mask individual differences and developmental changes in executive control. Dev. Sci. **16**(2), 269–286 (2013)
32. Cernich, A., Reeves, D., Sun, W., Bleiberg, J.: Automated neuropsychological assessment metrics sports medicine battery. Arch. Clin. Neuropsychol. **22S**(1), S101–S114 (2007)
33. Short, P., Cernich, A., Wilken, J.A., Kane, R.L.: Initial construct validation of frequently employed ANAM measures through structural equation modeling. Arch. Clin. Neuropsychol. **22S**(1), S63–S77 (2007)
34. Etra, J.L.: The effect of interactive metronome training on children's SCAN-C scores, Doctoral dissertation, Nova Southeastern University (2006)
35. Kuhlman, K., Schweinhart, L.J.: Timing in child development. High/Scope Educational Research Foundation (1999)
36. Mulder, R.: Training for the brain'technology yields academic gains at St. Thomas Aquinas HS. J. (Technol. Horiz. Educ.) **30**(3), 52 (2002)
37. Greitzer, F.L., Andrews, D.H.: Training strategies to mitigate expectancy-induced response bias in combat identification: a research agenda. In: Andrews, D.H., Herz, R.P., Wolf, M.B. (eds.) Human Factors in Combat Identification, pp. 173–189. Ashgate Publishing Company, Burlington (2010)
38. Abernethy, B., Maxwell, J.P., Jackson, R.C., Masters, R.S.W.: Skill in sport. In: Durso, F.T., Nickerson, R.S., Dumais, S., Lewdandosky, S., Perfect, T.J. (eds.) Handbook of Applied Cognition, 2nd edn. Wiley, Hoboken (2007)

39. Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., Camos, V.: Time and cognitive load in working memory. J. Exp. Psychol. Learn. Mem. Cogn. **33**(3), 570 (2007)
40. Kane, M.J., Engle, R.W.: Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to stroop interference. J. Exp. Psychol. Gen. **132**(1), 47–70 (2003)
41. Tang, Y.-Y., Posner, M.I.: Attention training and attention state training. Trends Cogn. Sci. **13**(5), 222–227 (2009)
42. LaBar, K.S., Gitelman, D.R., Parrish, T.B., Mesulam, M.M.: Neuroanatomic overlap of working memory and spatial attention networks: a functional MRI comparison within subjects. NeroImage **10**(6), 695–704 (1999)
43. Melby-Lervag, M., Hulme, C.: Is working memory training effective? A meta-analytic review. Dev. Psychol. **49**(2), 270 (2013)

Neurocognitive Temporal Training for Improved Military Marksmanship: Grouping and Zeroing

Gary L. Boykin Sr.(✉) and Valerie J. Rice

U.S. Army Research Laboratory, HRED-ATSD,
Army Medical Department Filed Element, JBSA, Fort Sam Houston,
San Antonio, TX 78234, USA
gary.l.boykin.civ@mail.mil

Abstract. Soldiers' must adjust their rifle so fired rounds travel on the correct trajectory to strike an intended target. This process, known as shot grouping and zeroing, is essential to weapon firing accuracy. The purpose of this study was to determine whether neurocognitive temporal training (NTT) would improve grouping and zeroing precision. Research volunteers were mostly male (93%) and between the ages of 18 and 44 (mean and SD = 28.16 ± 8.1). No differences in shot group size, height and width shot patterns, or the number of rounds expended among those receiving NTT (n = 19), compared with a control group (n = 24) (p > .05). These findings are contrary to other findings demonstrating improved eye-hand coordination following NTT. A greater impact of NTT may be seen in tasks that are novel, complex, or performed under duress, as opposed to the more familiar tasks selected for this research.

Keywords: Interactive Metronome · Marksmanship · Shooting · Accuracy

1 Introduction

U.S. military soldiers are required to train and demonstrate proficiency in marksmanship [1]. Soldiers receive training with the M16A4 rifle and must meet qualification standards using the M16A4 throughout their military career [2–4], including adjusting their rifle prior to firing using techniques known as grouping and zeroing (G&Z). Service members shooting the M16A4 for the first time, or those that have not fired the rifle for prolonged periods may have difficulty completing grouping and zeroing tasks. This is problematic due to their incurring additional training time (time on the shooting range and time for instructor involvement), along with increased financial costs for ammunition on live firing ranges. Shooting an M16A4 rifle is associated with multi-functional human performance tasks and sequential timing efforts [3, 5, 6], including establishing visual alignment between rifle sights and targets, coordinating breathing with trigger pull, depth perception, eye and hand coordination, body-movement correctness, target interpretation, and an even trigger pull [8, 9, 13, 22, 24, 32].

These behaviors require an innate sense of timing and rhythm, suggesting neuro-cognitive temporal training may improve shooting standards or increase

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_7

marksmanship accuracy. The focus of this research is to investigate M16A4 grouping and zeroing performance among military service members in a simulated shooting environment. It was expected there would be no difference in marksmanship outcome measures between volunteers completing twelve neurocognitive temporal training sessions in addition to routine military marksmanship training, and a control group receiving routine military marksmanship training only.

2 Literature Review

2.1 Interactive Metronome

The Interactive Metronome® (IM) is a computer-based neurocognitive training device designed to help improve timing, rhythm, motor coordination, and focused attention [8] The IM is based on the premise that all humans operate with an internal clock, while sleeping or awake [10]. During training, individuals attempt to match a reference tone as closely as possible by tapping a trigger using their hands, feet, or a combination of both hands and feet. Reaction times are measured in milliseconds, and score composites are created after each task is completed. The IM has been shown to be an effective augment intervention for attention deficit and hyperactivity disorder and gait rehabilitation [8, 9, 13, 22, 24, 32].

Possible explanations of why IM training influences complex motor performance rely on the following theories. First, complex human performance processes are dependent on an internal sense of timing and rhythm [10, 11]. An internal sense of timing and rhythm influences motor control (planning, sequencing, coordination) and attention (focus, ability to pay attention, concentration) [8, 15–17, 32]. Thus, it would be expected that a developed sense of timing and rhythm would positively influence marksmanship performance. For example, marksmanship deficits during grouping and zeroing may be due to a lack of procedural coordination, such as omnidirectional sight movement for acquiring a target. A second theory is that repeated neural sequences can develop new pathways or strengthen existing pathways, and these improvements might generalize to other tasks [11, 12, 18]. The areas of the brain activated during IM use include the prefrontal region, the cingulate gyrus, basal ganglia, and the medial brainstem [11]. The prefrontal area is responsible for planning, sequencing, analysis, decision-making, high-order motor control and working memory [10, 14]. The cingulate gyrus allows shifting of attention [18], the basal ganglia integrates thought and movement, and the medial brainstem is involved in neuro-motor control [12]. Theoretically, it would follow that IM training could lead to improvements in these areas of functioning, which are important for complex tasks such as firing the M16A4 rifle.

Recent literature suggests that the IM has the potential for "improving life satisfaction and reducing a loss of attention" among service members [19]. Although most studies investigating the use of the IM with military populations argue its utility in health care [19, 20], athletes have benefited from IM training [21, 23]. Research has shown improvements in golfing (golf ball placement accuracy) and soccer performance following IM training [21, 23]. The military task of weapon firing requires accurate, fast psychomotor performance, similar to that of professional athletes. Therefore, training that improves athletic performance may also improve marksmanship performance.

2.2 Military Marksmanship

Shot grouping and zeroing (G&Z) are used to assess the initial mechanical functionality of the M16A4 and the shooter's ability to strike a target in cluster patterns in the proximity of center mass [1, 25]. Although grouping and zeroing are independent of each other, together they comprise the sight alignments and individual adjustments needed for accurate shot placement [2, 25]. Literature from 1986 to 2012 suggests grouping and zeroing M16 class rifles (to include the A4 model) are essential for shooting accuracy, and that grouping and zeroing tasks should be completed before assessing accuracy measures [2, 3, 25–27, 29]. Grouping and zeroing ensure the point of aim is accurately aligned with the target. The standards for grouping and zeroing procedures are identical across three branches of the armed services [2–4].

Prior shooting experience is linked to improved shooting performance [30, 31]. Stated another way, cognitive aptitude regarding marksmanship develops with experience and correlates with shooting performance [6, 7]. Therefore, it is a reasonable expectation that service members shoot at different competency levels, and their time to complete each associated G&Z task will vary, along with the number of rounds they expend. That is, less experienced shooters may require more time to complete these tasks [2, 25, 26].

2.3 Simulated Shooting Shelter

The use of simulated firing ranges are prevalent in the military [28, 29]. While comparatively, concepts and procedures for M16A4 G&Z are similar for both a simulated environment and live fire range shooting [3, 28, 29]; there are differences for mechanically adjusting the rifle using the Engagement Skills Trainer (EST) 2000 (weapon-firing simulator shelter) and environmental conditions differ. During non-simulation G&Z, it is incumbent upon the shooter to manipulate the external components of the M16A4 rifle, for example, manually adjusting front and rear sights [2, 3]. Conversely, G&Z within the EST 2000 simulator does not require the shooter to make mechanical adjustments to the rifle. Shooters are only required to make changes to muzzle support elements when shooting in the supported prone position, e.g., sandbag(s). As such, the automated computer system identifies target bullet placement and automatically makes sighting adjustments after firing each shot group. Environmental conditions such as wind speed and direction are also preset using the same automated internal control mechanism. The procedure section further describes G&Z, and provides targeting illustrations. As part of this research study, the G&Z process was mandatory and satisfactorily completed before participants proceeded with precision and accuracy testing.

3 Methods

3.1 Participants

Participant recruitment of U.S. service members commenced under the authority, review and approval of the Institutional Review Board at the San Antonio Military Medical Center, Fort Sam Houston, Texas. Recruitment occurred within the Joint Base San Antonio area. After receiving full disclosure of the study requirements, including associated risks and use of the Engagement Skills Trainer 2000. Each participant rendered written volunteer consent to participate in the study and was randomized into either the neurocognitive temporal training (NTT) group or the control group. Both groups had received their annual routine marksmanship training, also both groups completed pre-testing requirements. The NTT group attended twelve IM training sessions. Post-testing was completed after the twelve NTT training sessions or after a comparable amount of time for the control group. Participants received no compensation (monetary or otherwise) for their involvement in the study, and could withdraw from the study at any point in time, without recourse. All participant documents and folders were giving specialized codes to prevent identity disclosure and secured in a double locked facility accessible only to researchers identified in the IRB protocol.

3.2 Questionnaires

The demographic questionnaire contained background information, i.e., age, gender, military status, rank, time-in-service (both active and reserve, if applicable). Participants in this study had completed basic training and basic rifle marksmanship training, therefore, their previous marksmanship qualification level of expertise, e.g., Marksman, Sharpshooter, Expert, was recorded.

3.3 Engagement Skills Trainer (EST) 2000 Outcome Measures

EST shooting performance outcome data was collected directly from the shelter's enhanced computer system consisting of: (1) Shot group size (a measure of distance between each shot (smaller shot group size is indicative of better marksmanship performance), (2) Bullet placement Height and Width (H&W) (Shots in proximity to target's center of mass), (3) The number of bullets (rounds) fired to obtain an acceptable grouping pattern, and (4) The number of bullets fired to obtain an acceptable zeroing pattern.

3.4 Procedures

Grouping (Description and Test Procedures). All participants performed grouping tasks in a prone supported position using the M16A4 rifle. Targeting distance consisted of 25-meters from the firing line to simulated target screen. Figure 1 shows the M16A4 rifle used by participants.

Fig. 1. M16A4, reprint from Global Security Organization, 2004. Rifle used by research participants

Although the EST 2000 firing line was capable of facilitating ten shooters at any given time, a maximum of six shooters were assessed (fired the rifle) at once. The rationale was to allow researchers (1 researcher per 2 shooters) sufficient time to manually record height and width data and the cumulative number of rounds fired. Environmental controls were implemented for improved shooting, that is, the simulated environment was lit with ambient lighting, and the shot sound emitted from the M16A4 was controlled at a moderate decibel level ($\sim < 60$ dB). The EST 2000 operator provided firing commands and operational instructions per their standardized procedure. For example, instructing participants to “assume a good prone support position”; “place the safety selector switch from safe to semi”; “watch your lane”; “your shot pattern is reflective of your breathing, so you should pause (in breathing) when squeezing the trigger.” This portion of the grouping process was critical, as the feedback imparted by the instructor operator during weapon firing can influence performance outcomes [1, 5, 7, 30]. The EST 2000 operator also ensured the rifles functioned properly, periodically assessing the overall firing condition of the simulated rifles. When detecting rifle defects or malfunctions that could potentially hinder shooting performance, the EST operator either made on-the-spot repairs, or removed and replaced the participant’s rifle with a fully functional rifle.

As defined by Army Field Manual 3-22.9, 2008 to successfully group rounds (bullets), the shooter must cluster six of six consecutive rounds at less than 4 cm. To achieve the grouping standard, a series of three rounds were fired in one set, with multiple iterations until a 4 cm cluster was formed. Figure 2 illustrates a cluster shot pattern after three rounds have been fired. Traditional live fire grouping methods would allow the shooter to triangulate and number their pattern on the target after each three round sequence [2, 3], however, this process was not necessary for simulated shooting. Figure 3 illustrates a grouping cluster after nine rounds (three consecutive sets). During the grouping phase, bullets may strike anywhere on the zeroing target to achieve an efficient cluster pattern. The <4 cm cluster pattern provides the shooter with sufficient

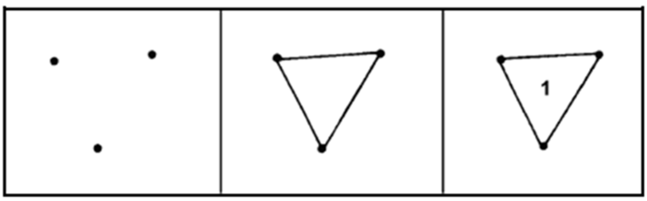


Fig. 2. Three Round Shot Group. Reprinted from Rifle Marksmanship M16-/M4 Series Weapon (pp. 5–4), by Headquarters Department of the Army, 2008, Washington, DC.

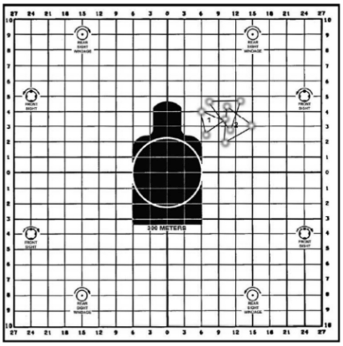


Fig. 3. Acceptable shot grouping performance. Reprinted from Rifle Marksmanship M16-/M4 Series Weapon (pp. 5–8), by Headquarters Department of the Army, 2008, Washington, DC.

information concerning sight alignment, shot group tightness, and physical factors associated with shooting the M16A4, such as firing patterns related to breathing inconsistencies during trigger squeeze. The grouping process is repeated until the participant meets the grouping standard or exceeds a maximum of 40 rounds to complete the task.

Zeroing (Description and Test Procedures). After achieving the grouping standard, participants zeroed the M16A4 rifle. While there are several zeroing processes for the M16A4 (each having a particular function), for the purpose of this study, only two zeroing procedures were applied: 1. Mechanical (MZ) and 2. Battle Sight (BSZ), each focused on aligning the rifle sights to the target. Firing line procedures and conditions were identical to those described in the grouping section. The following provides an explanation of the zeroing procedure standards for non-simulated shooting.

The purpose of MZ is to put the rifle in a neutral conformity [2–4], so Battlesight Zeroing can begin. Battlesight Zeroing requires the shooter to fire rounds at a zeroing target at a specified distance (25 m). Participants fired rounds in increments of three (equaling one set). Between each set, the EST computer system made adjustments to the front and rear sights based on where the rounds land on the zeroing target. Researchers collected height and width (H&W) data and the cumulative number of rounds fired between each set of zeroing shots. This process was repeated until five of six consecutive shots fell within 4 cm of the center mass. Figure 4 illustrates the final grouping and zeroing results after sight corrections. Notice six rounds are within the 4 cm circle, showing the Battlesight Zero criteria has been met. Both grouping and zeroing can be lengthy processes, the speed in which these tasks are completed depend on the shooter's ability to apply basic shooting fundamentals [27].

Statistics. Statistical analyses using IBM Statistical Package for the Social Sciences software produced the quantitative results with a significance level of 0.05. Specifically, the Mann-Whitney U test was used to examine between-group effects during pre-testing and post testing. The Mann-Whitney U was also used to examine within group pre/post scores for each outcome. Gain score analyses using a single factor ANOVA determined if the scores were significantly different between groups.

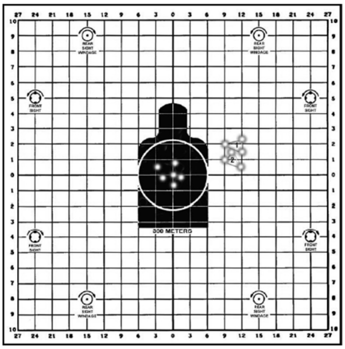


Fig. 4. Acceptable shot grouping and zeroing. Reprinted from Rifle Marksmanship M16-/M4 Series Weapon (pp. 5–8), by Headquarters Department of the Army, 2008, Washington, DC.

4 Results

4.1 Demographics

Forty-three active duty service members (19 treatment and 24 control) participated. The volunteer population was comprised of 93% men and 7% women. No differences were found in age for treatment group (mean age and standard deviation, 30 ± 8.8) and the control group (26 ± 7.2) ($p > .05$). Self-reported level of shooting expertise according to Army standards ranked as Expert, Sharpshooter, and Marksman. Over half of the participants reported themselves as Sharpshooter or Expert, with no difference between the treatment group (63%) and the control group (62%) ($p > .05$).

4.2 Pre- and Post-tests

There were no differences between groups during pre-testing and post-testing (see Tables 1, 2, 3, 4).

Table 1. Participants' average shot group size (Mann-Whitney U test).

	Pre-test		Post test	
	IM shooters $n = 24$	Control $n = 24$	IM shooters $n = 19$	Control $n = 24$
Avg. rank	25.37	19.33	25.13	19.52
	$z = -1.56, p > .05$		**$z = -1.45$, p > .05**	

Table 2. Participants' average height and width shot pattern (Mann-Whitney U test).

	Pre-test		Post test	
	IM shooters $n = 24$	Control $n = 24$	IM shooters $n = 19$	Control $n = 24$
Avg. rank	24.16	20.29	24.42	20.08
	$z = -1.00, p > .05$		**$z = -1.25$, p > .05**	

Table 3. Participants' shot grouping success by rounds expended (Mann-Whitney U test).

	Pre-test		Post test	
	IM shooters $n = 24$	Control $n = 24$	IM shooters $n = 19$	Control $n = 24$
Avg. rank	24.11	20.33	22.39	21.69
	$z = -1.32$, $p > .05$		**$z = -2.70$, p > .05**	

Table 4. Participants' shot zeroing success by rounds expended (Mann-Whitney U test).

	Pre-test		Post test	
	IM shooters $n = 24$	Control $n = 24$	IM shooters $n = 19$	Control $n = 24$
Avg. rank	22.29	20.98	24.53	20.00
	$z = -.646$, $p > .05$		**$z = -1.33$, p > .05**	

4.3 Change Score Analysis of Variance for All Outcome Measures

There were no differences in weapon firing performance among those receiving NTT and the control group using gain scores, for the four outcome measures ($p > .05$) (see Tables 5, 6, 7, 8). Therefore, the hull hypothesis was accepted. Figures 5 and 6 show the average number of rounds expended during grouping and zeroing.

Table 5. Single factor analysis of variance of group size

Source	*SS*	*df*	*MS*	*F*	*p*
Between groups	0.025	1	0.02	0.03	0.84
Within groups	26.90	41	0.65		
Total	26.93	42			

Table 6. Single factor analysis of variance of height and width

Source	*SS*	*df*	*MS*	*F*	*p*
Between groups	0.047	1	0.04	0.03	0.82
Within groups	39.89	41	0.97		
Total	39.94	42			

Table 7. Single factor analysis of variance of shot grouping success by rounds expended

Source	*SS*	*df*	*MS*	*F*	*p*
Between groups	62.49	1	62.4	2.14	0.15
Within groups	1194.5	41	29.1		
Total	1257	42			

Table 8. Single factor analysis of variance of shot zeroing success by rounds expended

Source	*SS*	*df*	*MS*	*F*	*p*
Between groups	10.32	1	10.32	0.26	0.61
Within groups	1620.1	41	39.51		
Total	1630	42			

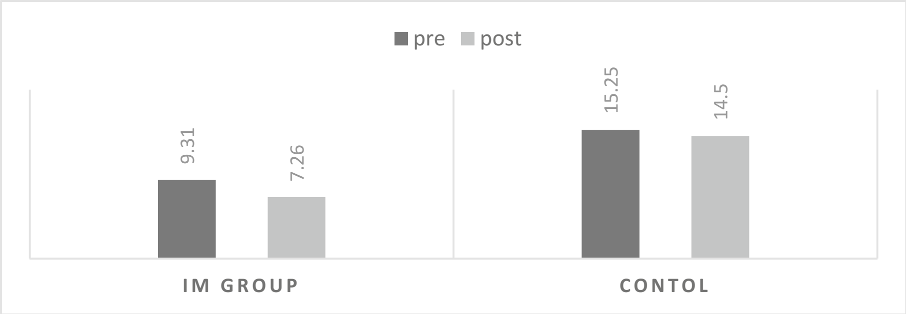


Fig. 5. Average number bullets used by participants to obtain a grouping pattern

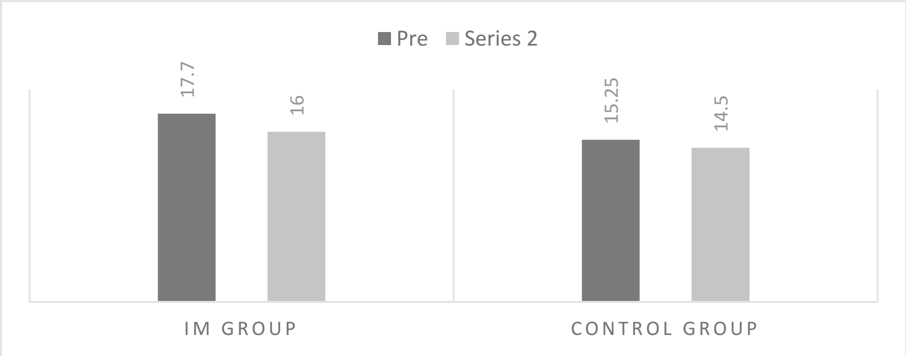


Fig. 6. Average number of bullets used by participants to obtain a zeroing pattern

5 Discussion

The purpose of this study was to evaluate the effects of neurocognitive temporal training on grouping and zeroing marksmanship performance among military service members. The effectiveness of the Interactive Metronome with grouping and zeroing shooting tasks among military participants after completing twelve training sessions did not improve shooting performance when compared to the control group. The results of this study diverged from other IM research involving eye-hand coordination and rhythmic timing [21, 23]. One plausible explanation for the results is that over half of the participants had scored either expert or sharpshooter and therefore, they were already proficient at grouping and zeroing the weapon. That is, grouping and zeroing in this scenario may not have presented a challenge for participants. As noted in the literature review, the more you shoot, the better you get [30, 31]. Although a significant difference was not found between groups, a practical significance was seen in the reduction of two rounds expended during grouping by the NTT group, compared with the control group. Hypothetically, firing two fewer rounds per person during annual weapon firing training would reduce both manpower and financial costs.

5.1 Limitations

The limitations of this study are the sample size and the military specific population. Care should be taken in generalizing the results to other populations.

6 Conclusion

The results of this research suggest neurocognitive temporal training has little benefit on grouping and zeroing marksmanship performance among soldiers already experienced in weapon firing. Future studies on temporal training might target more complex or novel weapon firing tasks, alternate weapons, different firing positions, with a larger number of participants.

Acknowledgements. We would like to thank the soldiers who volunteered for this research, as well as the members of their Chain-of-Command. We also thank our colleagues, Dr. Leah Enders, Mr. Cory Overby, Ms. Angela Jeter, Ms. Jessica Villarreal from DCS Corp, and EST operators Mr. Gill Gallardo and Mr. Shelton Reyes (CSRA Corp) for their outstanding support. The views expressed in this article are those of the author and do not reflect the official policy or position of the Department of the Army, Department of Defense, U.S. Army Research Laboratory or the U.S. Government.

References

1. James, D., Dyer, J.: Rifle Marksmanship Diagnostic and Training Guide. U.S. Army Research Institute for Behavioral and Social Sciences, Fort Benning (2011)
2. Department of the Airforce: Combat arms training program individual use weapon (2004)
3. Headquarters Department of the Army: Field Manual 3-22-9: Rifle marksmanship M16/M4 series weapons. Department of the Army, District of Columbia (2008)
4. US Marine Corp.: Rifle Marksmanship MCRP 3-01A, Washington D.C. (2012)
5. Amburn, C., Goldberg, B., Brawner, K.: Steps Towards Adaptive Psychomotor Instruction. U.S. Army Research Laboratory, Orlando (2014)
6. Evans, K., Osborne, A.: The Development and Implementation of Basic, Advanced and Unit Rifle Marksmanship Training Programs. U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Benning (1988)

7. Chung, G., Nagashima, S., Espinosa, P., Berka, C., Baker, E.: An Exploratory Investigation of the Effect of Individualized Computer-Based Instruction on Rifle Marksmanship Performance and Skill. National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles (2009)
8. Koomar, J., Burpee, J., DeJean, V., Frick, S., Kawar, M., Fischer, D.: Theoretical and clinical perspectives on the interactive metronome: a view from occupational therapy. Am. J. Occup. Ther. **55**(2), 163–166 (2001)
9. Shaffer, R., Jacokes, L., Cassily, J., Greenspan, S., Tuchman, R., Stemmer, P.: Effect of interactive metronome training on children with ADHA. Am. J. Occup. Ther. **55**, 155–162 (2001)
10. McGrew, K.: The science behind Interactive Metronome, MindHub (2013)
11. Alpiner, N.: The role of functional NRI in defining auditory - motor processes networks. In: Paper Presented at the National Learning & the Brain Conference at Harvard, Cambridge (2005)
12. Kandell, E., Schwartz, J., Jessell, T.: Principles of Neural Science in Principles of Neural Science, 4th edn. McGraw-Hill, New York (2001)
13. Namgung, Y., Son, D., Kim, K.: Effect of interactive metronome training on timing, attention and motor function of children with ADHD: case report, Inje University (2015)
14. Elhalal, A., Davelaar, E.J., Usher, M.: The role of the frontal cortex in memory: an investigation of the Von Restorff effect. Fron. Hum. Neurosci. **8**, 410 (2014)
15. Carr, K., Fitzroy, A., Tierney, A., White-Schwoch, T., Kraus, N.: Incorporation of feedback during beat synchronization is an index of neural maturation and reading skills, National Center for Research on Evaluation, Standards, and Student Testing. Elsevier, pp. 43–52 (2016)
16. Pascual-Leone, A., Amendi, A., Fregni, F., Merabet, L.: The plastic human brain cortex. Annu. Rev. Neurosci. **28**, 377–401 (2005)
17. Kane, M.J., Engle, R.W.: The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. Psychon. Bull. Rev. **9**(4), 637–671 (2002)
18. Davis, K., Hutchison, W., Lozano, A., Tasker, R., Dostrovsky, J.: Human anterior cingulate cortex neurons modulated by attention-demanding tesks. J. Neurophysiol. **83**, 3575–3577 (2000)
19. Ridenhour, K.: Evaluating the effectiveness of the Interactive Metronome in improving life satisfaction and reducing attention in Marines with mild traumatic brain injuries associated with post-traumatic stress disorder symptoms: a pilot study of protocols. East Carolina University (2011)
20. Trujillo, L., Painter-Patton, J.: The validation of the interactive metronome: a pilot study prior to implementation for post deployment service members. J. Neurol. Neurosci. **6**, 2–7 (2015)
21. Sommer, K.: Effect of timing training in golf and soccer players (2014)
22. Shank, T., Harron, W.: A retrospective outcomes study examining the effects of interactive metronome on hand function. J. Hand Ther. **28**(4), 133–137 (2015)
23. Libkuman, T., Otani, H.: Training in timing improves accuracy in golf. J. Gen. Psychol. **129**, 77–96 (2002)
24. Beckelhimer, S., Dalton, A., Richter, C., Hermann, V., Page, S.: Computer-based rhythm and timing training in severe, stroke-induced arm hemiparesis. Am. J. Occup. Ther. **65**, 96–100 (2011)
25. Taylor, C., Dyer, F., Osborne, A.: Effects of Rifle Zero and Size of Shot Group in Marksmanship. Research Institute for the Behavioral and Social Sciences, Fort Benning (1986)

26. Dyer, J., Schafer, P., Bink, M., James, D., Wampler, R., Dlubac, M.: Soldier Performance on a New Marksmanship Course of Fire. U.S Army Research Institute for the Behavioral and Social Sciences, Fort Benning (2010)
27. Dyer, J., Lipinski, J., Goodwin, G.: Assessment of New Marksmanship Strategies in 2010. U.S. Army Research Institute, Fort Benning (2011)
28. Schwetje, C.: Intergrating Intelligence and Building Teams Within the Infantry Immersion Trainer. Naval Postgraduate School, Monterey (2009)
29. Scribner, D., Wiley, P., Harper, W.: A Comparison of Live and Simulated Fire Soldier Shooting Performance. Army Research Laboratory, Aberdeen (2007)
30. Chung, G., Delacruz, L., Kim, J., Bewley, W., deSouza, A., Sylvester, R., Baker, E.: Determinates of Rifle Marksmanship Performance: Predicting Shooting Performance With Advanced Distribution Learning. Office of Naval Research, Los Angeles (2004)
31. Lipinski, J., James, D., Wampler, R.: Test of a Prior Marksmanship Knowledge Predictor Test. U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Benning (2014)
32. Taub, G., McGrew, K., Keith, J.: Effects of improvements in interval timing on the mathematics achievement of elementary school students, J. Res. Child. Edu. 352–366 (2015)

Audition and Workload

Neurofeedback for Personalized Adaptive Training

Jesse Mark[1], Neha Thomas[1], Amanda Kraft[2], William D. Casebeer[2], Matthias Ziegler[2], and Hasan Ayaz[1,3,4(✉)]

[1] School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, PA, USA
ayaz@drexel.edu
[2] Advanced Technology Laboratories, Lockheed Martin, Arlington, VA, USA
[3] Department of Family and Community Health, University of Pennsylvania, Philadelphia, PA, USA
[4] Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

Abstract. Learning of complex skills can be enhanced if the training can adapt to the learner. Earlier studies utilized behavioral performance based adaptation; however, behavioral performance assessment is limited and does not take into account the mental effort and the brain plasticity changes during the acquisition of complex skills. In this study, we utilized objective brain based measures for the assessment and adaptation of a four-day training program with three piloting tasks on four participants using a low fidelity flight simulator. Functional near-infrared spectroscopy (fNIRS) from prefrontal cortex was measured to adapt the difficulty levels of the training trials of the tasks with the aim of optimizing the cognitive workload. Participants also performed reference practice trials that had the same difficulty across sessions. Preliminary results identified specific brain areas within prefrontal cortex for each reference tasks that corroborates earlier task practice studies. Furthermore, the same brain areas were responsive to the adaptive training trials as well. The results overall suggest task specific brain areas are coupled with the behavioral performance. This study outlines a new neurofeedback based training paradigm using the wearable and portable fNIRS. Future uses of such personalized training could help prevent unnecessary over-training or insufficient under-training.

Keywords: fNIRS · Prefrontal cortex · Adaptive training · Cognitive workload

1 Introduction

Human mental workload is critical for understanding and augmenting complex human-machine teaming. Mental workload that indicates how hard the brain is working to meet task demands can be dissociated from performance output [1]. Accurate assessment of mental workload could help in preventing operator error and allow for pertinent intervention by predicting performance decline that can arise from either work overload or understimulation [1–3]. The augmentation of complex human-machine team performance and its transfer to improved functioning at work or in everyday

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_8

settings via alteration of the operator's underlying neurocognitive processes is a prime goal of neuroergonomics [4, 5]. Practice-based repetitive exposure of operator to cognitive task is common and one potential method for enhancement of such neurocognitive capacity. A growing body of neuroergonomics studies suggest that mental workload and skill acquisition for complex real-world tasks such as airplane piloting and medical surgery operations can be monitored for assessment and enhancement [6–8].

Traditionally, the metrics for decision of how and when a learner graduates or progresses to the next level depends on either experience (i.e. mostly measured by time on task) or behavioral performance as measured by theoretical knowledge exams and practical skill application tests. However, these measures do not take into account the neurocognitive processes, rendering them inefficient. This results in learners being undertrained when level advancement is too fast, or over-trained when advancement is too slow [9, 10].

To overcome these shortcomings, personalization of the training paradigm using neurocognitive information, also known as neurofeedback, is needed [11, 12]. Personalization in this context may also be called dynamic difficulty adjustment based on individual behavioral metrics in which the level of each assigned task is modified based on constant or regular (e.g. hourly, daily) performance. Neuroimaging studies have shown that this type of adaptive training leads to improved outcomes and less cognitive workload than otherwise [11, 13]. Past research includes an air traffic control task modulated by number of aircraft, a driving/dual task protocol, and neuroadaptive piano lessons [13–15].

Neurofeedback creates a closed-loop system with the operator's brain using cognitive workload as the dependent variable [16]. The neural efficiency hypothesis states that individuals, based on neural capacity, discriminately respond to situational challenges and economically use mental resources, which is a measurable metric using neuroimaging techniques [17, 18]. Furthermore, an inverted U-shape relationship is manifested for performance vs level of engagement relationship that also describes higher and lower levels of workload for a task [19]. The peak of this curve at center is where optimal performance is found. As engagement decreases with lower workload, gradual disengagement causes the participant to have lower performance. Likewise, as workload increases, limited processing capacity causes performance to drop. This can also be interpreted through the concept of flow, defined as "effortless attention, reduced self-awareness, and enjoyment that typically occurs during optimal task performance." [20] flow occurs when the challenge of a task and skill level are perfectly matched, and is equivalent to peak efficiency. By adapting the level of training for each learner informed by their skill and workload, performance, and thus learning, can be optimized within an experimental design.

Measurement of brain activation with practical and reliable sensors as well as accessing the acquired data online (during a task) for processing and information extraction are essential components of neurofeedback application. Functional near-infrared spectroscopy (fNIRS) is an emerging neuroimaging technique that utilizes wearable sensors and portable hardware to monitor cortical oxygenation changes [21, 22]. The sensors can be built battery operated and even wireless, which allows participants to be untethered and mobile [10, 23, 24]. fNIRS has been successfully utilized to monitor cognitive workload in a variety of cognitive task settings related to working memory, attention, and decision making, as well as complex realistic tasks such as air traffic control and driving [10, 19, 20, 25–29].

In this study, we have integrated fNIRS-based neurofeedback with personalized adaptive difficulty adjustment in realistic piloting tasks. Our aim is to investigate if fNIRS can capture neural adaptation-related prefrontal cortex activation changes throughout the training. We also describe a new neurofeedback-based training approach using wearable and portable fNIRS and utilizing integrated behavioral and neuroimaging (i.e. neurobehavioral) data to adapt the training. Future uses of such personalized training could help prevent unnecessary over-training or insufficient under-training.

2 Methods: Multi-session Flight Simulator Adaptive Training

2.1 Participants

Four participants between the ages of 21 and 28 (2 females, mean age = 25 years) volunteered for the study. All subjects confirmed that they met the eligibility requirements of being right-handed with vision correctable to 20/20, did not have a history of brain injury or psychological disorder, and were not on medication affecting brain activity. Prior to the study all participants signed consent forms approved by the Institutional Review Board of Drexel University.

2.2 Experimental Procedure

The experiment was performed over four one-hour sessions spaced over two weeks. Sessions were held in a room isolated from outside distractions. Subjects sat at a table three feet away from a twenty-five-inch monitor, with Thrustmaster HOTAS Joystick and Throttle controls within reach of each hand. The flight simulator program used was Lockheed Martin's Prepar3D (®). Three tasks, as described below, were presented in randomized block order. Each task block contained one reference trial which remained at the same difficulty between sessions (level 1) and four training trials (dynamically adapted difficulty based on performance and neurofeedback, Fig. 5). The final session contained one less training trial as well as two additional max level transfer trials for each task (Fig. 1). Each trial took about 120 s to complete, and each task block of five to six trials averaged fifteen minutes total.

Situation Awareness Task. Subjects sit passively and view a ninety second prerecorded flight simulation based on a previous study [30]. Subjects must devote attention to the values of several gauges (from three to eight depending on level) indicating airspeed, vertical angle, and so on. After the recording, subjects are shown a buffer screen and asked to perform mental subtraction, and then are asked multiple choice questions about the values of each gauge. Performance is calculated based on the percentage of correct answers (Fig. 2).

Landing Task. Subjects control a plane with both the joystick and throttle in a procedure based on a previous study [30]. The plane begins in the air and must land on a

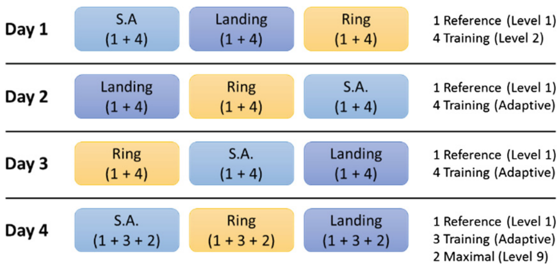


Fig. 1. Training protocol block diagram showing tasks in pseudo-randomized order and number of trials in parentheses.

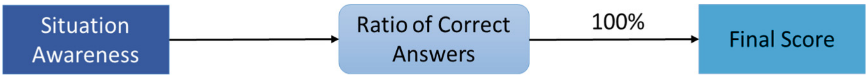


Fig. 2. Situation awareness score calculated using only answers to memory probes.

runway in front of and slightly offset to the starting position, and attempt to turn as few times as possible during approach. Early difficulties include autopilot for pitch (up/down angle) control. As level increases, control is given to the subject, the plane must land below a maximum vertical speed and/or airspeed, vision is obscured by fog or rain, and high winds cause turbulence. Performance is calculated based on ratio of successful landings, plane roll (left/right tilt) control, and pitch control (Fig. 3).

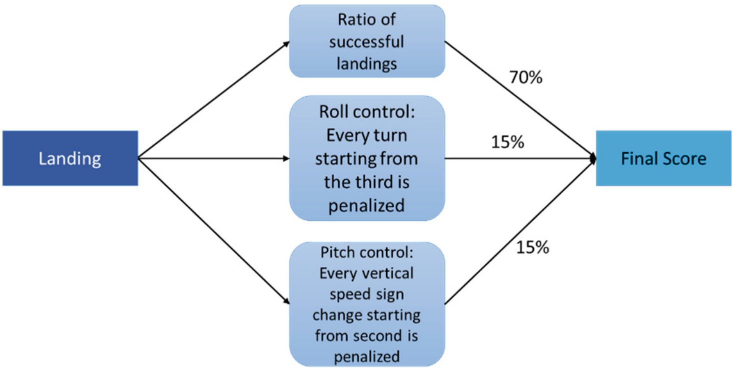


Fig. 3. Landing task score based primarily on success, and partially on steadiness of flight.

Ring Task. Subjects control a plane using only the joystick. Subjects must attempt to fly through the center of a series of seven rings floating in midair approximately twenty flight seconds apart from one another. As level increases, rings move back and forth laterally at incrementally increasing speeds. Performance is calculated based on ratio of successful ring hits, and a measure of plane control consisting of ring centroid accuracy, roll, and pitch at moment of ring hit (Fig. 4).

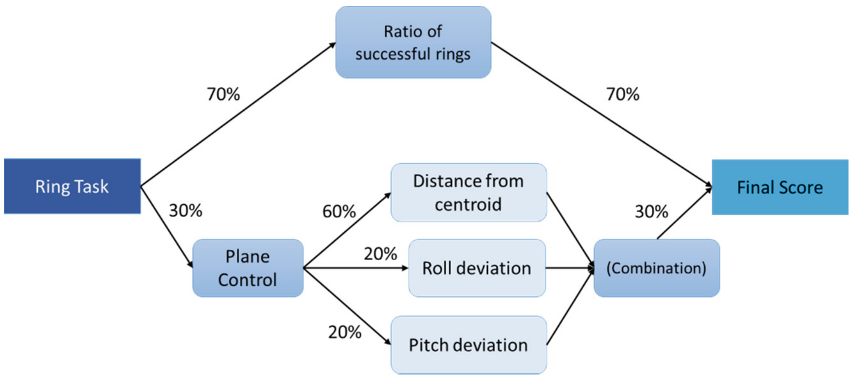


Fig. 4. Ring task score based primarily on success, and partially on ring navigation skill.

2.3 Metrics

The level for each training task over sessions two through four of the experiment were calculated based on a combination of behavioral performance and fNIRS neuroimaging measures.

Performance Analysis. A single score per trial was calculated based on the following flowcharts, and training trial scores were combined to determine overall performance within a session.

Performance data was statistically analyzed using repeated measures linear mixed models using Bonferroni correction with an alpha of 0.05. The Akaike information criterion was used to select the most parsimonious model of fixed effects.

fNIRS Acquisition and Analysis. Prefrontal cortex hemodynamics were measured using a continuous wave fNIRS system (fNIR Devices LLC, fnirdevices.com) that was described previously [25]. Light intensity at two near-infrared wavelengths of 730 and 850 nm was recorded at 2 Hz using COBI Studio software [31]. The headband system contains four LEDs and ten photodetectors for a total of sixteen optodes. Data was passed through a finite impulse response low pass filter of order 20 and cutoff frequency 0.1 Hz. Time synchronized blocks for each trial were processed with the Modified Beer-Lambert Law to calculate oxygenation for each optode. For optode numbers and corresponding locations please see Fig. 1 in [25]

Each block was subjected to simple linear regression to extract the overall slope and oxygenation change for each trial. The slopes of the four training trials within each session were then again processed using simple linear regression, and the final value was classified as either high workload, indicated by a positive slope, low workload, indicated by a negative slope, or neutral [32]. The statistical analysis was done using repeated measures linear mixed models with Bonferroni correction. The Akaike information criterion was used to select the most parsimonious model of fixed effects.

Adaptation Decision Tree. For each task training block during a session, subjects are first classified into one of four groups based on their performance score (green bubbles). Neutral classification then follows the black arrow, low workload follows the

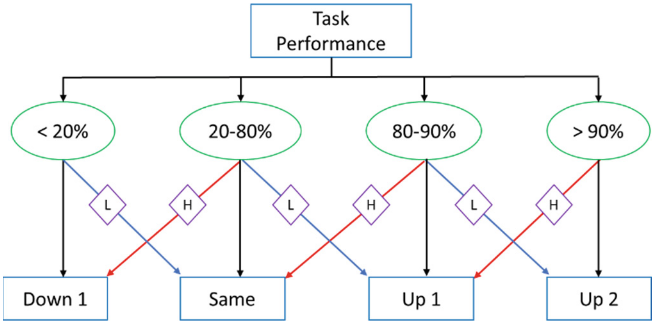


Fig. 5. Flowchart describing training progression between successive sessions.

blue arrow, and high workload follows the red arrow. The training level for the following session is decided by the blue boxes at the bottom of the flowchart.

3 Results

Each participant performed static reference trial (level 1) for each task during each of the four sessions, for a total of four trials per task overall. The average performance scores and dorsolateral prefrontal cortex oxygenation listed below for each task.

The ring task reference levels that had the same difficulty level throughout all four sessions' results are summarized in Fig. 6. The average behavioral performance had an almost significant effect with sessions ($F_{3,101.1} = 2.489$, $p < 0.065$) and showed an increasing trend over sessions. fNIRS based oxygenated-hemoglobin changes only from optode 8 resulted in a significant effect with session ($F_{3,104} = 2.908$, $p < 0.039$) and as expected showed a decreasing trend over sessions.

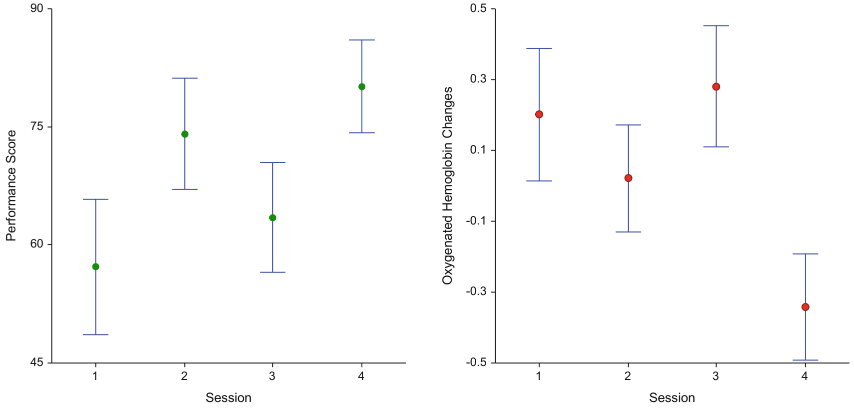


Fig. 6. Ring task behavioral performance (left) and prefrontal cortex oxygenation changes (right) (n = 4) indicate an increase in performance and decrease in task-related oxygenation changes. The whiskers are standard error of the mean (SEM).

The landing task reference levels at the same difficulty level throughout all four sessions produced similar results to the trends of ring task and are summarized in Fig. 7. The average behavioral performance did not have a significant effect with sessions although showed an increasing trend throughout the sessions There was a decrease in variability in performance for session 4 indicating enhancement of performance and is due to all subjects making successful landings. fNIRS based oxygenated-hemoglobin changes only from optode 5 resulted in a significant effect with session ($F_{3,10} = 4.242$, $p < 0.036$).

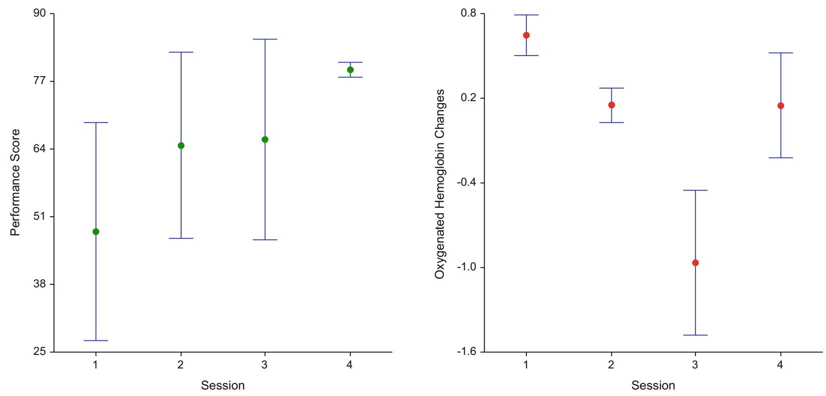


Fig. 7. Landing task behavioral performance (left) and oxygenation changes (right) indicate an overall increasing trend in performance and decreasing trend in task-related oxygenation changes. The whiskers are SEM.

The situation awareness task reference level results are summarized in Fig. 8. The average behavioral performance had no significant effect with sessions and indicated no

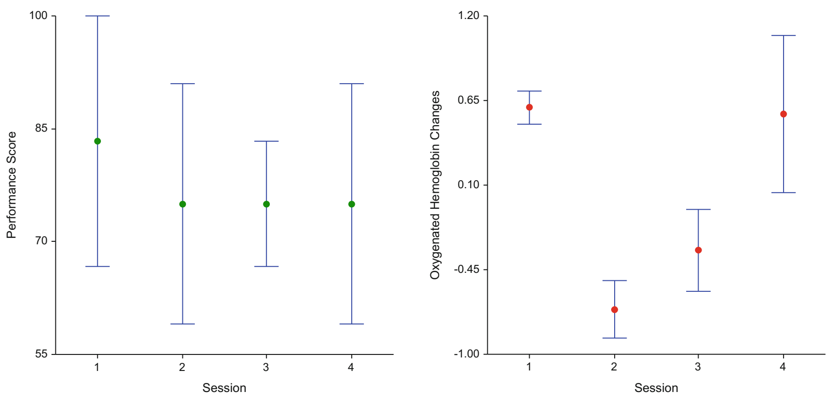


Fig. 8. Situation awareness task behavioral performance (left) and oxygenation changes (right) indicate no overall difference in performance across sessions and an increasing trend in task-related oxygenation changes. The whiskers are SEM.

change across sessions. This is partly due to the relatively easy nature of the task causing a ceiling effect for participants reaching a 100% score. However, fNIRS based oxygenated-hemoglobin changes still indicated a significant effect with session but only from optode 1 ($F_{3,10} = 5.194$, $p < 0.021$).

Finally, the adapted difficulty levels of landing task were analyzed (Fig. 9). For behavioral performance, there was a main effect of levels only ($F_{1,26.7} = 7.996$, $p < 0.01$). For fNIRS based oxygenated-hemoglobin changes at optode 5 (the same area identified with the static reference levels of the same landing task), there was a main effect of session ($F_{3,24.8} = 4.844$, $p < 0.01$) and level ($F_{1,23.1} = 7.823$, $p < 0.011$).

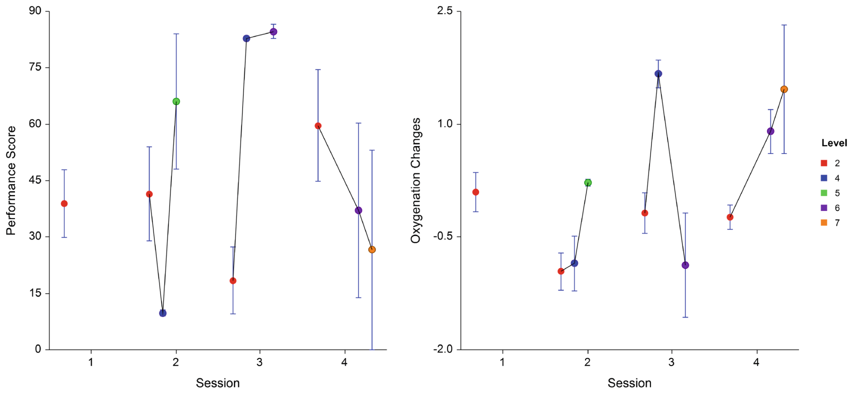


Fig. 9. Dynamic difficulty adapted landing task performance scores (left) and oxygenation changes (right) for each session with breakdown of level indicate an overall increased oxygenation with increasing task difficulty. The whiskers are SEM.

4 Discussion

In this study, we have investigated a new neurofeedback-based personalized adaptive training protocol. Participants performed piloting tasks in a low fidelity flight simulator while their anterior prefrontal cortex was monitored with fNIRS. There were three different tasks: situational awareness, landing, and ring tasks over four sessions (days). For each task, there were static (same difficulty) levels that were repeated at each session as well as dynamic (difficulty adapted) training levels.

Preliminary results of static reference levels across sessions indicated that the behavioral performance increased and anterior prefrontal cortex hemodynamic response decreased for landing and ring tasks. This result is consistent with previous literature of multi-session repetitive practice of motor and cognitive tasks [6, 11, 33] and also with similar ring and landing tasks in our previous studies [8, 25, 34, 35]. On the other hand, the situational awareness task did not yield any performance changes across sessions. This can be explained by the relatively simplistic nature of the task (multiple choice questions) and the ceiling effect that participants reached. Furthermore, in contrast to landing and ring tasks, there was no fine motor or continuous control requirements. The situational awareness task only requires sustained attention and engages working memory as during the task participants passively monitor aircraft parameters and report when asked.

Another outcome of the reference task fNIRS results is that the different brain areas (optodes) that were responsive for each task (ring, landing and situation awareness) across sessions is consistent with the different underlying neurocognitive demands of each. The situational awareness left lateral prefrontal cortex activation is consistent with earlier working memory findings. Moreover, the medial prefrontal cortex was active for ring and landing tasks as demonstrated earlier for reference tasks.

Due to the exploratory nature of the study, we used neurofeedback only from a predefined area for the subjects. Participants received adaptive training with fNIRS oxygenation measures in the left dorsolateral prefrontal cortex area, which is correlated with the development of task expertise. It is possible that there may be more optimal areas or additional factors that contribute to a change in workload which may recruit areas of the brain that were not being monitored in the current study. More information could be gained by monitoring parietal areas in addition to the frontal cortex, or by combining fNIRS with EEG. Moreover, signal processing algorithms for fNIRS may be further optimized to increase the sensitivity of dependent variables. It is possible that the hemodynamic response itself may not be sufficiently sensitive to pick up subtle trial by trial changes in workload. This is an area for future research.

In conclusion, this study demonstrated a neurofeedback-based closed loop adaptive training protocol and identified specific brain areas responsive to practice and difficulty levels. The study described here provides important albeit preliminary information about fNIRS measures of the PFC hemodynamic response and its relationship to mental workload, task training, and performance in a complex multitasking environment. Level of expertise does appear to influence the hemodynamic response in the dorsolateral/ventrolateral prefrontal cortices, at least for some complex tasks. Since fNIRS technology allows the development of miniaturized and wearable sensors, it has the potential to be utilized for continuous monitoring of trainees in future learning and training environments to personalize the training and/or to assess the mental effort of human operators in critical multitasking environments.

Acknowledgments. This investigation was in part funded by Lockheed Martin. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the funding agency.

References

1. Hancock, P.A., Parasuraman, R.: Human factors and safety in the design of intelligent vehicle-highway systems (IVHS). J. Saf. Res. **23**, 181–198 (1992)
2. Hancock, P.A., Verwey, W.B.: Fatigue, workload and adaptive driver systems. Accid. Anal. Prev. **29**, 495–506 (1997)
3. Parasuraman, R., Wilson, G.F.: Putting the brain to work: neuroergonomics past, present, and future. Hum. Factors **50**, 468–474 (2008)

4. Parasuraman, R., Christensen, J., Grafton, S.: Neuroergonomics: the brain in action and at work. Neuroimage **59**, 1–3 (2012)
5. Parasuraman, R.: Neuroergonomics. Curr. Dir. Psychol. Sci. **20**, 181–186 (2011)
6. Leff, D.R., Orihuela-Espina, F., Elwell, C.E., Athanasiou, T., Delpy, D.T., Darzi, A.W., Yang, G.-Z.: Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fNIRS) studies. NeuroImage **54**, 2922–2936 (2011)
7. Prakash, R.S., De Leon, A.A., Mourany, L., Lee, H., Voss, M.W., Boot, W.R., Basak, C., Fabiani, M., Gratton, G., Kramer, A.F.: Examining neural correlates of skill acquisition in a complex videogame training program. Front. Hum. Neurosci. **6**, 115 (2012)
8. 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**, 1–13 (2013)
9. Ayaz, H., Çakir, M.P., Izzetoglu, K., Curtin, A., Shewokis, P.A., Bunce, S.C., Onaral, B.: Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging. In: 2012 IEEE Aerospace Conference, pp. 1–11 (2012)
10. 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)
11. McKendrick, R., Ayaz, H., Olmstead, R., Parasuraman, R.: Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. NeuroImage **85**(Part 3), 1014–1026 (2014)
12. Wickens, C.D., Hutchins, S., Carolan, T., Cumming, J.: Effectiveness of part-task training and increasing-difficulty training strategies: a meta-analysis approach. Hum. Factors **55**, 461–470 (2013)
13. Yuksel, B.F., Oleson, K.B., Harrison, L., Peck, E.M., Afergan, D., Chang, R., Jacob, R.J.: Learn piano with BACh: an adaptive learning interface that adjusts task difficulty based on brain state. In: CHI 2016 (2016)
14. Ayaz, H., Willems, B., Bunce, S., Shewokis, P.A., Izzetoglu, K., Hah, S., Deshmukh, A., Onaral, B.: Estimation of cognitive workload during simulated air traffic control using optical brain imaging sensors. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) Foundations of Augmented Cognition. Proceedings of the 6th International Conference on Directing the Future of Adaptive Systems, FAC 2011, Held as Part of HCI International 2011, Orlando, FL, USA, 9–14 July 2011, pp. 549–558. Springer, Heidelberg (2011)
15. Solovey, E.T., Zec, M., Garcia Perez, E.A., Reimer, B., Mehler, B.: Classifying driver workload using physiological and driving performance data: two field studies. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 4057–4066. ACM (2014)
16. Shewokis, P., Ayaz, H., Izzetoglu, M., Bunce, S., Gentili, R., Sela, I., Izzetoglu, K., Onaral, B.: Brain in the loop: assessing learning using fNIR in cognitive and motor tasks. In: Schmorrow, D., Fidopiastis, C. (eds.) Foundations of Augmented Cognition. Directing the Future of Adaptive Systems, vol. 6780, pp. 240-249. Springer, Heidelberg (2011)
17. Rypma, B., Berger, J.S., Prabhakaran, V., Martin Bly, B., Kimberg, D.Y., Biswal, B.B., D'Esposito, M.: Neural correlates of cognitive efficiency. NeuroImage **33**, 969–979 (2006)
18. Neubauer, A.C., Fink, A.: Intelligence and neural efficiency. Neurosci. Biobehav. Rev. **33**, 1004–1023 (2009)

19. Sibi, S., Ayaz, H., Kuhns, D.P., Sirkin, D.M., Ju, W.: Monitoring driver cognitive load using functional near infrared spectroscopy in partially autonomous cars. In: 2016 IEEE Intelligent Vehicles Symposium (IV), pp. 419–425 (2016)
20. Afergan, D., Peck, E.M., Solovey, E.T., Jenkins, A., Hincks, S.W., Brown, E.T., Chang, R., Jacob, R.J.K.: Dynamic difficulty using brain metrics of workload. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, Toronto, Ontario, Canada, pp. 3797–3806. ACM (2014)
21. Bunce, S.C., Izzetoglu, K., Ayaz, H., Shewokis, P., Izzetoglu, M., Pourrezaei, K., Onaral, B.: Implementation of fNIRS for monitoring levels of expertise and mental workload. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) Foundations of Augmented Cognition. Proceedings of the 6th International Conference on Directing the Future of Adaptive Systems, FAC 2011, Held as Part of HCI International 2011, Orlando, FL, USA, 9–14 July 2011, pp. 13–22. Springer, Heidelberg (2011)
22. Ayaz, H., Shewokis, P.A., Bunce, S., Onaral, B.: An optical brain computer interface for environmental control. In: 2011 IEEE Proceedings of the Conference on Engineering in Medicine and Biology Society, pp. 6327–6330 (2011)
23. Mckendrick, R., Parasuraman, R., Murtza, R., Formwalt, A., Baccus, W., Paczynski, M., Ayaz, H.: Into the wild: neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. Front. Hum. Neurosci. **10**, 216 (2016). doi:10.3389/fnhum.2016.00216
24. McKendrick, R., Mehta, R., Ayaz, H., Scheldrup, M., Parasuraman, R.: Prefrontal hemodynamics of physical activity and environmental complexity during cognitive work. Hum. Factors **59**, 147–162 (2017)
25. 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**, 36–47 (2012)
26. Durantin, G., Dehais, F., Delorme, A.: Characterization of mind wandering using fNIRS. Front. Syst. Neurosci. **9**, 45 (2015)
27. Gateau, T., Durantin, G., Lancelot, F., Scannella, S., Dehais, F.: Real-time state estimation in a flight simulator using fNIRS. PLoS One **10**, e0121279 (2015)
28. Mandrick, K., Chua, Z., Causse, M., Perrey, S., Dehais, F.: Why a comprehensive understanding of mental workload through the measurement of neurovascular coupling is a key issue for neuroergonomics? Front. Hum. Neurosci. **10**, 250 (2016)
29. Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., Schultz, T.: Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. Front. Hum. Neurosci. **7**, 935 (2013)
30. Choe, J., Coffman, B.A., Bergstedt, D.T., Ziegler, M., Phillips, M.E.: Transcranial direct current stimulation modulates neuronal activity and learning in pilot training. Front. Hum. Neurosci. **10**, 34 (2016). doi:10.3389/fnhum.2016.00034
31. 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
32. Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., Perrey, S.: Utilizing slope method as an alternative data analysis for functional near-infrared spectroscopy-derived cerebral hemodynamic responses. Int. J. Ind. Ergon. **43**, 335–341 (2013)
33. Yumie, O., Jack Adam, N., Xian, Z., Yasunori, N., Tatsuya, S., Sotaro, S., Atsumichi, T., Shaw, B., Joy, H.: Motor learning and modulation of prefrontal cortex: an fNIRS assessment. J. Neural Eng. **12**, 066004 (2015)

34. Hernandez-Meza, G., Slason, L., Ayaz, H., Craven, P., Oden, K., Izzetoglu, K.: Investigation of functional near infrared spectroscopy in evaluation of pilot expertise acquisition. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) Foundations of Augmented Cognition, vol. 9183, pp. 232–243. Springer, Heidelberg (2015)
35. Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Woo, H., Shewokis, P.A., Bunce, S.C., Onaral, B.: Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. IEEE Trans. Hum.-Mach. Syst. **44**, 429–440 (2014)

SHELL Revisited: Cognitive Loading and Effects of Digitized Flight Deck Automation

Mark Miller and Sam Holley(✉)

College of Aeronautics, Embry-Riddle Aeronautical University Worldwide,
Daytona Beach, FL, USA
sam.holley@erau.edu

Abstract. Acknowledging the SHELL human factors model, authors examine interfaces among components and assess problems created when the model is aligned with modern digitized flight deck systems. Complacency and over-reliance on automated systems are evaluated, and cognitive load and potential for degraded situational awareness are examined. Authors present a SHELL overlay demonstrating where particular digitized functions and operations present challenges to operators and markedly influence effective SHELL interactions in highly complex flight deck systems. Human factors contributing to the Asiana Flight 214 accident are examined and correlates identified with the SHELL analysis. Implications for advanced crew resource management are presented, and human centered system training applications are proposed for addressing the workload challenges. Implications for working and prospective memory functions are examined, along with accompanying biases. Potential for adaptive automation technology concludes the SHELL overlay analysis with potential for reducing cognitive overload in the digitized flight deck environment.

Keywords: SHELL · Digitized flight deck · Human factors · Working memory · Crew resource management

1 Introduction

The technological world we live in is currently issuing the next generation of computer power and the aviation industry has been taking full advantage of both computer automation and information in the cockpit in the new millennia. With all these marvelous technologies working together, the industry has improved greatly in terms of safety and efficiency. Pilots can do more in the cockpit with less man power and companies can save money in terms of man power and training. However, as the industry has made giant strides with computer technology, one accident like the 2013 Asiana Airlines Flight 214 crash in San Francisco can be a sobering reminder that even with the best equipment and computers at hand (a Boeing 777), human error can still win the day. Accidents that are investigated by the National Transportation Safety Board (NTSB) like the Asiana Flight 214 crash are thoroughly processed using a meticulous scientific method of elimination to determine safety causation. The vast

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_9

majority of the time the NTSB has no problem narrowing down the culprit to a Probable Cause and then supports that Probable Cause with many human factors causations that contributed to that accident. In the case of the Asiana Flight 214 crash, the NTSB could not come up with one Probable Cause, but instead drew on several different opinions for Probable Causes from a slew of human factors computer automation errors that all contributed to the accident. Why the NTSB had such a difficult time in addressing the Probable Cause of this accident can be analyzed through an older tool of human factors analysis called the SHELL. The SHELL model was developed into a building block structure by Hawkens in 1984 [1] and has been in use since the 1980's to analyze human factors and human error in aviation. By using an updated model of the SHELL in 2017, new light can be shed into the effects of the many editions of the computer on human factors in the cockpit and the many cognitive issues caused by them.

2 The SHELL Model Revisited

The SHELL model of human factors analysis that Hawkins presented had the simple block layout of centering the human as represented by the Liveware (L) in the center and surrounding it by four other human factors interfaces: Software (S), Hardware (H), Environment (E) and Liveware (L) as in Fig. 1.

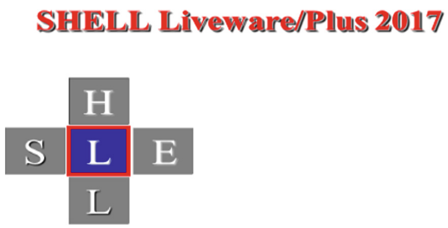


Fig. 1. SHELL model updated for 2017 with Liveware/Plus highlighted [1]

There was a time when aircraft had very few, if any, computers in the cockpit for automation and information usage to the pilots represented at the center of the SHELL, but over the last 30 years computers related to automation and information have slowly started to crowd the SHELL model's interfaces as indicated by the red surrounding the center Liveware (L) in Fig. 1. Surprisingly, it is not just the Hardware (H)-Liveware (L) interface where computers have invaded in the form of flight controls to disrupt that direct flow in the ergonomic design of the man-machine interface. In 2017, the computer is now involved in all Liveware interfaces. Multiple computers are now used in all the SHELL interfaces to include the new era of the Electronic Flight Bag (EFB) which has been introduced to aviation cockpits in the last decade. A closer examination of each interface will demonstrate how prolific the computer has become in the cockpit.

2.1 The SHELL Model 2017 and the (L)-(H) Interface

The Liveware (L)-Hardware (H) interface shown in Fig. 1. represents all the physical elements of the aircraft and the system including such things as: the wing of the aircraft, the control surfaces along with the entire hydraulic systems, the flight controls in the cockpit. Every part of an aircraft physically falls into this category, but none is a more direct connection than the ergonomic man – machine interface found in the cockpit. It is there where the crew not only utilizes the flight controls, but also continually assesses data in the form of displays to manipulate those controls. The computer has been integrated in the form of automation in the flight controls and the fuel systems to tightly manage the aircrafts control inputs and flight envelope. At the same time the pilots observe the digitalized computer flight information to insure the computer is maintaining the appropriate parameters. This automation and information has formed a well-managed barrier in the Liveware (L)-Hardware (H) interface that is depicted in Fig. 2.

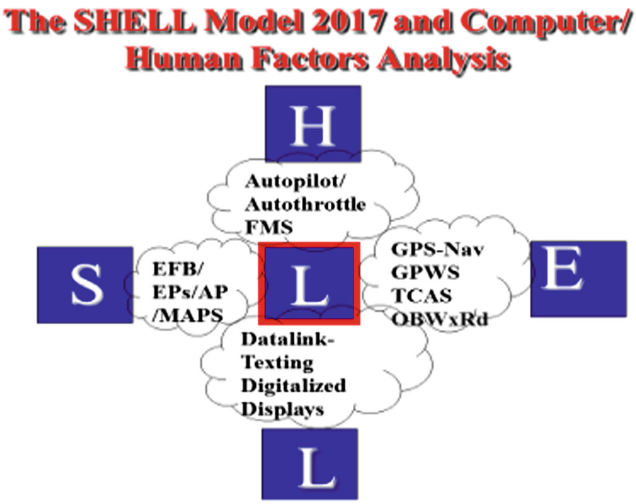


Fig. 2. The SHELL model 2017 with computer automation and information displayed by Miller in 2016 [2]

This computer barrier insists that the human work through it to manage the aircraft flight controls indirectly. By doing so the computer is now doing most of the flying thus eliminating a large amount of the direct human interaction. The ergonomic idea behind this is to limit the human's direct control of the aircraft as much as possible to prevent human error and inefficiency. The computer maintains the safe flight envelope and flies the best profile to conserve the most fuel. Here is where the flight management system and auto-throttle join the basic flight computer configuration in the increasing of the controllability of the aircraft. The use of fly by wire controls also has the potential to eliminate the pilot's direct feel for what the aircraft controls are doing and forces the pilot to manage the flight controls by visual reference only. Doing so can make the new

Liveware (L)-Hardware(H) flight control interface optical only. This is flight in the new ergonomically designed human-computer interaction loop by using the optical channel and calls for the pilot to manage the computer to fly the aircraft. The computer is now dominating this interface and it certainly has the potential of flying the aircraft by itself as demonstrated by the many Unmanned Aerial Vehicles (UAV) flying today.

2.2 The (L)-(E) Interface

The second interface to be dominated by the computer historically is that of the Liveware (L)-Environment (E) as depicted in Fig. 1. This interface is all about man's relationship to the different types of environment of the flight. This environment includes both inside the aircraft and outside the aircraft and is multifaceted by including such things as: the temperature, cabin pressure, day time, night time, and the weather. The huge growth of the commercial industry internationally after World War II was greatly aided by advent of the Jet age which boasted the safer, efficient and more reliable jet engine, but shifted the causation of most accidents to be human error related. Many of these human errors have been found to be related to the Liveware (L)-Environment (E) interface in four forms: flying the aircraft into terrain, flying the aircraft into another aircraft, flying the aircraft in bad weather and loss of control. To counteract the high numbers of accidents and fatalities, the industry became committed to reducing them by enhancing the Liveware (L)-Environment (E) interface with safety computer technologies in the cockpit in a campaign to reduce human error. This started with the Ground Proximity Warning System (GPWS) to prevent aircraft from Controlled Flight into Terrain (CFIT). With the skies becoming more crowded, the next safety device added was designed to prevent other aircraft from flying into one another and was called Traffic Collision Avoidance System (TCAS). With the all-weather flight capabilities in aircraft and the demand to fly in all types of weather, the industry eventually adopted Airborne Weather Radar to increase safety from the common hazards in the aviation weather environment. Stick shakers were added for Stall Warning. GPWS, TCAS, Airborne Weather Radar and Stall Warning are all tremendous safety devices and they have greatly enhanced aviation safety globally over the last 30 years. At the same time, they are all taking up room in the Liveware (L)-Environment (E) interface as computer generated emergency information to avoid hazards. Adding these with other important computers like the Inertial Navigation System that allow aircraft to navigate from point A to point B in the environment with enhanced computer navigation equipment and the Liveware (L)-Environment (E) interface becomes even more crowded as depicted in Fig. 2. Future Air Traffic Control Systems computers will soon be added to the cockpit in the form of Automatic Dependent Surveillance – Broadcast (ADS-B).

2.3 The (L)-(S) Interface

With the Liveware (L)-Hardware (H) interface and the Liveware (L)-Environment (E) interface filled with computers it would seem to be difficult to add more computer power to the Liveware (L)-Software (S) interface. When looking at the original Liveware (L)-Software (S) interface as depicted in Fig. 1, its main purpose was to cover everything that is non-physical used in the aviation system. This is a wide range of materials which includes such things as: procedures, maps, publications, documents, checklists and approach plates. These materials called for a hefty storage space in pilots' flight bags that they would carry with them in the cockpit. The invention of the I-pad would quickly change that allowing everything in those weighty bags to be carried on board electronically by the pilot in what is now called the Electronic Flight Bag (EFB). The Liveware (L)-Software (S) interface was no stranger to computer information, as most modern commercial cockpits produce plenty of computerized information that is digitized in nice LCD screens. However, in the new age of the EFB, all other important flight information that pilots had to carry with them in paper form is now adding more computer information to the Liveware (L)-Software (S) interface as depicted in Fig. 2. This is not only affecting the commercial industry as the EFB is a tremendous step forward for General Aviation (GA) pilots as they are loading everything they can with Jeppesen software into their own EFB's. The advent of pad and smart phone computer technologies taking the form of the EFB have also coincided with the rise of another technology that will offer up what may be the biggest change to come in the SHELL 2017 interfaces related to computers. Surprisingly, this one will be played out in the Liveware (L)–Liveware(L) interface.

2.4 The (L)-(L) Interface

The Liveware (L)-Liveware(L) interface accounts for the human interactions that occur in the flight. This can be between the cockpit crew and air traffic control or the cockpit crew and the flight attendants, but most often it focuses in on the heart of flying the aircraft and how the actual flying crew is interacting with one another. The industry has made giant strides in moving from a cockpit paradigm where the Captain was considered god in the cockpit (used up until the 1980s) to a paradigm that focuses on teamwork that is now called Advanced Crew Resource Management (ACRM). The new model focuses on tenants that include: communication, assertiveness, management, task delegation, teamwork, leadership/followership and decision making [3]. Although the computer related safety and efficiency enhancements have had a profound impact on human factors in the cockpit over the last 37 years; it is the non-computer human factors safety program of ACRM that has also had a strong impact to safety and efficiency through the Liveware (L)-Liveware(L) interface. Through this method of teamwork in the cockpit, many of the potential human errors in the Liveware (L)-Liveware(L) interface are being dealt with. However, the strong potential usage of EFB's and ADS-B digitized communication through the computer brings with it a whole new realm of communicating in the cockpit. Instead of communicating with radio, the computer technologies that are now in the cockpit offer a substitute from the radio called digitalized messaging. In terms of NextGen Air Traffic Control technologies this is called "Datalink". This is essentially the same as texting, but is now accomplished from the cockpit. Aircrew are now able to communicate to others outside the cockpit through digital messaging on the computer. Although the industry is at the beginning of implementing this new form of communication, it will be only a matter of time before this new way of communicating through the computer impacts the industry and ACRM. Certainly texting has changed the way people communicate on the ground profoundly. At the same time, it also surprised society by how dangerous it can be if done while driving. How big this computerized communication will become in aviation is something that only time will tell, however its existence and legitimacy as planned in NextGen make it very real. With digitalized communications affecting the Liveware (L)-Liveware (L) interface and the EFB transforming the Liveware (L)-Software (S) interface, the SHELL model has been completely transformed in all interfaces in 2017.

3 The SHELL Model 2017 and Asiana Flight 214

The repercussions of this new SHELL model to aviation human factors and human error have been well documented in aviation incidents and accidents over the last two decades and the San Francisco Asiana Flight 214 crash is a great example. The new 2017 SHELL Model in Fig. 2 depicts computer automation and information in all four interfaces of the SHELL and these computer innovations have been responsible for making aviation extremely safe and efficient in the last 20 years. The downside to these technological enhancements are those rare cases where computer/human factors causations line up causing human error from those SHELL interfaces and cause an incident or a rare accident. Some of the most common causes of computer-Liveware (L) error seen the most by the industry are: (1) Complacency in relying on computers, (2) Not understanding the computers, (3) Overly Focused on a computer and distracted from flying, and (4) Optical inside only with little outward scanning. In the Asiana Flight 214 crash, the pilots became complacent in the Liveware (L)-Hardware (H) and relied on their computer system to do the landings instead of being proficient at manual landings while using the computer and auto-throttles to help improve the manual landings [4]. They fell short in the Liveware (L)-Liveware (L) interface where they needed to be able to cross check normal landing checkpoints of altitude and airspeed to correct sooner when too low on glideslope and communicate this through good ACRM. In the Liveware (L) – Environment (E) interface they ignored the GPWS too long as they needed to be able to manually do a missed approach sooner in the landing. In the Liveware (L)-Hardware (H) interface the pilots also had a difficult time understanding the idiosyncrasies of the auto throttle system. By using proper ACRM training techniques like task Management and assertiveness, other members of the crew could have participated and even challenged the wrong auto throttle setting and possibly prevented the accident. In the Liveware (L)-Hardware (H) interface, when the computer technology is not working, the crew must be able to confirm the malfunction and assertively take the proper course of action. Lastly, from the Liveware (L)-Software (S) interface, the Asiana crew had become so reliant on the computer systems to land the aircraft safely that they focused more on their computers inside the cockpit then on basic landing parameters; the most important being the runway glide slope monitor telling them to execute a missed approach sooner. To avoid these errors with computers, pilots still need to aviate, navigate and communicate while scanning outside to use the runway and inside while still flying with the computer automation and information. They need to task delegate with the computers by being more human centered and managing the computer systems to fly the aircraft instead of being computer centered and letting the computer fly them.

4 The SHELL 2017 and the Potential Human Factors Cognition Issue

The NTSB recognized many different factors as the Probable cause in this Asiana Flight 214 accident. Usually one or two Probable Causes covers the main cause of NTSB Aviation Accident investigative reports, but the case of Asiana Flight 214 was different. The probable cause of the Asiana Flight 214 accident was determined by the NTSB to be the flight crew's mismanagement of the airplane's descent during the visual approach, the flying pilot's unintended deactivation of automatic airspeed control, the flight crew's inadequate monitoring of the airspeed, and the flight crew's delayed execution of a go-around after they became aware that the airplane was below acceptable glide path and airspeed tolerances [4]. These NTSB Probable Causes are all valid, but when looking at them from the SHELL model in 2017 the underlying cause of human factors cognition also needs to be addressed.

4.1 Cognitive Processing

The added interactions among SHELL 2017 components are largely cognitive. This involves considerably more continuous information processing, which suggests the potential for overload at some point. As stated earlier, Liveware activity in the SHELL framework is changing to readouts and text. With this evolution, and the emphasis on optical tasks, it is imperative to keep the pilot scan of the displays and instruments in continual engagement. As shown in Fig. 2, what may have been embedded and non-visualized software now requires more focused pilot attention, optical processing, and cognitive processing to understand the plethora of dynamic information presented. When pilots must look up and through the windscreen, the optical flow changes to actual external cues and associated cognitive processing to comprehend what is happening and the relevance to flight parameters. To some degree, use of head-up display symbology is similar to the screens used, and is consistent with the cognitive processing underway. However, real-world visual cues, particularly in non-standard events quickly overloads the pilot. Synthetic vision can introduce real world and virtual representations in the same display, which will consume far more neural resources in combination. The cognition difficulties inherent in the SHELL 2017 interfaces are both combinatorial and concatenated. These might best be illustrated with Gestalt perception principles [5]. For example, the combinatorial effects are seen with proximity (SHELL clouds approach overlapping), closure (premature conclusions), and primacy (initial optical indication). Concatenation effects are evident in the principle of continuation (current trajectory will continue). Cognitive processing clouds represented in the SHELL 2017 illustrations show these principles and their role in cognitive loading and potential for cognitive error. The SHELL process is essentially linear, and the original model did not initially contemplate simultaneous, multi-dimensional interfaces and interactions among the SHELL components. Consequently, as the optical and cognitive loads have increased with digitization, a concatenation effect can occur where information from one SHELL domain attaches to a second domain, and continues to build until the pilot either becomes confused or may reach a faulty conclusion or determine incorrectly a need for action. This can lead to one of the three forms of cognitive error [6]: faulty synthesis involving flawed processing of available information (with premature closure the most common element), faulty knowledge, and faulty data gathering. To assess the significant involvement of cognitive processing represented in the SHELL 2017 architecture, neural functions are key to understanding the safety implications. Prefrontal cortex connects in some fashion with every level of activity in the brain. In this regard, long-term memories distributed throughout inter-connected neural pathways are subject to disruption or rerouting as a result of excess beta wave activity. If a pilot becomes anxious when overcome with information from the cognitive clouds seen in SHELL 2017, this type of disruption occurs. An associated transmitter, acetylcholine, is highly involved in attentional processes, alpha wave production, and is influential in neural processing flow. When alpha wave activity is deficient, the brain lacks sufficient mental speed to connect perceptions and thoughts [7]. Thus, situations that require escalating optical activities, as SHELL 2017 has illustrated, add to the load placed on the prefrontal areas.

4.2 Cognitive Flow

When considering the varying functions and intentions of each SHELL component, the combined effects require synchronous neural processing. To achieve this process, information and associated neural actions must flow freely. This flow, however, is easily disrupted and affects intra-neural communication creating an imbalance in the brain's systems [8]. Flow relies on sustained attention to processing demands and a mental sequencing, or map, of what is required to complete actions with rapid updating of working memory [9]. Bowers et al. [10] demonstrated a strikingly different pattern associated with moving between tasks that would indicate post-workload transition effects might manifest with onset of a high level of workload. This is at the heart of the SHELL process. This is precisely the situation when the pilots must integrate the cognitive content and synthesize flow among the various SHELL components. A workload transition, such as shifting within the SHELL 2017 domains, constitutes a shift in cognitive task difficulty and could likely result in an increase in missed events or significant flight dimensions as described by the hysteresis effect which has been attributed to cognitive resource depletion [11]. Among perceptual tasks with SHELL 2017 is constructing a cognitive map of the operating environment and interacting influences. Recent findings [12] verified that neurons in the brain related to space-mapping react to virtual environments differently from the real world. The hippocampus is recruited when a person develops a cognitive map of the environment, including calculation of distances and space, and is further mediated through the post rhinal and entorhinal cortex. In virtual environments, results showed that as much as half the hippocampal neurons usually involved were actually shut down and the cognitive map was nonexistent. The implications for pilots are profound. This suggests a different region of the brain is involved in the spatial learning tasks and processes, compared with reduced digitalization, and is complicated when perceptual variances become intertwined (one using virtual cues and the other real-world cues). The implications for SHELL 2017 are that the cognitive areas encroaching the component boundaries are increasingly subject to perceptual disparities that consume evermore resources and deplete neural energy rapidly [13]. Combinations of virtual and real world perceptual input among the SHELL 2017 components acts directly upon this hippocampal processing with attendant reductions in comprehension by pilots. Taken together, the optical processing demands, coupled with cognitive processing functions, place a notable load upon the pilot. When combined perceptions and multidimensional interfaces are added, there is a likely potential for cognitive disruption. What becomes sacrificed, then, are cognitive maps and continuity of processing. These introduce the phenomenon of cognitive loading, which is examined next.

4.3 Cognitive Loading

The concept of cognitive loading was introduced in 1988 and developed further by Chandler and Sweller [14]. As the term came into use for attention and memory applications, references to information processing became prominent with particular emphasis on perception, memory, and reasoning. The amygdalae are constantly scanning incoming information to determine potentials for safety risks. Increased scanning activity, as with the SHELL 2017 cognitive tasks, has the potential to elevate conscious awareness of perceived threats. As is well known, the amygdalae can dominate the neocortex and obstruct clear thinking when most needed [15]. Related neural function is contained in the concept of continuous partial attention which is a compromised state of focus due to attending many information streams at the same time. This is directly indicated in the growing cognitive task clouds of SHELL 2017, and can lead to infoglut from continuously increasing information contexts. The concept of working memory developed by Baddeley [16] has proven robust and is widely accepted as a useful model for understanding how brief memory operates. As he described it, working memory is the cognitive process that allows moment-to-moment perceptions across time to be integrated, rehearsed, and combined with simultaneous access to stored information about previous experience, actions, or knowledge. As for the encoding aspect of the arriving and cycling information, proteins in the prefrontal cortex are essential to keeping optical tasks and neural processing active. As additional messages are added, the replenishment of synthesized protein is inhibited or diminished, thereby causing loss of some of the cycling elements active in working memory [17], which happens when strings of working memory connected with the SHELL 2017 tasks are compromised and no longer intact. Drawing upon the model developed by Baddeley for working memory, and including recent findings for bandwidth issues and protein cycling limits, it is apparent that pilots operating in the realm of SHELL 2017 cognitive loading can reach saturation of working memory buffering. Pilots in a digitized cockpit often are working with multiple screens and monitors. As would be expected, the visual component of perception is subject to saturation from stimuli and data. Accompanying visual input is the need to interpret the significance or urgency of the information. Perceiving what is critical, what is evolving, and sequences for actions becomes paramount. An element associated with reducing stress is the ability to resolve conflicts as they arise. A resulting uncertainty with regard to conflict resolution can occur between two states such as GPWS alert (L-E interface) and EFB map (L-S interface). While cognitive memory is generally mediated in the anterior hippocampi, affective memory is processed via the amygdalae. Often, amygdala-driven memories can take precedence in neural sequencing. For some pilots, amygdala-level situational appraisal may invite distorted pattern recognition and proneness to false alarms [18]. When considering automation influences between open-loop and closed-loop interactions, issues related to SHELL 2017 cognitive loading become evident. Evidence reported in a study [19] of nearly 2,000 pilots flying aircraft with advanced automation systems showed pilots described differences in "flying through the computer, which required more self-discipline, lags in anticipating aircraft behavior, and increased monitoring of mode annunciations. Each of these conditions increases the vigilance pilots direct to the cognitive tasks in the expanded SHELL 2017 environment. In our examination of SHELL 2017 effects, information being encoded from multiple related sources in an interactive interface can present a delta gap effect which delays correctly understanding the implications of the information received. As the cycling working memory elements from several SHELL domains are attempting to cross-communicate among various neural processing centers, the lag in neural responding and accurate information integration can be profound [20]. Consequently, when cross-neuralization is challenged (as in combined SHELL 2017 cognitive activities) – a delta gap may apply, with signal lag among brain components responding to inquiries from other brain sending locations. Potential for confusion, slowed comprehension, and related cognitive functions increases accordingly. With cognitive shifts from closed loop to open loop processing into the SHELL 2017 architecture, the pilot is likely to be cycling two or more scenarios in working memory, with rehearsal and encoding challenges continually involved to comprehend flight maneuvers and anticipated next steps. Where level of automation is not a significant factor, what has become evident is that although mental resources are not always completely expended for primary tasks, the presence of competing demands from secondary task components can strain the primary functions to the point of saturation [10]. It has been demonstrated that working memory deficits involving increased dopamine levels caused by stress can occur [9]. The obvious implications for SHELL 2017 are that, should pilots become confused or anxious, working memory deficits could manifest. Likewise, calcium sensitive kinases, calcineurin, and dephosphorylation have been shown as detrimental to working memory and are in evidence during high visual task loadings [21]. When neural processing capacities are exceeded, with particular reference to the digitized cockpit environment, degraded situational awareness and various cognitive errors may ensue. Further, this disrupts the prospective memory and remaining ahead of related incoming information from other components of the SHELL interactive process.

4.4 Situational Awareness

Research on mental workload, and especially overload, has focused on situational awareness, information processing, and decision making where they are simultaneously present. When too high or too low, cognitive load increases risk of error, more notably when abrupt bursts of a large amount of information must be processed quickly [22]. This would likely occur, for instance, during unanticipated events and rapidly changing information flow in the digitized cockpit environment. Correspondingly, the outcome manifests as a cognitive loading challenge. The researchers found that the nature of a non-linear task environment, like that with SHELL 2017, stimulated operator concerns about future states of the system. The SHELL 2017 domain has advanced the optical interactions significantly. Accompanying the optical loading, interaction effects can be expected. Endsley [23] describes situational awareness (SA) as a working memory bottleneck for pilots in novel situations. For more experienced pilots with skilled performance capabilities, increased recruitment of long term memory augments the SA process and results in fewer gaps or performance decrement. Consequently, the volume of mental processing to sustain high levels of SA require pilot access to embedded mental constructs in long term memory. Such increased and sustained activity would clearly result in a more rapid consumption rate of available brain glucose necessary for effective functioning. Protein that is fueling working and prospective memory is depleted rapidly [21]. A study of pilot decision making [24] affirmed that in low tempo operations extra cognitive resources are available, however, when uncommon or emergency events occur pilot time for reflection is substantially reduced. Consequences result in the need for understanding that the changed situation has compromised the system status. The researchers identified several cognitive breakdowns that occurred: delay in comprehending an event was occurring, fragmented scan of information sources, narrowed assessment, inability to commit to a course of action, and failure to re-check new courses of action to assure implementation as intended. As digital cockpit pilots become loaded near maximum working memory capacity, during especially challenging flight maneuvers or unanticipated procedures, deferring critical actions could be catastrophic. As the SHELL 2017 model illustrates, multiple, simultaneous optical processing tasks must be attended to and integrated with current flight parameters and aircraft systems. When exceeded, neural capacities become strained and, along with incipient cognitive error, mode confusion can result. This has a history, as highlighted in the Australian study of pilots [25] which also confirmed that workarounds highjacked cognitive resources. As optical processing demands have increased, many pilots view the flight management computer as more competent than themselves.

5 Conclusions

In summary, the SHELL 2017 representation reflects the combined influences of optical processing demands on pilots, and the related cognitive loading challenges associated with the proliferation of similar cognitive tasks within each SHELL domain. Efforts to contain or reduce cognitive error suggest better training for pilots so they are aware of potential environments that contribute to error. Other efforts are more within the adaptive automation spectrum to recognize potential error and adjust flight computations accordingly. Further assessment of the similarities and differences of cognitive demands within each SHELL domain is indicated to more thoroughly understand how to gain efficiencies and to reduce potential for unintended errors based on cognitive misperceptions.

References

1. Hawkins, F.H.: Human Factors in Flight, 2nd edn. Ashgate, Aldershot (1987)
2. Miller, M.D.: Aviation human factors: the SHELL model 2017 and computer/ human factors analysis. In: Presentation to FAA Aviation Safety Conference, Honolulu (2016)
3. Seamster, T., Boehm-Davis, D., Holt, R., Shulz, K.: Developing advanced crew resource management (ACRM) raining: a training manual. Federal Aviation Administration Office of Chief Scientific and Technical Advisor for Human Factors, AAR-100 (1998)
4. National Transportation Safety Board.: Descent Below Visual Glide Path and Impact with Seawall Asiana Flight 214 NTSB/AAR-14/01 (2013)
5. Wagemans, J., Elder, J.H., Kubovy, M., Plamer, S.E.: A century of gestalt psychology in visual perception. Psy. Bull. **138**(6), 1172–1217 (2012)
6. Dunbar, M., Helms, S.E., Brodell, R.T.: Reducing cognitive errors in dermatology: can anything be done? J Am. Acad. Dermatol. **69**(5), 810 (2013)
7. Blokland, A.: Acetylcholine: a neurotransmitter for learning and memory? Brain Res. Rev. **21**, 285–300 (1996)
8. Lind-Kyle, P.: Heal Your Mind: Rewire Your Brain. Energy Psychology Press, Santa Rosa (2010)
9. Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., Rush, G.: Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum. Factors **40**(1), 79–91 (1998)
10. Bowers, M.A., Christensen, J.C., Eggemeier, F.T.: The effects of workload transitions in a multitasking environment. In: Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, pp. 220–224 (2014)
11. Johns, M., Inzlicht, M., Schmader, T.: Stereotype threat and executive resource depletion. J. Exp. Psychol.: Gen. **137**(4), 691–705 (2008)
12. Aghajan Z., Acharya, L., Moore, J., Cushman, J., Vuong, C., and Mehta, M.: Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. Nat. Neurosci. (2014)
13. Ungar, M.: The social ecology of resilience: addressing contextual and cultural ambiguity of a nascent construct. Am. J. Orthopsychiatry **81**(1), 1–17 (2011)
14. Chandler, P., Sweller, J.: Cognitive load theory and the format of instruction. Cogn. Instr. **8** (4), 293–332 (1991)
15. Woods, D.D., Patterson, E.S., Roth, E.M.: Can we ever escape from data overload? A cognitive systems diagnosis. Cogn. Technol. Work **4**(1), 22–36 (2001)
16. Baddeley, A.D.: Working Memory, Thought and Action. Oxford Univ. Press, Oxford (2007)
17. Cottrell, J., Levenson, J., Kim, S., Gibson, H., Richardson, K., Sivula, M., Li, B., Ashford, C., Heindl, K., Babcock, R., Rose, D., Hempel, C., Wiig, K., Laeng, P., Levin, M., Ryan, T., Gerber, D.: Working memory impairment in calcineurin knock-out mice is associated with alterations in synaptic vesicle cycling and disruption of high-frequency synaptic and network activity in prefrontal cortex. Science **33**(27), 10938–10949 (2013)

18. Arden, J.: Rewire Your Brain. Wiley, Hoboken (2010)
19. Rudisill, M.: Line Pilots' attitudes about and experience with flight deck automation. In: Proceedings of 8th International Symposium on Aviation Psychology. Ohio State University Press, Columbus (1995)
20. Clough, B.: Unmanned aerial vehicles: autonomous control challenges, a researcher's perspective. J Aerosp. Comput. Inf. Commun. **2**, 327–347 (2005)
21. Dash, P.K., Moore, A.N., Kobori, N., Runyan, J.D.: Molecular activity underlying working memory. Learn. Mem. **14**, 554–563 (2007)
22. Huttunen, K., Keranen, H., Vayrynen, E., Paakkonen, R., Leino, T.: Effect of cognitive load on speech prosody in aviation. Appl. Ergon. **42**(2), 348–357 (2011)
23. Endsley, M.: Situation awareness misconceptions and misunderstandings. J Cogn. Eng. Decis. Making **9**(1), 4–32 (2015)
24. Pruchnicki, S.A.: Increasing general aviation safety through data analysis. In: Conference on Proceedings of the Aviation, Aeronautics, Aerospace International Research Conference (2016)
25. Hutchins, E., Holder, B., Hayward, M.: Pilot attitudes toward automation. Collected Papers, Department of Cognitive Science, University of California, San Diego (1999)

Workload Assessment and Human Error Identification During the Task of Taking a Plain Abdominal Radiograph: A Case Study

Cesar Alferez-Padron[1], Aide Aracely Maldonado-Macías[1(✉)], Jorge García-Alcaraz[1], Liliana Avelar-Sosa[1], and Arturo Realyvasquez-Vargas[2]

[1] Autonomous University Ciudad Juárez, Del Charro Avenue 450 Partido Romero, 32310 Juarez City, Mexico
al150682@alumnos.uacj.mx, {amaldona, jorge.garcia, liliana.avelar}@uacj.mx
[2] Technological Institute of Tijuana, Calzada Tecnológico, S.N., 22414 Tijuana, Mexico
arturo.realyvazquez@tectijuana.edu.mx

Abstract. The following research, presents a case study of cognitive analyses of the tasks made by radiology technicians in a Mexican hospital. These professionals attend patients with different health conditions; likewise, they need to prepare the x-ray machine and position the patient correctly. For this reason, this study will determine the mental workload assessment and identify human errors of the procedure of taking a plain abdominal radiography. As results, a hierarchical task analysis diagram is presented, NASA TLX method found highest scores on physical and temporary demands. Recurrent human errors were the incorrect voltage and amperage adjustment and the incorrect identification of the patient using TAFEI method. Action for human error prevention is recommended. In addition, some digital support may be required to input patient's data and provide additional training sessions on those tasks with the greater mental workload and human error probability.

Keywords: Radiology technician · Hierarchical task analysis · Mental workload assessment · NASA task load index · Human error identification · Task Analysis for Error Identification

1 Introduction

Task analysis is a tool that allow engineers to describe the worker's activities more precisely [1]. Moreover, the task analysis provide knowledge to the engineer to teach workers what they will do [2]. This study was focused on the task analysis of radiology technician. This professionals need to be skilled to use a variety of machines and equipment for medical purposes like capture images of internal organs, bones and tissues to detect an illness, injury, and other medical conditions [3, 4]. In addition, it is necessary that the radiology technicians understands general anatomy principles [3]. Thus, revolutionary technological innovations have affected the ways in which

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_10

radiologists work, and have contributed to the changing pattern of this activity [5]. This case study was conducted in a Mexican private second level hospital located in Juarez City in northern Mexico. The hospital is one of the main local hospitals and one of the hoariest in Juarez City. There is an emergency, laboratory, blood bank, intensive care, surgery and x-ray department. The X-ray department operates 24 h and 7 days a week. A basic procedure performed by radiology technician is to take a plain abdominal radiography. This study focused on this procedure.

1.1 Problem Statement

The radiology technicians attend several patients with different health conditions. It is important that technicians have expertise in advanced anatomy, be aware of the different types of machines, maintain aseptic and sterile techniques, and know how to position the patient during the task of taking a plain abdominal radiograph. The technicians should give assistance to patient when is needed. Likewise, they should evaluate the amount of amperage and voltage of the equipment for obtaining a clear imaging; also, they must count with theoretical knowledge of radiology security.

Because of the complexity activities done by the radiology technicians, it is possible that they could be suffering of work stress symptoms and mental workload. For this reason, this study will analyze the mental workload of the procedure for taking plain abdominal radiograph.

1.2 Objectives

This study is analyzing the cognitive tasks in order to estimate the mental workload and human error that may be generated on the procedure for taking plain abdominal radiography.

The specific objectives are:

1. Develop the Hierarchical Task Analysis (HTA), to deploy all task's goals and sub goals.
2. Perform a mental workload assessment of tasks having in the procedure for taking plain abdominal radiography using the global index of NASA TLX method.
3. Identify potential errors according to the Task Analysis for Error Identification (TAFEI) method.
4. Make some recommendations in order to improve the procedure of taking a plain abdominal radiograph.

1.3 Justification

Human error identification and the assessment of mental workload will help propose recovery strategies and determine consequences associated with them. In addition, the study describes potential errors, displays potential consequences and recovery actions. It can determine the probability and criticality of errors, and offers associated remedies to improve workplace design and find strategies to reduce mental workload. It promotes the more efficient use of time during work, the enhancement of infrastructure and material resources usage; another associated benefit is to minimize the replication of radiographs to prevent the unnecessary exposure of patients to x-ray beam.

1.4 Delimitation

This study was held in a private second level Mexican hospital located in the Juarez City, Chihuahua. Results of this research were focused on taking a plain abdominal radiography, since it is the most frequent activity performed by the radiology technicians according to hospital records.

2 Literature Review

2.1 Hierarchical Task Analysis

One of the most important Human Factors (HF) methods is the Hierarchical Task analysis (HTA). Originally developed in response to the need for greater understanding of cognitive tasks [6]. HTA involves describing the activity under analysis in terms of a hierarchy of goals, sub-goals, operations and plans. The majority of HF analysis methods require an initial HTA, to explain information of the task, it is a good way define the most important thing of the task [6].

The HTA refers to a family of techniques used to describe and analyze operator performance within a human-machine system in manufacturing technologies, workload assessment and evaluation of the interface human- machine [7, 8]. For example, key skills now include perceptual skills for monitoring equipment, diagnostic skills for interpreting computerized information, and communication skills required for problem solving and co-ordination in distributed decision environments [9]. The literature highlights at least twelve additional applications to which HTA has been utilized, including interface design and evaluation, training, allocation of functions, job description, work organization, manual design, job aid design, error prediction and analysis, team task analysis, workload assessment and procedure design [6].

HTA is a framework for the analysis of tasks and make it clear that the methodology is based upon a theory of human performance, therefore can be used to describe both teamwork as the tasks that man made in the system [10]. HTA methods involved significant cognitive components (such as monitoring, anticipating, predicting, and decision making), a method of analyzing and representing this form of work was required [10].

2.2 Mental Workload

Mental workload is one of the most widely invoked concepts in ergonomics research and practice [11]. Mental workload as an “a measurable quantity of information processing demands placed on an individual by a task” [12]. In addition, other definition of mental workload is "that portion of the operator's limited capacity actually required to perform a particular task." All mental processing requires some resources [13]. Several authors have shown several definitions of mental workload; however, until now there is no clear definition of the construct and its measurement unit.

An advantage of using mental workload assessment techniques is to determine how much mental load is too much or hazardous for individuals; also determination of sub workload is important to define if individuals are sufficiently challenged to sustain useful levels of output [11]. The mental workload can be broadly comprised of two components: task demands and stress [8]. Mental workload becomes a product of the resources available to meet those task demands [14].

NASA TLX. NASA-TLX is a subjective mental workload assessment tool. NASA-TLX or National Aeronautics and Space Administration Task Load Index allow users to perform subjective workload assessments on operator(s) working with various human-machine systems. NASA-TLX is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales [15]. These subscales are Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level; each dimension is rated on a visual analogue scale, with five-point steps between 0 and 100 [6]. The participants evaluate each subscale during the experimental task. The analyst asks participants to determine workload in a scale divided into 20 intervals, ranging from low to high. The participants should perform an evaluation of the most effect of mental workload during the task under analysis through comparisons of pair of potential sources [15].

2.3 Human Error

Human error is defined as "a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency" [16]. Human errors occur during the lifetime of a product or system; or from design to the end of its useful life. Human beings are fallible to mistakes, either by omission in action, lack of training, lack of experience, among others [16, 17].

Human error occurs when human behavior and its influence on the system exceeds the limit, that is, when you do wrong actions [17]. The purpose of human error identification (HEI) techniques is the definition of points in the interaction between humans and artifacts, or systems [18]. The HEI objectives are [18]:

- Representing the full range of operations that people can perform using the artifact or system;
- Determining the types of error that are likely to occur;
- Assessing the consequence of errors for system performance;
- Generating strategies to prevent, or reduce the impact of errors.

Task Analysis for Error Identification (TAFEI). TAFEI has been used of to predict user performance [19]. It is a method to predict errors in the use of a device to model the interaction between the user and the device [6]. It is assumed that people use the device for a purpose such that the interaction is developed as a cooperative effort with this procedure problems arise [18]. The method assumes that the product restricts every action and provides information to the user about its functionality.

Thus, the interaction progresses while the user selects the action that is more relevant to reach your goal. The principal metrics of user performance are time and error, but TAFEI offer a means of predicting performance time in terms of possible errors [6]. In the initial model, it is assumed that erroneous actions can be corrected in one move (by returning to the start state, e.g., by pressing a Cancel button) [19]. However, it is a simple matter to develop the model to allow people to move to the previous state, e.g., by pressing an Undo button.

3 Methodology

According with the specific objectives, the methodology focuses on three stages with several steps that will be explained in the following paragraphs.

- Stage 1 Development of the hierarchical task analysis.
- Stage 2 Mental workload assessment.
- Stage 3 Human error identification.

3.1 Stage 1 Development of the Hierarchical Task Analysis

HTA is the principal technique using in the human factors. The procedure to create are [6]: 1. Identify user's primary goals, 2. Describe in detail the steps that the users must perform to accomplish their goals and 3. Optimize these procedures.

3.2 Stage 2 Mental Work Assessment

The next line describe the methodology according for NASA TLX [6].

1. Define the task under analysis.
   This is a very important part, because the analyst selects the task to will be analyzing. This part depends of the analyst focus.
2. Create an HTA for the task.
   Section of the analyst shows that the elements of the task.
3. Select participants.
   The subject will be randomly.
4. Inform participants about the mental workload method.
   This step refers to describe and inform participants about the purpose of the study and the NASA TLX method.
5. Perform the task under analysis.
   At this point, it is advisable that the analyst perform the NASA TLX when the participant has finished the task.

6. Give a weight procedure.
   The participant gave a weight to 15 pairwise comparisons of six sub scales. It is to identifying the scale with greater weight.
7. NASA TLX rating procedure.
   The participant gives a score of the scale between 1 (low) and 20 (high)
8. Score calculation.
   This score was calculated by multiplying each rating by the corresponding weight given by the participant. Then, Eq. 1 obtains the aggregation procedure of weighted ratings for each task. Finally, the workload will have a score between 0 and 100.

$$IC = \frac{1}{5}\sum_{i=1}^{6}(P_i X_i) \quad (1)$$

IC: load index
Pi: weight obtained for each dimension (first phase)
Xi: score for the dimension (second phase)

3.3 Stage 3 Human Error Identification

The TAFEI method shows transition states between human and machine that leads to human error identification. This technique was adopted to identify human errors in this study.

It comprises of three steps introduced below [6]:

1. Hierarchical task analysis (HTA) development
   HTA is the first step in TAFEI technique, if the HTA is accurate, the better result we can get from TAFEI technique. HTA refers to aims, operations and plans. Having completed this step, we turn to SSDs.
2. Deployment of State Space diagrams (SSDs)
   SSD is the list of states that it may happen in a machine and the interaction with human operator or another machine. The SSD consist of a representation of states that machine passes through task, from starting state to the finish the task. The state has a set of possible exits from the other states.
3. Determination of the transition matrix
   Transition matrix is an important step in TAFEI technique. All the possible states are established into this matrix. Transitions states of SSDs are included into the cells of the table.

4 Results

4.1 Stage 1 Development of the HTA

Two radiology technicians participated during the morning journey. Average age of all participants is 33 years old. The equipment used to take the x-ray radiograph is a Multix Pro® machine by Siemens® (Machine 1, M1), for analyzing the digital plate is used a desktop computer with the Agility® software provided by Agfa® (Machine 2, M2) and for printing the plate is used Axys® machine by Agfa® (Machine 3, M3). Using HTA the procedure of taking a plain abdominal radiograph is as following:

0. Procedure for taking plain abdominal radiography (Do: 1, 2, 3 and 4).
1. Patient Income (Do: 1.1, 1.2, 1.3, 1.4 and 1.5).
   1.1 Patient presents a medical request for plain abdominal radiograph.
   1.2 Receptionist asks to the patient if have a pacemaker or is pregnant.
   1.3 Receptionist gives a request to the radiology technician.
   1.4 Technician reads the request to identify the patient and study will perform.
   1.5 The technician transports the patient to a dressing room (the patient may be in a wheelchair or not) as well as the patient may be accompanied or not.
      1.5.1 The technician asks protocol questions.
      1.5.2 The technician provides a gown.
      1.5.3 The technician requests the patient to put on a gown.
      1.5.4 The technician applies for the patient to remove the metallic objects and that must place those in the dressing room.
2. Taking digital plate (Do: 2.1, 2.2, 2.3 and 2.4).
   2.1 The patient guided to the x-ray room by the technician.
   2.2 The technician checks that the patient has removed the metal objects.
      2.2.1 If the task not complete (Do task 1.5.4).
   2.3 The patient is lying on the x-ray table face up by the technician.
      2.3.1 If the patient is man, the technician provides a lead shield placed over the testes to protect against the radiation.
   2.4 The technician closes the door.
   2.5 The radiology technician prepares the x-ray machine (Do: 2.5.1, 2.5.2, 2.5.3, 2.5.4, 2.5.5, 2.5.6, 2.5.7, 2.5.8, 2.5.9, 2.5.10, 2.5.11 and 2.5.12).
      2.5.1 Position the x-ray tube near the abdominal area.
      2.5.2 Place the digital plate under the x-ray Table
      2.5.3 Re-position the patient, if it is necessary.
      2.5.4 Prepare the set-up for amperage and voltage.
      2.5.5 Turn off the lights.
      2.5.6 Stand behind a protective wall x-ray.
      2.5.7 Ask the patient to remain and hold your breath.
      2.5.8 Press the x-ray emission button.
      2.5.9 Remove the finger from the button.
      2.5.10 Request the patient to breathe.
      2.5.11 Turn on the lights.
      2.5.12 Ask to the patient if he/she feels well.
3. The technician prints the radiography (Do: 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10 and 3.11).
   3.1 Remove the digital plate under the Table
   3.2 Take the digital plate to the analysis station.
   3.3 Analyze the digital plate with a computer program.
   3.4 Capture patient data.
   3.5 Capture data from the request.

3.6 Check digital radiography.
3.7 Review the brightness and contrast of radiography using the computer software menu.
3.8 Adjust the brightness and contrast if it is necessary.
3.8.1 Is radiography clear? (Yes, Do 3.9) (No, Do task 2).
3.9 Store the information in the computer.
3.10 Print the radiograph.
3.11 Take the patient to the waiting room.
4. Delivery of patient studies (Do: 4.1, 4.2 and 4.3).
4.1 Technician packs radiography.
4.2 Technician delivers radiograph to the receptionist.
4.3 Receptionist asks the patient's name and type of study.
4.4 Receptionist delivers the radiograph to the patient.

4.2 Stage 2 Mental Work Assessment

Results of mental workload are shown in the Table 1. Format of NASA TLX method was used. The assessment was made when technicians completed the task of taking the plain abdominal radiograph. Results of workload index by participant were obtained for each workload dimension to identify those tasks with more mental workload. The average weighted NASA TLX index for mental workload was of 66.83, and the dimensions with highest scores were the physical and temporary demands.

These results indicate that this task is laborious and demanding and it must be done in short periods.

Table 1. Results of NASA TLX.

Dimension	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration	NASA score
Participant 1	120	270	210	160	110	65	62.33
Participant 2	150	300	210	130	140	140	71.33
Mean	135	285	210	145	125	102.5	66.83

4.3 Stage 3 Human Error Identification

According with face-to-face interviews with the radiology technicians, the state space TAFEI diagrams appear in Fig. 1. Additionally, the transition matrix by TAFEI is shown in the Table 2 and error descriptions appear in Table 3. The results of the TAFEI method identifies that the main human error during the task is the one concerning to the "incorrect voltage and amperage adjustment", the one concerning to the "incorrect identification of the patient" and the "kind of medical study to be done". Human errors can be identified by the letter E and by number.

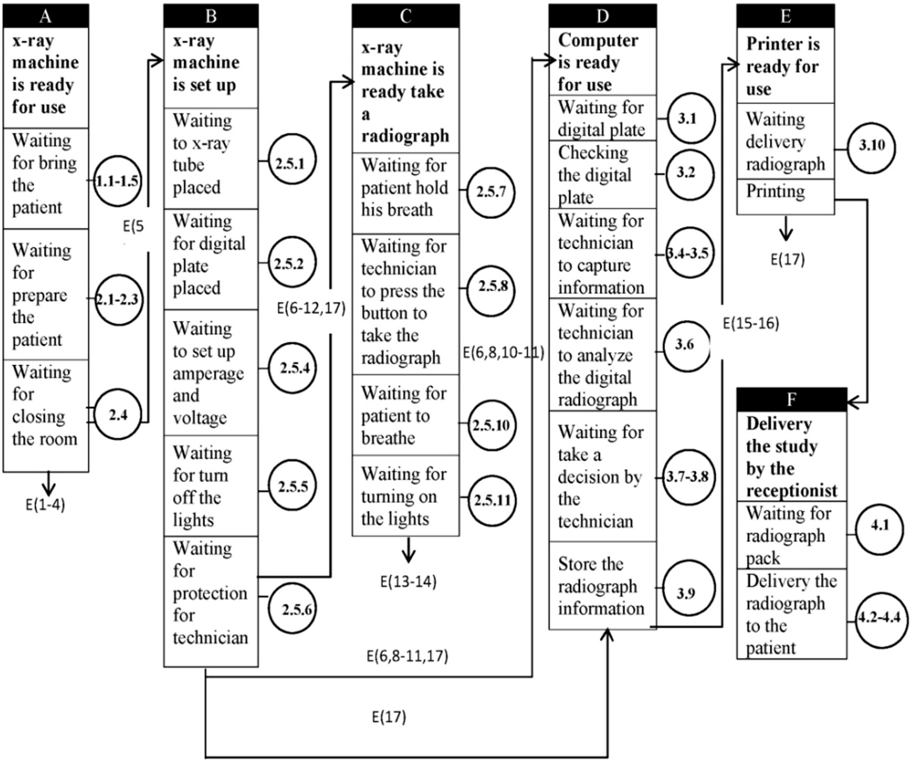


Fig. 1. State-space TAFEI diagram

Table 2. Transition matrix TAFEI

State	A	B	C	D	E	F
A	I E(1–4)	L E(5)	–	–	–	–
B	–	–	L (M1) E(6–12, 17)	I E(6,8–11, 17)	–	–
C	–	–	I E(13–14)	L (M2) E(6, 8, 10–11)	–	–
D	–	–	–	–	L M3 E(15)	–
E	–	–	–	–	I E(17)	L
F	–	–	–	–	–	–

L = legal; desirable and appropriate; I = illegal; undesirable and inappropriate;
– = Impossible; cannot happen; M = Machine, E = Error

Table 3. Error descriptions and design solutions

Error	Transition	Error description	Design solution
1	A to A	Incorrect patient	Check with the patient for his name Check the name with the patient's family
2	A to A	Incorrect procedure	Check with the patient for his request Check the name with the patient's family
3	A to A	Forget to read protocol	Give training
4	A to A	Forget to provide a gown	Give a checklist Give a procedure manual Give training
5	A to B	The patient carries a jewelry or metal things	Check if the patient carries a jewelry or metal things
6	B to D	The patient is not positioned correctly	Give a procedure manual
7	B to C	Forget to provide a lead shield for man.	Give a checklist Put a procedure sign on the wall to remind the technician and the patient
8	B to D	Incorrect part of the patient's body	Place the x-ray tube is place near the abdominal area Give a procedure manual
9	B to C	Forget to place the digital plate	Put a station of digital plate near Put a procedure sign on the wall to remind the technician and the patient
10	B to D	Patient moved.	Tell the patient not to move, because of the risk of give more radiation Ask for help for placing the patient
11	B to D	Amperage and voltage adjustments are incorrect.	Check the manual of technician and put correct amount of voltage and amperage
12	B to C	Forget to turn off the lights	Give a procedure manual
13	C to C	The technician remained pushed x-ray button	Only push the necessary x-ray beam
14	C to C	Forget ask for the patient's state	Check patient's state
15	D to E	Incorrect adjust of brightness and contrast	Give a procedure manual Ask for help
16	D to E	The radiography is not clear	Check the brightness and contrast Ask for help to adjust the brightness and contrast Take radiograph again Check the maintenance service Make a quality manual
17	E to E	Incorrect patient's radiograph	Ask the patient's name and request

5 Conclusions

Methods used in this study were adequate to describe the tasks, analyze those ones with the highest workload demand and identify potential human errors. Some conclusions can be made from their application. The HTA, shows that only one person performs multiple tasks, concerning input of data, management of patient and set up adjustments. As example, the task 2.5 involves twelve subgoals. The Table 1 shows that the dimensions with the highest workload are physical and temporal demands. According to participants, the physical demand shows an average value of 285, which may be related with those tasks that requires a significant amount of physical activity i.e. when technical helps the patient to lay on the X-ray table (task 2.3). In addition, temporal demand was the second highest score with an average value 210; this means that the technician feels some pressure of time to do the procedure and requires multiple subgoals to be completed. Based on the analysis of human error on TAFEI method, there are 17 possible human errors, shown in Table 3. Critical errors in the procedure are those that might damage the patient. The main error is the incorrect amperage and voltage adjustment. This is a critical error because in some cases it would be necessary to repeat almost the complete procedure from subtask 2. Another error is the incorrect identification of patient, when the radiology technician omits to ask the patient's name and the kind of requested study. Other important errors are when patient moves or is not correctly positioned, when the x-ray button remains pushed and when the adjustment of brightness and contrast is omitted or incorrect. Finally, when the procedure is repeated because the radiograph is not clear, there is a higher risk for the patient.

5.1 Recommendations

Some recommendations can be given for this task. Redistribution of tasks may be considered to involve a technical assistant. Additionally, some guides can be provided about body mass, height and weight to determine correct amperage and voltage adjustments. In addition, improvements must consider increase the focused attention on the activity, ask for the patient's name and give information about the risk of the procedure, provide a checklist of activities and duties. The design of a procedure flow diagram is highly recommended. Offer some training of radiology security; implement a training course dedicated on those tasks with greater mental workload and high potential of errors, and design a visual display for the procedure will help diminish the probability of human errors during this task.

References

1. Baber, C., Stanton, N.A.: Human error identification techniques applied to public technology: predictions compared with observed use. Appl. Ergon. **27**(2), 119–131 (1996). https://doi.org/10.1016/0003-6870(95)00067-4
2. Clark, R., Estes, F.: Cognitive task analysis for training. Int. J. Educ. Res. **25**, 403–417 (1996). https://doi.org/10.1016/S0883-0355(97)81235-9

3. Creus, A.: Fiabilidad y Seguridad. Marcombo Ediciones Técnicas, Spain (2005)
4. Fridell, K., Edgren, L., Lindsköld, L., Aspelin, P., Lundberg, N.: The impact of PACS on radiologists' work practice. J. Digit. Imaging **20**(4), 411–421 (2007). https://doi.org/10.1007/s10278-006-1054-1
5. Gael, S.: Job descriptions. In: The Job Analysis Handbook for Business, Industry and Government, vol. 1, pp. 71–86 (1988)
6. Hilburn, B.: Engineering psychology and cognitive ergonomics. In: Harris, D. (ed.) Proceedings of 7th International Conference, EPCE 2007, Held as Part of HCI International 2007, Beijing, China, 22–27 July 2007, pp. 497–503. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73331-7_54
7. Human Performance Research Group: NASA TASK LOAD INDEX (TLX), Moffett Field, California (1988)
8. Militello, L.G., Hutton, R.J.: Applied cognitive task analysis: a practitioner's toolkit for understanding cognitive task demands. Ergonomics **41**, 1618–1641 (1998)
9. Parasuraman, R., Hancock, P.A.: Adaptive control of mental workload. In: Desmond, P.A., Hancock, P.A. (eds.) Stress, Workload and Fatigue, pp. 305–320. Lawrence Erlbaum, New Jersey (2001)
10. Radiology Tech School. What kind of work does a radiology tech do day to day? (2012). http://onlineradiologytechnicianschools.com/what-kind-of-work-does-a-radiology-tech-do-day-to-day/. Accessed 15 Apr 2016
11. Radiology Technician. Job Description (2016). http://www.radiology-technician.com/job-description/#.Vw-xjPkrLIV. Accessed 14 Apr 2016
12. Reason, J.: Human Error. Cambridge University Press, Cambridge (1990)
13. Sanders, M., McCormick, E. (eds.): Human Factors in Engineering & Design, 7th edn. McGraw Hill, New York (1993)
14. Stanton, N.A.: Behavioral and cognitive methods. In: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H.W. (eds.) Handbook of Human Factors and Ergonomics Methods. Taylor & Francis Group, Boca Raton (2006)
15. Stanton, N.A., Baber, C.: On the cost-effectiveness of ergonomics. Appl. Ergon. **34**(5), 407–411 (2003). https://doi.org/10.1016/S0003-6870(03)00060-7
16. Stanton, N.A., Salmon, P., Rafferty, L., Walker, G., Baber, C., Jenkins, D.: Human Factors Methods: A Practical Guide for Engineering and Design, 2nd edn. Ashgate, Farnham (2013)
17. Tsang, P.S.: Mental workload. In: Karwowsky, W. (ed.) International Encyclopedia of Ergonomic and Human Factor, pp. 809–812. Taylor & Francis Group, Kentucky (2006)
18. Welford, A.T.: Mental workload as a function of demand, capacity, strategy and skill. Ergonomics **21**, 151–167 (1978)
19. Young, M., Stanton, N.: Mental workload. In: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods, pp. 39.1–39.7. CRC Press, United States (2005)

Spatial Perception

The Study of Sound and Shape Effects on Design

Tyan-Yu Wu(✉) and Alan Lee

Department of Industrial Design, Chang Gung University, 259, Wenhua 1st Rd, Guishan Dist, Taoyuan City, 33302, Taiwan (R.O.C.)
tnyuwu@mail.cgu.edu.tw, alanlee1990@yahoo.com.tw

Abstract. Exploring our environment in everyday life, multiple sensors are frequently applied in the learning processes. Previous design research mainly considered users' visual perception in terms of how they perceive an object visually and how they feel about the appearance of a product. Little research was focused on the investigation of the relationship between sound and shape characteristics. This study attempts to focus on the exploration of the cognitive relationship between object shape and sound effects. 54 participants (male 25, female 29; ages 18–25 years) were invited to participatory the experiment. Two variables experiment (i.e., shapes x sounds) were conducted in this study. Six types of stimuli including human, cars, animals (i.e., natural objects), cones, cubes, and spheres (i.e., artifacts) were adopted in the experiment. Each type of stimulus was displayed in pairs (e.g., large volume and small volume) on the screen. Audio stimulus was played after visual stimulus to help participants match the audio and visual stimuli. The results showed that there are significant congruent relationships between shape and sound. Participants were able to perceive and identify the larger volume of objects and to associate them with bigger sounds; in contrast, the smaller volume of objects to associate with the smaller sounds. Moreover, a small volume of objects is congruent with high pitch and low volume of sound. In visual identification, the shape with sharp angles is in accordance with large volume and high pitch of sound, while the curved shape appears to be congruent with small volume and low pitch of sounds. However, participants are likely to have difficulty in matching circle and square shapes with the sounds assigned. The results of this study should benefit designers in the development of interfaces enhancing shape and sound and attempting to have consistent sensations to users' mental model.

Keywords: Visual-audio · Cross-modal · Perception · Congruent

1 Introduction

Cross-modal correspondence plays a substantial role in product design. Some researchers have paid attention to multi-sensories design. In studying a variety of experience sensories, the research result showed that people could receive more information and elicit positive emotional response [1]. Moreover, an object or event which conveyed information emphasizing both visual and hearing sensations has a great potential to catch users' attentions consistently [2] and further possibly elicit their

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_11

positive emotions [3]. Conversely, if the product delivers incongruent sensory information, the users may have an incongruent cognition processes which leads to a confused reaction and further elicited negative emotion.

Cross-modal correspondence is a phenomenon concerning of one sensory experience to be connected to the other sensory experience [4]. For instance, without seeing, people could estimate the size, shape [5, 6], speed [7] and texture [8, 9] of the objects by judging the sounds. Some sensory connections can learn and build the semantic correlation through repeated experiences, such as dogs and barking, cats and meowing; this is called a principle of the up-button reaction [10]. However, some connections between different sensories can’t be explained clearly. For instance, brighter color balls connect to the higher pitch [3], the smaller size connects to the brighter color, the bigger size connects to the darker color [11], and the lower number locate in the left of mind space [12]. Previous design researches were mainly focused on visual perception only. Few studies focused on user perception of the integration between visual and audio sensations. Specifically, the present cognitive study attempts to focus on users’ perception to the relationship between visual (i.e., size & shape) and sound (i.e., sound-volume and pitch).

2 Methods

The aim of this study is to investigate the cognitive relationship between shape and sound of objects. In this cognitive and congruent study, two types of sensitive objects including natural basic geometry shape were investigated during the experiment. Their shapes and sizes were controlled by using six shapes with different sizes such as large and small.

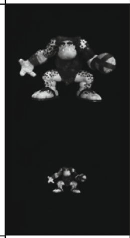
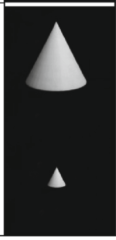
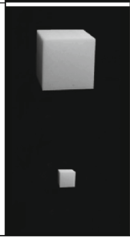
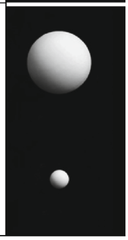
2.1 Participants

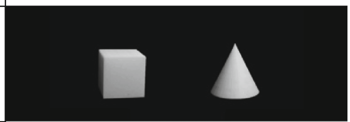
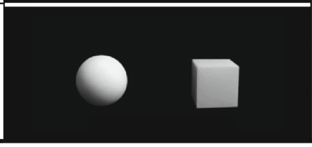
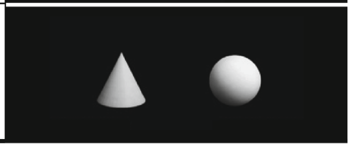
57 participants were randomly selected from three different departments in Chang Gung University. Three participants were eliminated due to their invalid questionnaires. Total 54 participants (i.e., 25 males and 29 females) including 18 medical students, 19 engineering students and 17 business management students (age ranging between 18–25 years old) participatory in this study.

2.2 Stimuli

In this experiment, we investigate two types of arrangements among stimuli: size and sound, shape and sound (see Table 1). For size and sound arrangement, three categories of objects (i.e., human, appliance and animal) were adopted as stimuli. Selecting stimuli (i.e., shape and sound), we adopted the objects which represent a variety of shapes and (along with appropriated) sounds, and are familiar to users in daily life. According to expert opinions, we selected monster to represent human, jeep to represent appliances, and chicken to represent animals because their shapes and sounds are familiar and easy to recognize by people. In addition, we selected three basic shapes

Table 1. The pictures of visual stimuli

Size vs. Sound							
Code	Pic	Code	Pic	Code	Pic	Code	Pic
Ex1		Ex2		Monster		Car	
Chicken		Cone		Cube		Sphere	

Shape vs. Sound			
Code	Pic	Code	Pic
Ex3		Cu_Co	
Sp_Cu		Co_Sp	

including cone, cube and sphere to represent the abstract artifact shapes. For shape and sound arrangement, there are three comparisons lay out: cube_cone, sphere_cube, and cone_sphere.

Three basic geometry shapes were created into 3D model formats by using 3D modelling software and we printed them out by using a 3D printer. To avoid bias, all visual stimuli was captured and edited by using Adobe Photoshop in order to have the same background on each image. At the end, the images were controlled as grayscale tone concisely (See Table 1).

The audio stimuli were classified into four types of sounds in order to match with visual stimuli; they were roar for monster, engine sound for the jeep, cluck for chicken and beep for the basic geometry shapes. Each type of sound matched with two sound-volumes (i.e., loud and low) and two pitches (i.e., high and low). Those sounds were adopted from Audio Database website [13]. Audacity was utilized to edit four different characters of sounds which covered high pitch, low pitch, high volume and low volume. Each sound was adjusted to avoid distortion, and to match with the quality in real environment and to avoid hearing impairment. (See Table 2).

Table 2. Audio stimuli settings

Object	Audio	Length	High pitch	low pitch	Loud volume	Low volume
Monster	Roar	3.8 s	4800 Hz	600 Hz	75 dB	55 dB
Jeep	Engine	3.4 s	3310 Hz	413 Hz	75 dB	60 dB
Chicken	Cluck	1.1 s	1911 Hz	735 Hz	75 dB	58 dB
Basic geo. shapes	Beep	2 s	3130 Hz	521 Hz	75 dB	55 dB

2.3 Questionnaire

12 pairs of stimuli (see Table 1) were included in the questionnaire. For the size and sound study, each pair (big size vs. small size) of stimulus appeared on the top of each page and had three check (options) boxes ☐ which appeared underneath of each stimulus image. Each check (options) box ☐ was indicated with "big", "uncertain", and "small" texts. In the test, participants are asked to check on one of the boxes after hearing the sound. For shape and sound study, each pair (i.e., two different geometrical shapes) of stimulus appeared on the top of each page and had three check (options) boxes ☐ appeared underneath of each stimulus image. Participants had to check a box to indicate which image is associated with sound attributions after the action.

2.4 Experimental Process

Researchers started with the explanation of research purpose, definition of terminologies and experimental instructions. Then, participants started to work on the first part of the questionnaire, the basic information. The experiment started with two examples for participants' practicing before six formal displayed stimuli. Next, participants were asked to judge the in/congruence between the object size and sound, and also between shape and sound after observing each image and listening to the sound. After finishing the experiment, each participant received 100 NTD as a reward.

3 Results

To investigate the relationship between size and sound, the data was run through Goodness of Fit test to confirm the significant match between size and sound. To do so, two steps were adopted to identify if there was significant difference. Firstly, the data of three options on matching test were run through Goodness of Fit test. At first step, if there are significant difference on three options; secondly, out of three, two options were further examined if there have significant difference (see Table 3). However, three pairs of stimuli (i.e., monster-volume, chicken-volume and sphere-volume) were not included in first step due to the zero option indicated. In Table 3, the statistic results showed that there are significant differences among each pair of size and sound. The results indicated that participants were able to identify most of size and sound significantly. In other words, participants perceived and showed congruent relationship between sizes (big and small) and sounds (loud volume and low volume, high pitch and low pitch) consistently. The details results are shown in Table 3.

Table 3. t-Test on stimuli with size and sound match

	Chi-square test	F	p
Monster_Volume	32.667	1	.000
Monster_Pitch	24.923	1	.000
Car_Volume	48.077	1	.000
Car_Pitch	18.000	1	.000
Chicken_Volume	38.208	1	.000
Chicken_Pitch	26.741	1	.000
Cone_Volume	8.647	1	.000
Cone_Pitch	25.920	1	.000
Cube_Volume	19.692	1	.003
Cube_Pitch	30.083	1	.000
Sphere_Volume	7.407	1	.006
Sphere_Pitch	24.020	1	.000

To investigate the relationship between shape and sound, similarly, the Goodness of Fit test was utilized to confirm the significant match between shape and sound. By using the Chi-square test, the result shows that there are significant different on three pairs: cube-cone and the volume ($p = 0.006$), cube-cone and the pitch ($p < 0.001$), cone-sphere and the volume ($p < 0.001$). However, sphere-cube and the volume ($p = 0.607$) and sphere-cube and the pitch ($p = 0.066$) were not significantly different (See Table 4).

Table 4. t-Test on stimuli with shape and sound match (pitch and sound volume)

	Chi-square test	F	P
Cu_Co_V	10.333	2	.006
Cu_Co_P	62.111	2	.000
Sp_Cu_V	1.000	2	.607
Sp_Cu_P	5.444	2	.066
Co_Sp_V	24.111	2	.000

Notes: Cu = Cube, Co = Cone,
Sp = Sphere, V = Volume, P = Pitch

Follows up the last step, two options out of three indicated significant differences were run through the Chi-square test repeatedly. The result showed that four pairs of stimuli indicated significant difference, which were cube-cone and the sound volume ($p = 0.022$), cube-cone and the pitch ($p < 0.001$), cone-sphere and the pitch

($p < 0.001$) and cube-cone and the volume ($p < 0.001$). The results indicated that participants are able to identify shape and sound consistently. However, two pairs showed no significant differences, which including sphere-cube and the volume ($p = 0.317$) and sphere-cube and the pitch ($p = 0.763$) (See Table 5).

Table 5. t-Test on stimuli with shape and sound match (pitch and sound volume)

	Chi-square test	F	p
Cu_Co_V	5.233	1	.022
Cu_Co_P	25.830	1	.000
Sp_Cu_V	1.000	1	.317
Sp_Cu_P	0.091	1	.763
Co_Sp_V	15.364	1	.000
Co_Sp_P	35.852	1	.000

Notes: Cu = Cube, Co = Cone, Sp = Sphere, V = Volume, P = Pitch

4 Discussion

The present study has examined the perception of audio-video congruent and incongruent and elucidated the multisensory integration effect. Table 6 summarized the results of Tables 3 and 5, which showed that participants enable to identify the most of positive relationship between visual and audio consistently. For instance, participants identify a big size of object as big when perceived higher sound volume and lower sound pitch, which reflects our cognition processes of understanding physical size and sound volume congruently. Conversely, participants perceived a smaller size of object as small, when perceived lower sound volume and higher sound pitch. Prior results can be applied to natural (i.e., monster and chicken), artifact (i.e., car), and geometry (i.e., cube, cone and sphere) objects. For instance, participants are able to identify the huge monster to loud sound, while small jeep to low sound volume. More details are shown at Table 6.

Table 6. Summary of significant congruent relationships between visual (size, shape) and audio (volume, pitch)

			Sound			
			Volume		Pitch	
			Loud	Low	High	Low
Visual	Size	Monster	Big	Small	Small	Big
		Jeep	Big	Small	Small	Big
		Chicken	Big	Small	Small	Big
		Cone	Big	Small	Small	Big
		Cube	Big	Small	Small	Big
		Sphere	Big	Small	Small	Big
	Shape	Cube-cone	Cone	Cube	Cone	Cube
		Cone-sphere	Cone	Sphere	Cone	Sphere

Moreover, when visually perceiving an object with a sharper shape (e.g., cone), participants were able to associate the shape with higher volume and pitch congruently. When perceiving an object with a rounded shape (e.g., cube and sphere), they were able to associate with lower volume and pitch congruently (See Table 6). However, the results showed that participants have difficulty in distinguishing between cube and sphere, and sound. Prior results are inconsistent with Walker's et al. [14] results, which indicated that 2D sharp-angle graphic was congruent with high pitch, while round angle was congruent with low pitch. We suspect that, in this experiment participants perhaps pay more attention to the whole 3D volume and less attention on individual angle features, compared with 2D graphic object. Prior results might explained by the Gestalt theory. However, it is suggested to have further studies on these details in the future.

5 Suggestions

This study has tested the congruency between visual and audio. The results may suggest that designers follow a rule when making decisions to the product shape, size and their sound consistently. The study allows designers to create an object that's fit into users' cognitive process, and will provide users with a better useful affordance in product design in terms of ease to understand and to operate the design interface, when followed the results of this study. We understand that there are strong interrelations between product emotions and the cognitive processes [15]. It means that users' emotional responses are influenced by the understanding of the object affordance (i.e., perceiving the way to use objects subconsciously). Thus, we believe that if an audio-visual object is designed congruent then it should be able to elicit users' positive emotions immediately. On the other hand, the object with audio and visual incongruence may elicit negative emotions due to the confusion of operating the objects. Currently, some research had applied event-related potential (ERP) to investigate the relationship of Cross-modal and emotion. Thus, for the future research, we may use ERPs to explore brainwaves when observing the shape and sound in/congruency, and perhaps also the emotional response when observing audio-video interface objects.

References

1. Norman, D.A.: The Design of Future Things. Basic Books, New York (2007)
2. Spelke, E.: Infants' intermodal perception of events. Cogn. Psychol. **8**, 553–560 (1976)
3. Mondloch, C.J., Maurer, D.: Do small white balls squeak? Pitch-object correspondences in young children. Cogn. Affect. Behav. Neurosci. **4**, 133–136 (2004)
4. Parise, C.V., Spence, C.: Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. Exp. Brain Res. **22**, 319–333 (2012)
5. Carello, C., Anderson, K., Kunkler-Peck, A.: Perception of object length by sound. Psychol. Sci. **9**(3), 211–214 (1998)
6. Grassi, M.: Do we hear size or sound: balls dropped on plates. Percept. Psychophys. **67**, 274–284 (2005)

7. Houben, M.M.J., Kohlrausch, A., Hermes, D.J.: Perception of the size and speed of rolling balls by sound. Speech Commun. **43**, 331–345 (2004)
8. Hermes, Dik J.: Measuring the perceptual similarity of pitch contours. J. Speech Lang. Hear. Res. **41**, 73–82 (1998)
9. Giordano, B.L., McAdams, S.: Material identification of real impact sounds: effects of size variation in steel, glass, wood, and plexiglass plates. J. Acoust. Soc. Am. **119**(2), 1171–1181 (2006)
10. Bien, N., ten Oever, S., Goebel, R., Sack, A.T.: The sound of size: Crossmodal binding in pitch-size synesthesia: a combined TMS, EEG, and psychophysics study. Neuroimage **59**, 663–672 (2012)
11. Kadosh, R.C., Henik, A., Walsh, V.: Small is bright and big is dark in synaesthesia. Curr. Biol. **17**, R834–R835 (2007)
12. Dehaene, S., Bossini, S., Giraux, P.: The mental representation of parity and number magnitude. J. Exp. Psychol. Gen. **122**, 371–396 (1993)
13. Freesound. http://www.freesound.org
14. Walker, P., Bremner, J.G., Mason, U., Spring, J., Mattock, K., Slater, A., Johnson, S.P.: Preverbal infants' sensitivity to synesthetic cross-modality correspondences. Psychol. Sci. **21**, 21–25 (2010)
15. Desmet, P.M.A., Hekkert, P.: Framework of product experience. Int. J. des. **1**, 57–66 (2007)

Research and Analysis on the Influence Factors of Spatial Perception Ability

Minxia Liu(✉), Chengtong Tang, Lin Gong, and Qing Xue

Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China
liuminxia@bit.edu.cn

Abstract. This paper mainly analyzes the influence of the spatial cognitive ability from the factors such as shape (strip shape, block shape, irregular shape), inherent thought, and gender. Through the study of human spatial perception ability processing mechanism, and combine the results of a large number of experimental data to explore the influence of these factors on the spatial cognitive ability. The results show: The block shape reaction time is the shortest and the irregular shape's is the longest; When the order of shape appearance accord with people's inherent way of thinking what has the shortest reaction time; the effect of gender difference is not obvious, which is different from the existing research results, which need to be further studied.

Keywords: Spatial perception · Shape · Inherent way of thinking · Gender

1 Introduction

Spatial perception is a perception which can reflect objet shape, size, distance and etc., because of that human obtain perceptual knowledge through spatial perception procedure, and then attain the essential rules of natural things by thinking process. Therefore, spatial perception is important in human cognition activity.

Spatial perception ability means the thinking ability of human by using 3-dimension pattern, means complicated perception what is integrated of vision, audition, sensation and kinanesthesia, the perception conclude shape perception,size perception, distance perception, depth perception and position perception. Spatial perception is acquired by study, not innate, and the study is carried on during early years, might not be aware of and might not be remember.

Spatial recognition ability is a main dimension in intelligence structure, and is closely related with profession performance of some certain group. In recent years, research on vision perception ability and recognition on vision perception became more deep-going and improved, and is also applied in some other field, this 2 example below state how to apply partial perception in daily lives.

(1) Apply of Montage in cinematics: Montage method is a method what combines two or more elements into a new content. Montage thinking is a creative thinking which based on vision and audition. Relationship between segmentation and combination of camera lens in Montage film, is a relationship between separation

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_12

and combination of different elements, through perception effect of analysis and compositive, and must rely on psychology affections of association at last. Association could cause memory of the past and imagination of the future, this important psychology phenomenon is just a bridge between the front camera lens and the back camera lens in Montage structure for picture communication. This psychology phenomenon cause memory and enlighten imagination, during a film process, induce audience produce artistic thinking about Montage structure from analysis to synthetic, from part to integral. Therefore, Montage method mainly comprehensive utilize human sense, and explained space through combination progress of different senses, make human spatial perception Pluralistic and stereoscopic. This method erects a bridge between human experience and spatial, makes the distance between human and spatial closer, enhance human ability for experience spatial.

(2) Apply of spatial perception in drawing field: drawing establish human spatial perception mainly by vision space imitating, and left human imagine space for imagination. Vision is the most important sense impression, the most direct way for human come into contact with the spatial. Just about vision what maximize the communication between human and space, makes human experienced space, and makes space accept us.

We conducted experiments with space perception experiment instrument, and the analysis results are based on the collection data of 50 undergraduate students' experiments, and we used SPSS software to analyze 106 experimental number [1]. This paper mainly discusses and analyzes the influencing factors of spatial cognition ability from four aspects:

(1) The influence of shape on spatial cognitive ability;
(2) The influence of inherent thinking on spatial cognitive ability;
(3) The influence of gender on spatial cognitive ability;
(4) The correlation between the number of errors and the reaction time in the experimental process of spatial cognitive ability.

2 Experimental Description

(1) The shape of stimulation signal: 4 × 4 box formed stimulus display, there are 4 stimulus squares appear one time, which are composed three patterns of bar, box-shaped, irregular shape, and it would be 2 stimulus display of one pattern (a total of 6 categories), as shown in Fig. 1.
(2) Set the number of experiments to 20 times for each category (a total of 6 categories);
(3) After selecting a stimulus category, press the "Start" button to experiment. The corresponding way is random between the stimulus position appeared on the screen with the keyboard number, such as 3-A, 4-B, 2-C, 1-D, or 1-A, 3-B, 4-C, 2-D or other corresponding way.

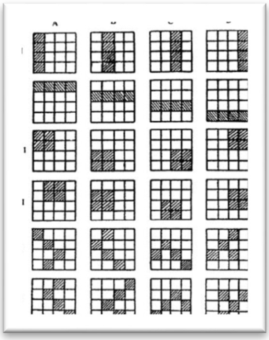


Fig. 1. Stimulate shape classification

(4) Once the stimulate signal appeared on the instrument screen, the subjects should press the test keyboard of a number key as soon as possible. If the reaction is consistent with correct way, then the stimulate signal will disappear. On the other hand, the subjects should immediately press the other keys until the response is correct.
(5) The experiment ends automatically when the number of experiments reaches the set number of times, and the experimental test results will be turn on the screen, including the average reaction time, keyboard mode, the number of errors, etc.

3 Analysis and Processing of Experimental Data

We use excel and SPSS software to analyze the experimental data, part of the experimental data collation results as shown in Tables 1, 2, 3, 4, 5 and 6 below:

Table 1. The experimental data collation of bar I

The experimental data collation of bar I							
Correct reaction numbers (N)	27	26	12	22	24	26	……
Average correct reaction times (s)	1.152	1.121	1.523	0.824 s	1.624	1.043	……
Wrong reaction numbers (N)	3	4	18	8	6	3	……
Average wrong reaction times (s)	2.679	2.28	3.243	1.925 s	1.632	2.196	……
Total reaction numbers (N)	30	30	30	30	30	29	……
Average reaction times (s)	**1.305**	**1.276**	**2.555**	**1.117**	**1.626**	**1.162**	**……**
The last number of wrong reaction (N)	25	23	29	16	24	25	……
The last number of continuous correct reaction (N)	5	7	18	14	6	4	……
Average reaction times of the last continuous correct reaction numbers (s)	1.225	1.647	1.623	0.862 s	1.004	0.844	……

Table 2. The experimental data collation of bar II

The experimental data collation of bar II							
Correct reaction numbers (N)	24	29	20	27	22	26	……
Average correct reaction times (s)	0.97	0.678	1.212	1.153 s	0.925	0.768	……
Wrong reaction numbers (N)	6	1	10	3	8	3	……
Average wrong reaction times (s)	2.028	2.276	1.231	2.188 s	1.241	1.348	……
Total reaction numbers (N)	30	30	30	30	30	29	……
Average reaction times (s)	**1.182**	**0.731**	**1.218**	**1.257**	**1.009**	**0.828**	**……**
The last number of wrong reaction (N)	30	2	27	19	23	14	……
The last number of continuous correct reaction (N)	0	28	3	11	7	15	……
Average reaction times of the last continuous correct reaction numbers (s)	1.231	0.664	1.225	1.141 s	1.235	0.705	……

Table 3. The experimental data collation of block I

The experimental data collation of block I							
Correct reaction numbers (N)	24	27	25	28	25	27	……
Average correct reaction times (s)	0.684	0.86	1.23	0.588 s	1.564	0.761	……
Wrong reaction numbers (N)	6	3	5	2	5	3	……
Average wrong reaction times (s)	1.868	2.171	1.33	1.752 s	1.679	1.826	……
Total reaction numbers (N)	30	30	30	30	30	30	……
Average reaction times (s)	**0.921**	**0.991**	**1.247**	**0.665**	**1.583**	**0.867**	**……**
The last number of wrong reaction (N)	28	25	27	3	24	22	……
The last number of continuous correct reaction (N)	2	5	3	27	6	8	……
Average reaction times of the last continuous correct reaction numbers (s)	1.025	0.629	1.236	0.588 s	1.124	0.892	……

Table 4. The experimental data collation of block II

The experimental data collation of block II							
Correct reaction numbers (N)	20	27	28	23	26	29	……
Average correct reaction times (s)	1.1	0.799	1.33	0.753 s	2.615	0.917	……
Wrong reaction numbers (N)	10	3	2	7	4	1	……
Average wrong reaction times (s)	2.695	2.059	1.2	1.721 s	2.451	4.484	……
Total reaction numbers (N)	30	30	30	30	30	30	……
Average reaction times (s)	**1.632**	**0.925**	**1.32**	**0.979 s**	**2.593**	**1.036**	**……**
The last number of wrong reaction (N)	28	11	29	5	28	1	……
The last number of continuous correct reaction (N)	2	19	27	25	2	29	……
Average reaction times of the last continuous correct reaction numbers (s)	1.623	0.787	1.31	0.568 s	2.419	0.917	……

Table 5. The experimental data collation of irregular I

The experimental data collation of irregular I							
Correct reaction numbers (N)	22	23	28	26	45	22	……
Average correct reaction times (s)	1.104	1.142	1.32	0.916 s	2.431	0.758	……
Wrong reaction numbers(N)	8	7	2	5	5	8	……
Average wrong reaction times(s)	1.685	2.898	1.25	2.342 s	2.498	2.463	……
Total reaction numbers (N)	30	30	30	31	50	30	……
Average reaction times (s)	**1.258**	**1.551**	**1.31**	**1.146**	**2.438**	**1.213**	**……**
The last number of wrong reaction (N)	22	30	27	29	47	19	……
The last number of continuous correct reaction (N)	8	0	3	2	3	11	……
Average reaction times of the last continuous correct reaction numbers (s)	1.089	1.346	1.3	1.041 s	2.601	0.777	……

Table 6. The experimental data collation of irregular II

The experimental data collation of irregular II							
Correct reaction numbers (N)	20	24	30	26	44	26	……
Average correct reaction times (s)	1.755	1.103	1.35	0.795 s	2.364	1.073	……
Wrong reaction numbers (N)	10	6	0	4	6	4	……
Average wrong reaction times (s)	2.742	3.011	0	1.699 s	2.785	1.761	……
Total reaction numbers (N)	30	30	30	30	50	30	……
Average reaction times (s)	**2.084**	**1.485**	**1.35**	**0.916**	**2.414**	**1.165**	**……**
The last number of wrong reaction	25	29	0	11	46	11	……
The last number of continuous correct reaction (N)	5	1	30	19	4	19	……
Average reaction times of the last continuous correct reaction numbers (s)	1.466	1.357	1.35	0.834 s	2.019	1.095	……

Tables 1, 2, 3, 4, 5 and 6 above represent the results of the experiment under different shapes of stimuli, which lays the foundation for the analysis of the influence of different shapes on the spatial perception ability.

3.1 The Effect of Shape on Spatial Perception

On the basis of the above data and analysis of the experimental data, the average response time under different shapes is compared and analyzed, as shown in Fig. 2.

As can be seen from the diagram: the average response time of the block is the shortest, 0.9943 s, the average response of irregular graphics is 1.2562 s, the average response of the strip is centered, 1.1028 s. The analysis of the results is as follows:

(1) Strip and block shape are simple graphics, exist in the person's memory who is in test, the person do not need to pay attention to the shape of the special memory, just need remember the location. Bar graphics test at first, and block graphics test

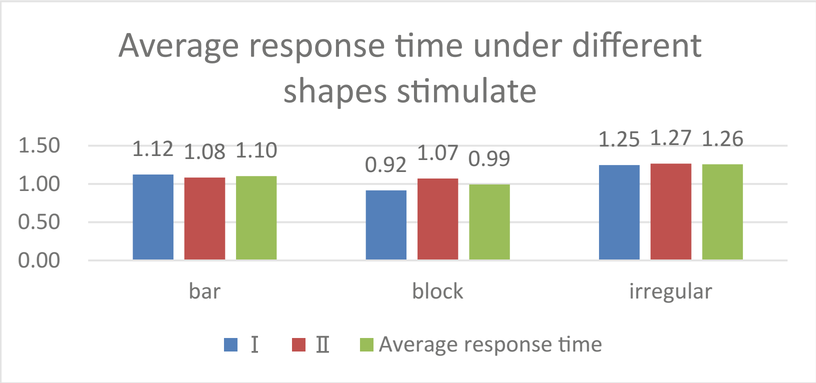


Fig. 2. Average response time under different shapes stimulate

second. The person in test is not familiar with the rules of the experiment at the beginning, he will take some time to get familiar with the rules. When the person gets familiar with the rules, the guy will focus their attention on the experiment itself, so the response time costs of the block is shorter than that of the strip;

(2) The sensory organs received external stimuli through brain processing, analyzing and processing, to form the overall understanding of the stimulus, which is perception. Perception is based on feeling, it is the result of the interaction between the newly accepted stimulus and the stored experience and knowledge in the brain [2]. But perception is not the sum of all the senses simply. Except sense of perception, there also includes memory, thinking and speech activities, etc. in perception, perception is higher period than the perception of perceptual knowledge stage. When the person in test observed the graphics on the screen, graphics shape, spatial location is mapped to the brain, the person tries to guess the representative figures, and format the relationship memory between correct number and its representative range in brain. When the figure appears on the screen again, the process of the recognition for the graphics is generally divided into three stages: analysis, comparison and decision making. The first stage of the recognition process is to analyze the information extracted from the sensory storage and break it down into components. The second stage is the comparison of the information to be extracted and the memory stored in the mind. After comparison, which one can give the best match of the code, the person need to make a decision. This is the task what the decision-making stage need to do, and finally determine the output of the identification system, that is, to recognize the object.

(3) For simple graphics, such as bar and block shape, with our previous experience, these two graphics already exists in memory, we only need to remember their position and by encoding, which is relatively simpler. But for irregular graphics, the person in test first saw the subjects and try to remember them, scan them with eyes, and format a scan route, establish memory traces in mind, the trace record perception and action activities. In the first stage of recognition, the person will change the irregular graphics into two familiar graphics (slash and box) to

remember. When the person sees the same graphics again, I will match the memory with it to make a decision. For complex graphics, it takes more time for the brain to encode and distinguish, so he need longer encoding time and processing time.

3.2 Influence of Intrinsic Thinking on Spatial Perception

The inherent thinking or thinking refers to [3]: In psychological view, with the premise that without the effect what is strong enough to change the current view of the evidence, many times to strengthen the same consciousness can make human thinking to maintain the original direction of thinking, which produced the inertia of thinking. Similarly, people in practice in the application of the law is the same, the solution to the practical problems permeated the same measures that encountered formerly. This is the direct cause of inertial thinking. As to the root cause of its evolution, it is related to biological evolution. The generation of habitual thinking is the result of the interaction of internal and external factors, external factors are objective things, and internal factors are the special functions of human brain. Therefore, the inertial thinking is still essentially consciousness, is a unique product of the human brain, it is rooted in the evolution of the advanced biology.

In addition to the ability of human beings to reflect the condition of the animals, the brain also has the ability to produce consciousness, which is the result of evolution, but also conforms to the law of the development of things from low to high-grade. The human brain can not only feel the outside stimulation, but also can select valuable factors from the stimulation, and format theory by thinking and, this is the process from perceptual knowledge to rational knowledge. The correct theory Among them has become the truth through practice, and it will naturally become the guiding ideology or basic principle of dealing with one or more events in a long time. As a result, repeated use of one certain truth, is actually repeated apply of a certain law, Is equal to stimulate the human brain repeatedly, it will create a strong conditioned reflex, so people in the future practice in similar situations will instinctively give priority to the original solution. Although the human deal with new information according to the different actual situation, discard the dross and select the essential, discard the false and retain the true, but completely get rid of the influence of conditioned reflex is impossible, because people are not machines, cannot abandon the inherent function as biology.

Space perception test process in this paper, at the beginning, person in test was easily affected by inertia thinking to response: he habitually thinks digital key sequence and graphics on the keyboard should be relative, that is in accordance with the “1-A 2-B 3-C 4-D” mode. But the actual experimental set is random, generally it’s not presented in accordance with the “1-A 2-B 3-C 4-D” mode, therefore, at the beginning of the experiment, the reaction cost more time than later, and more errors occured, only when the graphics sequence that appeared in the screen was of the order and the corresponding graphics appeared on the screen is in consist with the corresponding number in keyboard for knock on (i.e. when the keyboard is “1-A 2-B 3-C 4-D”) the time cost of reaction is the shortest and the number of error is the least.

In this paper, we study the influence of keyboard mode on the spatial perception ability by using the data of graphics in group II stimulate (presentation pattern was shown in Fig. 3). The results of the average response time of the different keyboard mode in the course of strip II experiment were calculated as shown in Table 7:

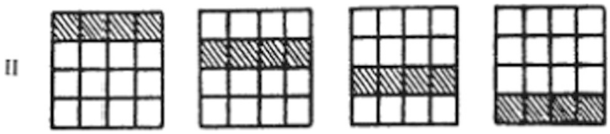


Fig. 3. Bar II graphics stimulus presentation pattern

As shown in Table 7, the average response time is significantly less than the other way when the keyboard is “1-A 2-B 3-C 4-D”, as a result of our habitual thinking (group stereotype). Normally the brain’s way of thinking and the sequence of visual observation is from top to bottom, our inertia thinking will let us feel that the number from top to bottom is 1, 2, 3, 4, therefore, when the keyboard mode is “1-A 2-B 3-C 4-D”, the response time is the shortest.

Table 7. Average response time table under different keyboard mode

Keyboard mode	**1-A**	1-A	2-A	2-A	2-A	3-A	3-A	3-A	3-A
	2-B	4-B	1-B	3-B	4-B	2-B	1-B	2-B	4-B
	3-C	3-C	4-C	4-C	3-C	4-C	4-C	1-C	3-C
	4-D	2-D	3-D	1-D	1-D	1-D	2-D	4-D	2-D
Total mean reaction time/s	**0.497**	0.829	1.096	0.701	0.784	1.559	1.012	1.087	0.967

3.3 The Effect of Gender on Spatial Perception

6 groups of boys and 6 groups of girls (52persons/group) for statistical analysis, get the average reaction time difference is shown in Table 8:

Table 8. Average reaction time of different groups

		Reaction time of girls	Reaction time of boys
Group	1	0.9992	1.0027
	2	0.9474	0.9752
	3	0.9219	0.8351
	4	0.9337	0.9213
	5	1.1424	1.0925
	6	1.1510	1.1328

Assuming a significant level of 0.005, using statistical software SPSS for 6 groups boys and 6 groups girls (52persons/group) for statistical analysis, the results is shown in Tables 9 and 10 below:

Table 9. Cross table of gender and reaction time

Count		Reaction time						
		.8351	.9213	.9219	.9337	.9474	.9752	.9992
Gender	Male	1	1	0	0	0	1	0
	Female	0	0	1	1	1	0	1
Total		1	1	1	1	1	1	1

Table 10. Symmetric measures

		Value	Approx. sig.
Nominal by nominal	Contingency coefficient	.707	.364
N of valid cases		12	

Table 9 and 10 shows: the significant coefficient of the 2 variables (gender and reaction time) is 0.364 > 0.05, so the correlation of spatial perception reaction time and gender is not significant; the boys average response time is 0.9933 what is less than girls 1.0159. The experimental results showed that there was no significant correlation between gender and the ability of space perception, it's different with the existing research results (many literatures shows that there is a certain relationship between the two), this may be because the experimental data might be fake or experimental samples is not wide enough, pacific analysis of the causes are as follows:

Physiological Causes. The gender differences in spatial cognitive ability are affected by physiological factors at first. Modern physiology, genetics and psychology studies show that the different chromosome characteristics of male and female, the asymmetry development of the two hemispheres of the brain function and sex hormones are relevant with the cognitive style differences about different gender.

The developmental speed and level of functional asymmetry in the two hemispheres of the brain are different. Women left hemisphere functional is more specialized, so they are good at calculation, classification and speech activities associated with intelligence activities, they are good at decompose a task into multiple details, and then processing meticulous; and men right hemisphere function specialization is better than women, so they are better at dealing with nonverbal aspects including spatial related information, and good at dealing with problems in a comprehensive way.

Different Cognitive Characteristics of Men and Women. The visual perception of the male is higher than that of the female, and the hearing, smell, pain, touch of female is more sensitive than that of the male. Therefore, the female has better auditory and speech perception, while the male has better visual and spatial perception than the female.

4 Analysis of the Correlation Between the Number of Errors and the Reaction Time in the Process of Spatial Perception

This paper makes a further analysis of the correlation between the number of errors and the reaction time in the process of spatial perception ability test. The results of correlation analysis using the statistical software SPSS are shown in Table 11 and Fig. 4 below:

Table 11. Correlation between the number of errors and reaction time

		Error response times (N)	Total mean reaction time (s)
Error response times (N)	Pearson correlation	1	.348**
	Sig. (2-tailed)		.000
	N	624	624
Total mean reaction time(s)	Pearson correlation	.348**	1
	Sig. (2-tailed)	.000	
	N	624	624

**Correlation is significant at the 0.01 level (2-tailed).

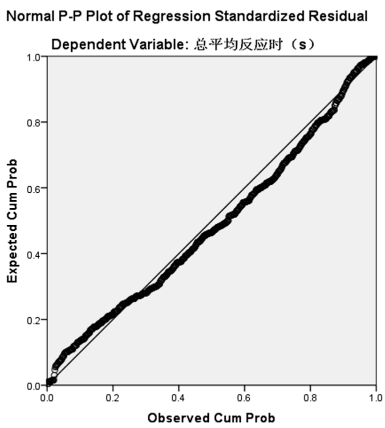


Fig. 4. Standardized residuals of normal regression

The above analysis shows that the tail probability of P is 0.000 < 0.05, which shows that there is a significant correlation between the number of errors and the average response time. Pearson correlation coefficient r = 0.348, indicating that they are moderately correlated. In other words, the more the number of errors, the longer the average reaction time costs, but the change is not particularly obvious. Specific analysis of the reasons are as follows:

(1) The beginning of the experiment there will be a learning process, the subjects initially did not know what the corresponding figure in different buttons, so need random attempt at first, there may be many times to special circumstances, so costs in the process of learning error number is more, reaction time is long. On the contrary, if fortunate, the first attempt could find out the corresponding graphics button, then the number of errors and reaction time will be reduced. This may be one of the reasons for the two variables.
(2) At the same time, the more the number of errors, the person in test will be relatively cautious in the subsequent experiments, the time cost of the reaction will increase. If the previous experiments carried out smoothly, then the person in test will be more confident, the reaction time will be shortened accordingly.

5 Conclusion and Discussion

Through the analysis of a large number of experimental data, we can draw the following conclusions:

(1) Different shapes of the stimulus has a certain impact on the spatial perception ability, standardized rules and pattern stimuli required shorter reaction time, irregular graphics stimulus response time is relatively long, this is because the human brain information processing is generally going through analysis, comparison and decision three stages, for complex graphics, because of the brain encoding memory and longtime costs for the identification processing, so person in test needs a longer reaction time;
(2) The inherent thinking will have a certain impact on the spatial cognitive ability, when the mode appeared during the experience with stimulus presentation of the inertia of thinking, people can be faster and more accurate to complete the experimental task by inertia of thinking, Conversely, the reaction time and error times should increase;
(3) Based on the analysis of the experimental data, the conclusion is that there is no correlation between gender and spatial perception, which is different from previous studies;
(4) The results of this study shows that the number of error response and the total average reaction was significantly correlated between the correlation coefficient, Pearson Correlation Coefficient shows that they were moderately related, overall the more the number of error response is, the total average reaction time is longer, but the change is not very obvious.

Acknowledgments. This paper is supported by National Natural Science Fund project, project approval number: A0920132002. I would like to thanks to my family, colleagues and graduates for their valuable comments and suggestions.

References

1. Guo, J., Sang, Y., Jiang, K., Li, J.: Application of flash fusion frequency test method in beijing public transit driver. Saf. Sci. Technol. **10**, 9–12 (2011)
2. Xue, Q., Pang, Y.: Comprehensive training tutorial of human factor. Beijing Institute of Technology Press (2007)
3. Chen, Q.: Out of misunderstanding of inertia thinking. J. Maoming Coll. **2**, 15–18 (2009)

Cognitive Modeling for Robotic Assembly/Maintenance Task in Space Exploration

Yanfei Liu[1], Zhiqiang Tian[2(✉)], Yuzhou Liu[3], Junsong Li[1], Feng Fu[1], and Jing Bian[1]

[1] Department of Computer Science and Technology, Zhejiang Sci-Tech University, No. 928, 2nd Avenue, Xiasha, Xiasha Higher Education Zone, Hangzhou 310018, China
yliu@zju.edu.cn, ljscg@sina.com, {fufeng,bj}@zstu.edu.cn

[2] National Key Laboratory of Human Factors Engineering, Astronaut Research and Training Centre of China, No. 26, Beiqing Road, Haidian District, Beijing 100094, China
tianzhiqiang2000@163.com

[3] Community College of Allegheny County, 3 Allegheny Center, Pittsburgh, PA 15212, USA
rcdl@me.com

Abstract. A cognitive control model for space assembly/maintenance task is proposed based on a double-arm robot in this paper. The two arms are designed to implement fixing assemble component and to execute assembling operations respectively. There is one hand and connecting rods articulated by three knots for each arm. The connecting knot retains the rod rotating in one plane. The arm is installed on fixed device which allow arm move with a 2 degree of freedom. By analyzing specific task, the procedural knowledge for cognitive modeling is constructed. Based on ACT-R architecture, a cognitive model for space assembly/maintenance task is constructed, and cognitive control simulation for space assembly is implemented. As an example, to replace solar panel maintenance task is chosen to investigate model's validation and cognitive procedure's rationality. The result shows that both the model's simulating cognitive process and model's running result satisfy human's cognitive characteristic.

Keywords: Space assembly/maintenance task · Cognitive modeling · Cognitive architecture · Cognitive robot · Space exploration

1 Introduction

The pace of human space exploration has been sped up since announcement of several plans for space in the 2000s. In recent past, more nations such as China and India have joined the efforts in exploring space. China even has announced plans to have a 60-ton multi-module space station in orbit by 2020.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_13

Ambitions to explore and develop space call for the assembly and servicing of diverse in-space facilities. Experience shows that past and current space assembly and maintenance operations have used the approach of extensive preflight mission planning and training to prepare the flight crews for the entire mission [1]. For astronauts boarding on space station, they are exposed to numerous stressors during spaceflights that may impair their cognitive capabilities, and even sometimes are very dangerous. Telerobotics is one of the most successful and versatile space technologies for space maintenance. Space telerobot may perform some sort of telemanipulation for spacecraft, orbital object or planet surface with more or less degree of autonomy. However, it is very hard to guarantee that a complex control task can be successful implemented due to it needs to cope with (a) large time delays, which furthermore are (b) variable, producing therefore (c) communication jitter and (d) data losses. Therefore, robots in space are expected to be capable of understanding its environment, generalizing ideas, predicating faults and working like human being to achieve task's accomplishment.

Recently, cognitive robots have been prompted by researchers to realize robotic intelligent control [2–8]. Cognitive robots are intelligent machines that are capable of performing tasks by themselves without explicit human control. They have capability of perception, information processing, decision-making and operation execution like human being. They also have sensors for perceiving various environmental factors such as relative position, posture and status. Based on its situation cognitive robot can perform some special purpose manipulations having been learned. Cognitive Robotics Systems need to reach a level of cognition that will allow them to understand and effectively operate in corresponding environments, interact with humans, and adapt their actions to an ever-growing range of situations.

The field of cognitive robotics has made significant progress in equipping robots and software agents with high-level cognitive functions. Preda et al. [2] address a complete automation of surgical robot by combining advanced sensing, cognition and control capabilities. Kelley et al. (2011) depict theoretical underpinnings of developmental change by using a cognitive architecture that implemented on a robotic system and how the theories of knowledge representation relate to critical periods of infant development. D'Mello and Franklin [8] describe computational modeling required for cognitive robotics and illustrate possible experiments. Morse et al. suggest that considerations of embodiment, situation awareness, and autonomy, intrinsic to cognitive robotics, provide an appropriate basis for the integration and theoretic accumulation. Morse et al. (2011) conclude that cognitive robotics provides a natural progression from cognitive modelling where issues of integration and autonomy lacking from cognitive modelling. Perico et al. (2016) present RoboFEI-HT Simulator developed for the RoboCup Humanoid Soccer domain, which allows cognitive algorithms to be implemented, simulated and transferred to real humanoid robots.

In addition, cognitive architecture and hybrid architecture have been becoming one of the hottest topics in cognitive modeling domain [9]. Cognitive architecture can refer to theories about the structure of the human mind. Its main goal is to summarize the various results of cognitive psychology in a comprehensive computer model. By applying cognitive architecture as a basic specification, it can provide a ready-made tool and theoretical basis for cognitive modeling, make the cognitive modeling normalizing, and achieve better results for cognitive modeling [10]. A hybrid framework developed based on different architectures can take advantage of all the relative strengths of the integrated components.

This study explores mechanical design for space assembly/maintenance robot. By analyzing task's scenario, construct cognitive model based on ACT-R cognitive architecture for a maintenance task of replacing solar panel, and implement model's simulation and verification.

2 A Cognitive Assembly/Maintenance Robot Design

Cognitive robotics is concerned with endowing a robot with intelligent behavior by providing it with a processing architecture that will allow it to learn and reason about how to behave in response to complex goals in a complex world.

In order to build a cognitive robot, the most important is to decide what it is that you want the robot to do. Is it going to be mobile or will it being able to do its tasks from one place? Depending on its functions, one will need to determine what kind of sensors it will need, what it will look like and how it will perform control task. The more complex tasks the robot will implement, the more powerful the processor have to be. To get all the parts needed to build the robot, a cognitive robot need sensors, motors and a central processor. A cognitive robot for space assembly/maintenance has its visual perception part, mechanical execution unit and motion control mechanism.

2.1 Vision Perception Design for Cognitive Robot

The vision system ensures that the robot can find the working target in the working area, locate object accurately, discover operation faults and feedback them to the decision center. Thus, vision is essential source of information while robot interacts with environment [11]. As for the studying cognitive robot, the vision sensor is two cameras installed in the head of robot, it is the eye of the cognitive robot, it aims to detect operating objects, and the behaviors' constrain of external environment.

As the robot vision sensor working, the camera captures the image from its scenario. Then the process unit transforms the image signal to symbolic signal by image process and pattern recognition algorithms so as to the system can make response to the outside vision's stimuli. In this study, the vision information will be the basic basis for cognitive model making decision through sending them to cognitive model.

As a complex process, human vision involves image acquisition with the eye, interpretations and decision making in different regions in the brain. Vision perception for a cognition model is to implement the functions of human vision system. Thus, the main function for a cognitive model is to capture the scenario information and make it the requirement of decision-making.

ACT-R designs the vision module to provide model with information about what seen in current vision device and provides the model with a visual attention system. There are two subsystems which comprise ACT-R's vision module with "where" and "what" knowledge. The two subsystems work together, but each retains its own buffer and accepts specific requests. The important thing is what the vision module does not model eye's movements in reality. It is only a representation of visual attention abstracted away from what is occurring with the eye sensor. Providentially, the robot's eye provides what is occurring in the visual attention, and ACT-R then makes an abstraction of robot's visual attention as the vision perception for cognitive model.

2.2 Mechanical Design

6R (six-revolute) robot is a flexible high-tech industrial robot with multi-joint, multi-degree of freedom and characterized by a wide range of applications, changeable action pattern, and easy control. The robot designed for space assembly/maintenance is equipped with two arms, and there are two connecting rods articulated by knot for each arm. One end of the arm links to robot's body and the other end concatenates robot's hand both by knots. The connecting knot retains the rod rotating in two planes within 360-degree limitation. The function of the left hand is to hold assembly/maintenance part and the right hand is designed to implement operations. The operations of the right hand are designed now to tighten and to remove bolt with screwdriver in the case studying example task of replacing solar panel for spaceflight. Figure 1 shows the basic mechanism and mechanical structure of a 6R robot for the space assembly/maintenance task.

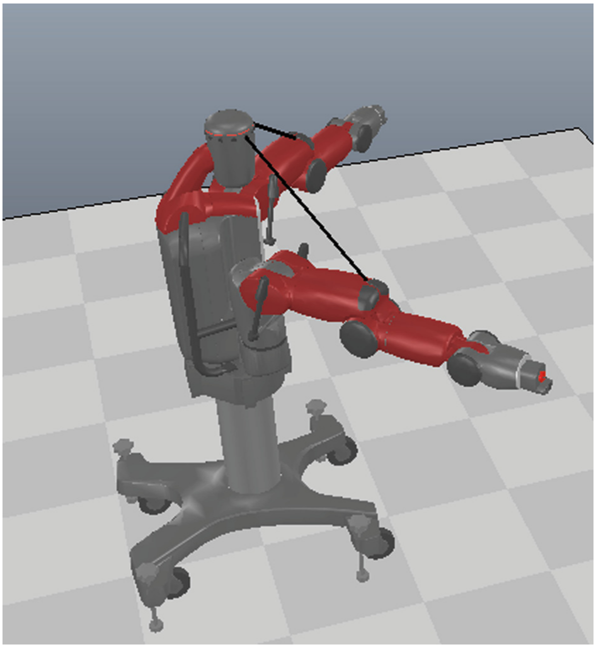


Fig. 1. The basic mechanism and mechanical structure of a 6R robot for the space assembly/maintenance task.

2.3 Motion Control Algorithm

The inverse kinematics problem (IKP) of the 6R robot manipulators has been discussed widely [12]. The IKP for 6R robot manipulator is to determine the joint values given the position and orientation of the end-effector relative to the base and the values of all link parameters. Problem solving of the inverse kinematics is complicated than forward kinematics, and there may be uncertain solutions or no solution, however it is more common by comparing to forward kinematics problem in the real engineering. Therefore, the IKP has been recognized as a more important problem for robot workspace analysis, trajectory planning, motion control and off-line programming. Figure 2 shows the basic mechanism and mechanical structure of space assembly/maintenance robot.

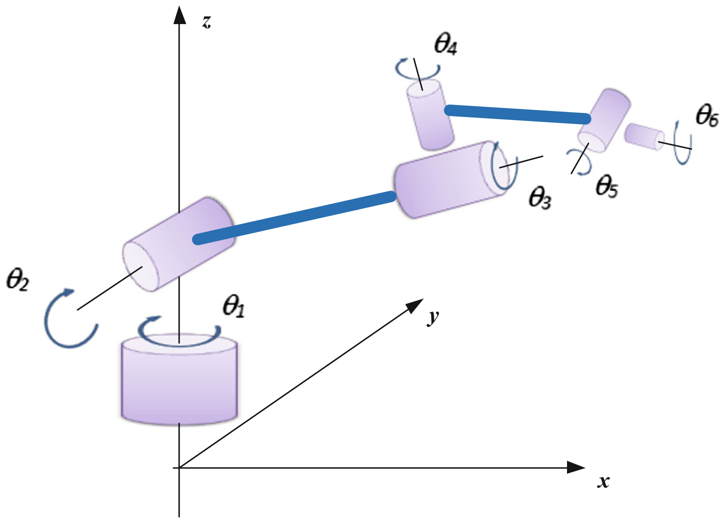


Fig. 2. The basic mechanism and mechanical structure of space assembly/maintenance robot.

During the movement of the robot arm, it is necessary to constantly change the rotation angle of each knots so that the robotic hand reaches the corresponding position. If let the coordinates of a point in a three-dimensional space be (x, y, z), then manipulator rotating angle corresponding to the six joints are (θ_1, θ_2, θ_3, θ_4, θ_5, θ_6) respectively, thus the relationship between (x, y, z) and (θ_1, θ_2, θ_3, θ_4, θ_5, θ_6) are manipulator movement model. If the manipulator's initial rotating angle and posture are known, then the manipulator's corresponding position and posture can be obtained by matrix transformation.

As considering the constraints of the knots with mechanism structure, one can get the moving range corresponding to each knot rotating in angle. In order to simplify operation and for convenience of robot removal/mounting parts, the study let θ_5 of the left arm be 0° and θ_5 of the right arm be 90°, to let the replacement part be parallel and the screwdriver be vertical to the *xoy* plane. By joining this constraint, it can ensure the IKP obtain a unique solution. Solution about IKP of a general 6R manipulator has been reported in many literatures and it is unnecessary to go into details here.

2.4 Cognitive Architecture of Assembly/Maintenance Robot

There is a debate in cognitive science between architectures that are called symbolic and those that are called connectionist. Many researchers of the larger cognitive science community tend to regard ACT-R as an instance of symbolic architecture. However, like all theories, both symbolic and connectionist architecture has advantage and disadvantage. For Addition, most researchers nowadays think that the facts of human cognition are neither completely symbolic nor fully connectionist. It is argued that some circumstance it may be symbolic and sometimes it may be connectionist, even most cases are both or hybrid and more. A concept of hybrid cognitive architecture was introduced to practical applications recently, For example, Christian Lebiere et al. proposed a hybrid cognitive architecture SAL (Synthesis of ACT-R and Leabra) which is a integration of two theories of cognitive functioning, ACT-R a predominantly symbolic production-rule based architecture and Leabra a neural modeling architecture.

For the mainly related technics for the visual attention of the robot are image processing and pattern recognition, and studies on artificial neural networks being used in image recognition and classification have been for a long time. It demonstrates that neural networks are a powerful technology for classification of visual inputs. Recently, research on convolutional neural network (CNN) has become a new hottest topic in artificial intelligent, and it achieved great success in classification of high-resolution images with [13]. Hence, this study uses CNN to be the vision recognition algorithm and ACT-R to be the production system for robot's cognition processing. Figure 3 shows a combined architecture of ACT-R and CNN for assembly/maintenance robot.

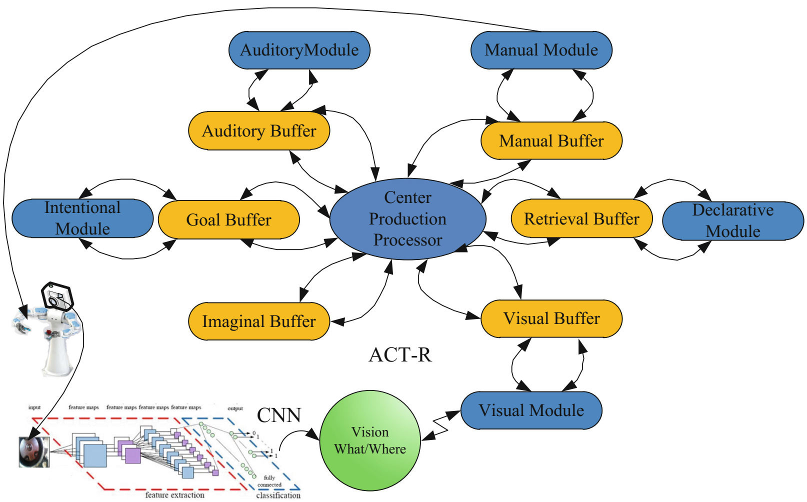


Fig. 3. A combined hybrid cognitive architecture for space assembly/maintenance robot.

CNN is a kind of deep learning algorithm that specifically designed for image recognition and classification based on multi-layer neural network. The neurons with learnable weights and biases form a CNN. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. In addition, the whole network still have a loss function on the last layer and all the tips/tricks we developed for learning regular Neural Networks still apply. CNN architectures make the explicit assumption that the inputs are images, which allows algorithm to encode certain properties into the architecture. These then make the forward function more efficient to implement and vastly reduce the amount of parameters in the network.

By using a structure distinct from artificial neural network (ANN) for image recognition, it can achieve a higher speed for CNN training. These make it easy for adopting multi-layer neural network, while which is benefit to accuracy of classification for multi-layer structure. Through making the output of CNN as its input, ACT-R makes response to the vision perception and conducts its vision module's work. The combination of the two architectures allows for rich dynamics that take advantage of neural and symbolic aspects of cognition and provides mutual constraints that promote convergence towards models that are both neurophysiologically and psychologically valid.

3 Cognitive Model for Assembly/Maintenance Robot

In ACT-R, there are two different categories of long-term memory: declarative and procedural. Declarative memory consists of facts and procedural memory consists of our knowledge of how to do things. ACT-R represents declarative knowledge using chunks and represents procedural knowledge by adopting productions that are collections of production rules. ACT-R defines syntax to represent chunks and productions. An ACT-R model is a computer file written in the LISP programming language, and can be executed on ACT-R platform.

3.1 Procedural Knowledge

Taking the task of replacing solar panel an example, the procedural knowledge is as follows.

- Start an assembly/maintenance task
  - Task 1: Disassemble the old solar panel
    - Right hand takes screwdriver
    - Left hand holds the old solar panel
    - Vision scan to find the screw position
    - Right hand places the screwdriver in screw
    - Right hand removes screw and brings back screw
    - Make sure all screws removed
    - Left hand takes back the old solar panel
  - Task 2: Locate the new one solar panel
    - The left hand grasps the new solar panel
    - The left hand puts back the new solar panel to the old one's place
    - The left hand aligns the new solar panel
    - The robotic vision check is at its right position for installation

  - Task 3: Assemble the new solar panel
    - The right hand takes a screw
    - The vision scans the right position for screw installation
    - The left hand locates the right place for screw installation
    - Right hand places the screw in screw hole
    - Right hand takes screwdriver
    - Right hand tighten the screw
    - Check all the screws being installed
- Finish assembly/maintenance task

3.2 Cognitive Model

An ACT-R model is a text file written in Lisp [14]. A typical model file will have the following structure:

```
(clear-all) ; reset ACT-R's state to a clean state
{Lisp functions for presenting an experiment, data col-
lection or other support needs}
(define-model robot-assembly ; model's name goes here
    (sgp {parameter value}*) ; set parameters
    {chunk-type definitions} ; declare chunk type
    {initial chunks are defined}; chunks goes here
    {productions are specified} ; production rules
    {any additional model set-up commands} ; first goal
    {additional model parameter settings} ; others
)
```

The ACT-R commands appeared in ACT-R model and the model components can reference to ACT-R *reference manual*.

3.3 Cognitive Control for Space Assembly/Maintenance Task

This study implements robotic assembly/maintenance simulation over the V-REP platform and the robotic cognitive model runs on ACT-R cognitive architecture. The camera installed in robot captures the image of its view scope, and by image recognition (CNN), the image is converted to robotic vision information. The visual information is sent to the cognitive model as vision perception for cognitive model, and finally the model is responsible for making decisions. During the model's run, the ACT-R software details the cognitive procedural vividly, such as visual location, conflict resolution, production fire, and motor execution etc. The motor part of the model running results are transferred to the simulation platform and the simulation side performs controls of the motor behaviors from the cognitive model side.

4 Model Verification and Control Simulation

The model's validation is at two stages. The one is model implement simulation space assembly/maintenance task driven by cognitive model and the other is to investigate the cognitive details, which those of the model are coincidence with the true control.

4.1 Task Accomplishment Verification and Simulation

The robot simulator V-REP, with integrated development environment, is based on a distributed control architecture: each object/model can be individually controlled via an embedded script, a plugin, a ROS (robot operating system) node, a remote API client, or a custom solution. This makes V-REP very versatile and ideal for multi-robot applications. V-REP is used for fast algorithm development, factory automation simulations, fast prototyping and verification, robotics related education, remote monitoring, safety double-checking, etc. Controllers can be written in C/C++, Python, Java, Lua, Matlab or Octave. Therefore, a robot and related components for space assembly/ maintenance task are implemented over the V-REP platform.

For ACT-R has been used successfully to modeling a variety of behavioral phenomena and has proven particularly successful at modeling tasks, this study employ ACT-R cognitive architecture as modeling tool for cognitive model.

The simulation of robot assembly/maintenance task and cognitive model are both run on the computer at the same time, and the simulation detects the motor command from cognitive architecture and model check the vision perception from robot simulation side. By communicating among robot simulation and running cognitive model, the cognitive-model-based assemble/maintenance control task is accomplished.

4.2 Cognitive Details Cognitive Processes Verification

As the model run on ACT-R, the cognitive processes can be displayed in every 50 ms cognitive periods. This study contrasts the processes to human control behavior. The comparison for a fragment shows in Table 1.

Table 1. The processes comparison of model running and human behaviors.

Time	Model	Time	Human control behavior
0.000	SET-BUFFER-CHUNK GOAL	0.00	Vision focus on screw hole
0.000	VISUAL-LOCATION REQUESTED		/
0.050	PRODUCTION-FIRED FIND…		/
…	…	…	…
0.735	PRODUCTION-FIRED RESPOND…	0.74	Take screwdriver
…	…	…	…

This table only shows fragment of cognitive process, however the entire comparisons indicates that the model process is consistent with that of human behavior.

5 Conclusions

This paper bring cognitive model to robotic control to realize the robot's automation control for telemanipulation of spacecraft. A hybrid cognitive architecture mixed symbolic system and neural network is proposed, a 6R robotic arms for space assemble/maintenance is applied, the motion algorithm is discussed. Through analyzing specific task, the procedural knowledge it is necessary for cognitive model are built. Based on ACT-R cognitive architecture the assembly/maintenance cognitive model is constructed. By using simulation, a robot for replacing solar panel task is designed. Depending on vision perception from robot and model's goal, the cognitive model makes decision according to designed cognitive behavior. In the same time, by obtaining the output of the model decision the robot performs the corresponding operations–motor behavior. From two stages which one by accomplishing task and another by details cognitive processes the model's validation is verified.

The simulation result shows that by applying cognitive model to create a cognitive robot is feasible and cognitive robot can perform some human intelligent control especially under the circumstance of time delay exists of remote control. The model's verification at the stage of cognitive details indicates that cognitive model based ACT-R cognitive architecture reflects the human cognition.

Acknowledgments. This work is supported by the China Advanced Space Medico-Engineering Research Project (No. 2012SY54A1705) and the Zhejiang Provincial Natural Science Foundation (No. LY12C09005).

References

1. Rembala, R., Ower, C.: Robotic assembly and maintenance of future space stations based on the ISS mission operations experience. Acta Astronaut. **65**(7), 912–920 (2009)
2. Preda, N., Ferraguti, F., De Rossi, G., Secchi, C., Muradore, R., Fiorini, P., Bonfè, M.: A cognitive robot control architecture for autonomous execution of surgical tasks. J. Med. Robot. Res. **1**(04), 1650008 (2016)
3. Nalpantidis, L., Detry, R., Damen, D., Bleser, G., Cakmak, M., Erden, M.S.: Cognitive robotics systems-concepts and applications. J. Intell. Robot. Syst. **80**(Supplement-1), 3–5 (2015)
4. Ho, S.B.: Cognitive architecture for adaptive social robotics. In: Kubota, N., Kiguchi, K., Liu, H., Obo, T. (eds.) International Conference on Intelligent Robotics and Applications, 549–562. Springer International Publishing, Heidelberg (2016)
5. Maniadakis, M., Trahanias, P.: Modelling robotic cognitive mechanisms by hierarchical cooperative CoEvolution. Int. J. Artif. Intell. Tools **16**(06), 935–966 (2007)
6. Aly, A., Griffiths, S., Stramandinoli, F.: Towards Intelligent Social Robots: Current Advances in Cognitive Robotics (2016)
7. Kurup, U., Lebiere, C.: What can cognitive architectures do for robotics? Biologically Inspired Cogn. Architectures **2**, 88–99 (2012)
8. D'Mello, S., Franklin, S.: Ccomputational modeling/cognitive robotics complements functional modeling/experimental psychology. New Ideas Psychol. **29**(3), 217–227 (2011)

9. Jilk, D.J., Lebiere, C., O'Reilly, R.C., Anderson, J.R.: SAL: an explicitly pluralistic cognitive architecture. J. Exp. Theor. Artif. Intell. **20**(3), 197–218 (2008)
10. Liu, Y., Tian, Z., Liu, Y., Li, J., Fu, F.: Cognitive architecture based platform on human performance evaluation for space manual control task. In: Hale, K., Stanney, K. (eds.) Advances in Neuroergonomics and Cognitive Engineering, pp. 303–314. Springer International Publishing, Heidelberg (2017)
11. Bittermann, M.S., Sariyildiz, I.S., Ciftcioglu, Ö.: Visual perception in design and robotics. Integr. Comput.-Aided Eng. **14**(1), 73–91 (2007)
12. Fu, Z., Yang, W., Yang, Z.: Solution of inverse kinematics for 6R robot manipulators with offset wrist based on geometric algebra. J. Mech. Robot. **5**(3), 031010 (2013)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105. (2012)
14. Anderson, J.R.: How Can the Human Mind Occur in the Physical Universe?. Oxford University Press, Oxford (2009)

Vision and Memory

Analysis of Relationship Between Impression of Video and Memory Using fNIRS

Suguru Azehara[1(✉)], Toru Nakata[2], and Toshikazu Kato[1]

[1] Chuo University, 1-13-27, Kasuga, Bunkyou-ku, Tokyo, Japan
azeharu0521@gmail.com
[2] Industrial and Systems Engineering, 2-3-26, Aomi, Koto-ku, Tokyo, Japan

Abstract. Experiment was conducted on the memory in the impression of video. I used a TV commercial (TVCM) to use it as an video. Based on the author's previous research, we conducted experiments using NIRS this time. Focusing on the impression of TVCM, the presence or absence of memory fixation in impression was measured by using a task of removing visual/audio infor-mation other than impression factors. As a result, impression and suggestion of brain region related to memory was obtained.

Keywords: Video · Impression · Memory · fNIRS

1 Introduction

Commercials on television is one of the most famous forms of advertisements. It is used to promote products, services, corporation, and brand. It is taken purchasing behavior to audience. Memory of the contents of advertisements triggers purchasing behavior. It is the most important element of an advertisement.

In addition, TVCM is an advertisement including many elements. For example, various factors such as music, performers, length of broadcasting time, etc. can be considered. In this experiment, we focus on the impression of TVCM that comprehends them, and clarify what impression of TVCM remains in memory.

Many of the conventional memory research is based on questionnaires. However, there are many subjective factors in the questionnaire, and it is difficult to judge whether you actually remember it. Therefore, in this experiment, change in cerebral blood flow is used as an objective index.

In this study, we watch subjects to watch TVCM and find relationships between good and bad impressions and memory from the viewpoint of brain science.

2 Previous Research

Yoshida's research on TVCM's visual method and memory relationship has been studied. Yoshida has been studying expressive techniques to raise recall of TVCM from the viewpoint of producers [1]. Yoshida hypothesized that expression method has an influence on the memory of the consumer. He revealed that basic expressions such as "Fade in", "Fix", etc. increase the recall under certain conditions.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_14

However, the TVCM used in Yoshida's experiment was not actually broadcasted but was made by himself. The memory test was done just after viewing, and it was not mentioned during the period when memory was established.

Also, as a previous research, the authors focused on "Repeated calling", "Number of cuts", and "Frequent appearances" of TVCM as a whole, which are elements of TVCM, and experimented as to which TVCM containing many elements remained in memory.

In this experiment, it was expected that forgetting will proceed by leaving a period after watching TVCM. To find a TVCM that remains in memory even while forgetting goes on, a three-step model of memory proposed by Hayashi [2]. As a three-step model of memory, "Recognition", "cued recall", "free recall" existed in order of weakness of fixation, and the degree of fixation of memory in three stages was measured. At that time, analysis was performed using the index of information transmission efficiency, and a graph of the result is shown in Fig. 1.

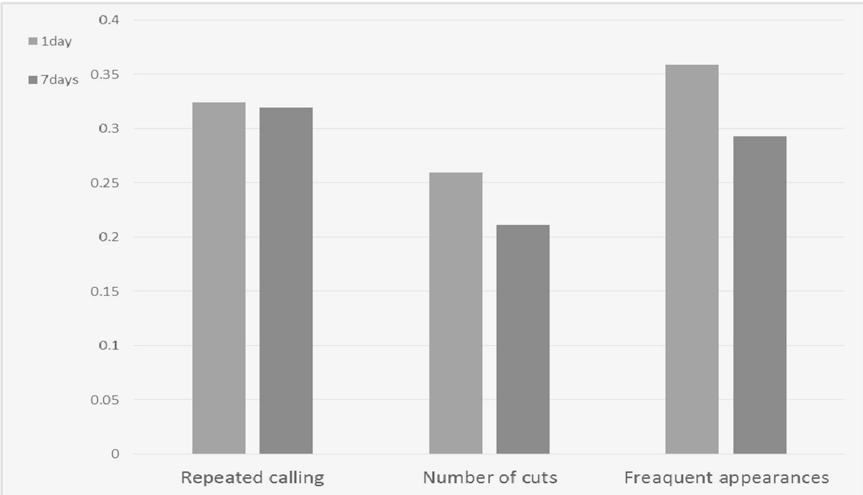


Fig. 1. Information transmission efficiency of "Repeated calling", "Number of cuts" and "Frequent appearances". Blue: 1 day span. Orange: 7 days span.

As shown in Fig. 1, as an effective element in short-term memory, frequent appearance times of the product is large, and the degree of recognition of memory after one day was stronger than the other. Effective elements in long-term memory, it is understood that those with a "number of calling" have a long-term memory of consolidation in terms of the fact that the value of information transmission efficiency has not changed. In addition to using information transmission efficiency, in addition to describing what was left in memory by questionnaire. As a result, there were many descriptions of TVCM's music with a large "Repeated calling". Therefore, it was thought that music, music made for TVCM among them, impacted the audience and was easy to set in memory.

In this research, we study what kind of factors of TVCM affect memory. TVCM used in the experiment is actually broadcasted, and in the previous experiment indexes such as information transmission efficiency are for. A questionnaire was used as the main means of the experiment. The questionnaire will be answered after thinking in once, so there is a fear that you will unintentionally answer differently from your true thought. Therefore, we think that it is difficult to judge viewer's latent memory only by the questionnaire. In order to solve this problem, in this experiment, using the gaze and brain activity without depending on the questionnaire. We will analyze recollection of memories concerning TVCM by finding elements that are established in TVCM memory. Throughout the experiment, and will support it in the future TVCM production.

3 Measuring Equipment

We used the ETG-4000 optical topography system (HITACHI Medical Corporation, Japan) to measure brain activity in the frontal cortex. The ETG-4000 system allows for neuroimaging based on near infrared spectroscopy; it measures the change in the oxygen levels in cerebral tissue [3]. Light of two different wavelengths—695 nm and 830 nm—is incident on the scalp through optical fibers and detected every 0.1 s.

4 Preliminary Experiment

4.1 TVCM Used in Experiment

Because this experiment is about memory, we do not use TVCM which I have seen in the past. TVCM to be used was broadcasted before subjects were born, and those that have never been viewed have been selected. Also, the length of TVCM was unified in 30 s [4].

4.2 Impression Evaluation

In order to select the TVCM used in this experiment, we had TVCM impression assessment as a preliminary experiment. The subjects were 15 college students in twenties (12 men, 3 females, average age 22.93 years old). I asked the TVCM (Table 1) prepared for viewing and evaluated likes/dislikes in 7 stages of −3 to 3. The experimental flow is shown in Fig. 2 below.

Table 1. TVCM list used in preliminary experiment.

TVCM number	Corporation
1	Panasonic Corporation
2	Tokio Marine & Nichido
3	Fuji Film
4	Sony Corporation
5	Sony Corporation
6	Lotte
7	Ajinomoto
8	Panasonic Corporarion
9	Panasonic Corporation
10	Shiseido Company
11	Shiseido Company
12	Ryukakusan Co., Ltd.
13	Nissin Food Products Co,. Ltd.
14	Lion Corporation
15	Nikka Whisky
16	Renown Incorporated
17	Mftlji
18	Nissin Food Products Co.. Ltd
19	ROHTO Pharmaceutical Co., Ltd.
20	Sony Corporation
21	Panasonic Corporation
22	Yamaha Corporation

Fig. 2. Flow of preliminary experiment.

Z-score was calculated from the calculated score, and clustering by the k-means method was performed. After processing, three groups with “Excellent”, “Normal”, “Poor”, TVCM were selected for each group from the value of Z-score, and TVCM group by impression used for this experiment was created. The result of clustering Z-score by the k-means method is shown in Fig. 3 below, and the TVCM group by impression used in this experiment is shown in Table 2 below.

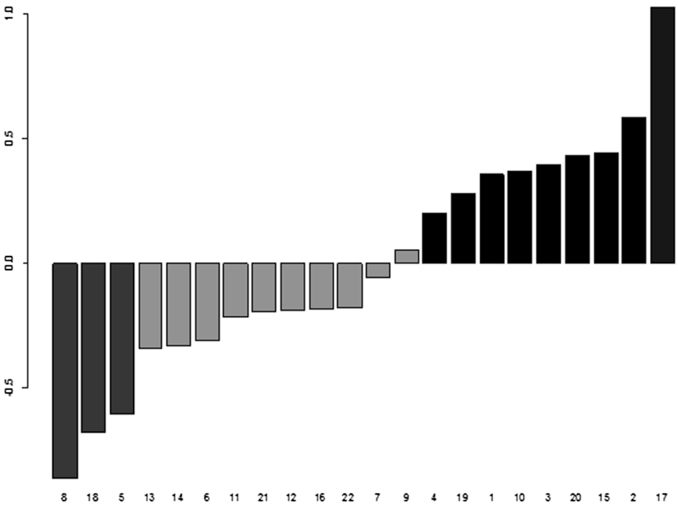


Fig. 3. Z-score(mean) of all subjects and glaph by method of k-means. I selected TVCM used experiment. “Excellent”, “Normal”, “Poor” groups.

Table 2. TVCM group by impression. TVCM number of 17 is as a outliers.

	VCM Njmsor	Corporation	Z-scoro
Excellent	2	Tokio Marine & Nichido	0.394G0G
	15	Nikka Whisky	0.445012
	20	Sony Corporation	0.432881
IMcriral	7	Ajinomato	−0.0593b
	9	Panasonic Corporation	0.C53378
	22	Yamaha Corporation	−0.1/9
Poor	5	Sony Corporation	−0.60075
	8	Panasonic Corporation	−0.86174
	18	Miss n Food Products Co., Ltd.	−0.67747

5 Method of Experiment

5.1 Subjects

Subjects were 11 healthy 20-year college students (8 males, 3 females, average age 23.8 years old). All subjects were right-handed.

5.2 Flow of Experiment

We showed TVCM have selected in 4.2 to all subjects. Ask them to wear NIRS during viewing and make sure the brain activates at which impression from the change in the cerebral blood flow at that time. After viewing, we set up a period of seven days and conducted playback tests as to whether we remember the TVCM actually watched. In the playback test, I showed all TVCM in Table 1 of and confirmed whether I remember it actually. The overall flow of experiment is shown in Fig. 4 below.

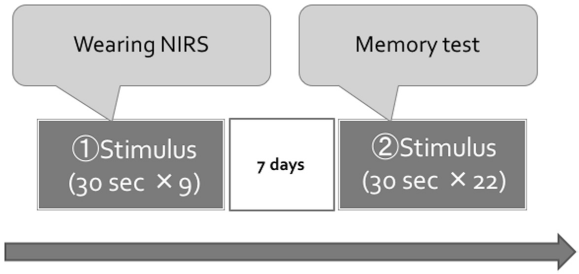


Fig. 4. Overall flow of experiment. At first, subjects watch TVCM. Finally, they take memory test.

5.3 Presentation Stimulus

In the experiment of this time, the objective is to ascertain the presence or absence of memory retention from the viewpoint of brain science using NIRS, due to the difference in impression of TVCM. As mentioned above, TVCM includes visual and audio information, and it is not known whether the impression has actually affected the memory at the time of measurement. Therefore, work was done to remove elements other than impressions.

Elimination of Visual Information. As the elimination of visual information, we established a rest between TVCM main contents. As a rest, the same mosaic animation was created for RGB values of TVCM main volume. Cut the main part into 900 images, and create a mosaic image by arranging each pixel randomly for all the images. After that, we joined 900 mosaic images and made it a 30 s movie. An example of the created mosaic image is shown in Fig. 5 below.

Elimination of Audio Information. As the elimination of the voice information, the voice of the reverse reproduction of the main part voice was created. Then, by matching the mosaic animation created in 4.3.2, it was made to exclude visual and sound. Let this be the control task (CT) of visual and audio information.

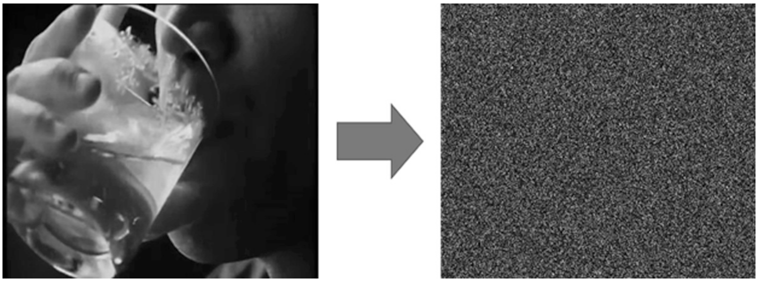


Fig. 5. Created a mosaic image by arranging each pixel randomly for all the images.

Detail of Experiment. Experiment participants were seated in the chair and the distance to the display where the stimulus was presented was unified to 600 mm. Instructions were carried out before the start of measurement, and experiments were carried out after participants fully understood the contents of the experiment.

We prepared a movie with gaze point around TVCM main story, then asked to watch CT created in 4.3.1 and 4.3.2. Detailed experimental flow is shown in Fig. 6 below.

Fig. 6. Subjects watch fixation, TVCM, fixation, CT videos in turn.

Experiments were carried out with the flow as in the above experimental flow. The gazing point, TVCM, and CT were 30 s each. Since nine movies were watched, the above experimental flow was repeated nine times (i = 1 to 9).

After viewing, after establishing a period of seven days, have TVCM of Table 1 all watched and listen to what subjects remember and remember what I do not remember. We measure the retention of memory from the response rate at that time.

6 Result of Experiment

6.1 Brain Activity of Each Impression

For the data used for analysis this time series data of TVCM section and CT section including 25 s of gazing point moving image before stimulus presentation and after 5 s is used (Fig. 7). The difference between the TVCM section and the CT section was taken and the t test was conducted using the difference between the stimulation section and the average value in the rest section.

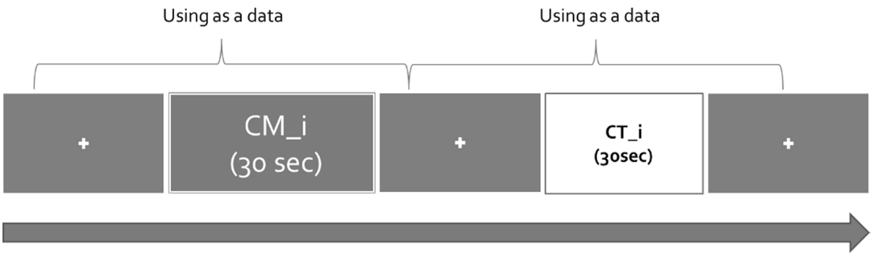


Fig. 7. The data used for analysis

In this time, t test was conducted from each data of impression (Excellent, Normal, Poor). As a result, when watching TVCM with impressions, significant differences were found in brains 37 and 48 ch ($p < 0.05$ *).

Next, at the time of Nomal, there is a significant difference in 12 ch, and the site here is a site called a broca field. It is said that Broca is involved in understanding language processing and spoken language.

Finally, at the time of Poor there were no significant difference channels (Table 3).

Table 3. Channels with significant difference by impression

Impression	$p < 0.05$
Excellent	37 ch, 48 ch
Normal	12 ch
Poor	

6.2 Memory Test

Whether you remember TVCM as seen at the time of the experiment, we made a period of 7 days after the experiment and conducted a regeneration test. As a result, the correct answer rate of TVCM in Excellent group and Poor group was 100% in all subjects. However, the correct answer rate of the Normal group TVCM was about 90%.

7 Consideration

The brain activity of the Excellent group with high impression evaluation of TVCM showed a significant difference in t test at 37 ch and 48 ch. It was found that the part here is very close to the part considered to be related to recollection of impression [5]. From this experiment it seemed that 37 ch, 48 ch may be involved in impression recall. In the Normal group, only significant difference was found in 12 ch. As described above, 12 ch is a site called broker field. If the impression of TVCM is Normal, it seems that the contents of the TVCM are processed linguistically. That is, I thought that I was trying to understand the contents of TVCM. In the Poor group there was no significant difference channel. From this, it is thought that people are not interested in TVCM with bad impression.

From the result of the regeneration test, the correct answer rate was not 100% for only the Normal group TVCM. As mentioned above, there is a possibility that the Normal group's TVCM may be processing with language, so there is a possibility that it was canceled by the linguistic activity in the daily life in 7 days with the period.

8 Summary

In the experiment of this time, when watching TVCM by impression, it was found that the part with good impression responded to the part related to impression recall, and the one with ordinary impression was broker field responding. In the regeneration test, it was found that the TVCM group in which the impression is normal is not remembered in comparison with other impression groups.

In the analysis conducted this time, t test could be performed only in each impression group. In the future, each factor is analyzed by performing multiple comparison of three factors. And by using NIRS also during storage test, we clarify which part is related to memory.

References

1. Smith, T.F., Waterman, M.S.: The factors to improve commodity favorable rating in TV advertising. Art Eng. J., 129–136 (2010)
2. Hayashi, M., Ohta, N.: A comparison of three hypermnesia tasks employing remember/know judgments: recognition, word-stem recall and free recall. Tsukuba Psychol. Res. **26**, 39–51 (2003)
3. Villringer, A., Planck, J., Hock, C., Schleinkofer, L., Dirnagl, U.: Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neurosci. Lett. **154**, 101–104 (1993)
4. Muto, T., Tsumori, S.: Mou ichido mitai nihon no CM 50 nenn (Japanese TV commercial in 50 years which we want to watch once again). Avex Entertainment Inc. (2010)
5. MItsui, S.: Analysis between First Impression of Likability and Change of Oxy-Hb in Frontal Cortex during Watching Video (2012)

Evaluate Fatigue of Blue Light Influence on General LCD, Low Blue Light LCD and OLED Displays

Yunhong Zhang[1](✉), Yingbao Yang[2], Yilin Chen[2], Yong Yang[2], and Yu Chao[3]

[1] AQSIQ Key Laboratory of Human Factor and Ergonomics, China National Institute of Standardization, Beijing, China
zhangyh@cnis.gov.cn

[2] China Star Optoelectronics Technology Co., Ltd, Wuhan, China

[3] Academy of Psychology and Behavior, Tianjin Normal University, Tianjin, China

Abstract. Visual search tasks are mainly test methods of user performances for smartphone displays, which are usually used to assess the subjective quality of the visual display terminal. This study investigated the effect of different smartphone displays on visual search performance. A within-subject factorial design was used in this experiment. The visual search contents were Landolt Ring visual search performance, the search target of which was a ring with a gap in one of the four directions (up, down, right, left). The experiment was carried out in three smartphone displays. The search time and accuracy of each participant were recorded. The difference in search performance and subjective experiences among different smartphone displays was significant. Post hoc comparisons found that the search performance and the subjective experiences under the OLED was significantly better than the other smartphone displays, the low blue light LCD was second and he general LCD was the worst. These results revealed that visual search performance and the subjective experience were sensitive to different smartphones. The obtained results could be a reference to decide the visual search efficiency on different smartphone displays.

Keywords: Visual fatigue · Comfort · Visual research · Low blue light LCD · OLED display

1 Introduction

With the rapid development of science and technology, electronic products have spread to every aspect of our lives and human society has begun to enter the “screen time”. The time people facing screen in daily work and life becames longer and the dependence on it becomes higher. But using these electronic products for a long time can lead to eye fatigue and physical discomfort, which includes dry eyes, blurred vision, sense of tension and tear of eyes, head, neck and back pain and other symptoms. Asthenopia becomes a big threat to people’s health. Therefore, improving the comfort of the display has become an important problem that many manufacturers focus on.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_15

The damage and influence to the eyes caused by the high-energy visible light in display particularly attracts manufacturers and researchers' attention among them. The wavelength of this kind high energy visible light is 400–500 nm and its light frequency is slightly lower than ultraviolet ray. The energy of it is in the highest part of the visible light and its color is blue and purple, which is known as "blue light". Many manufacturers directly increased the intensity of the blue light to improve the display brightness, which caused the problem of "blue-ray excess", as the increase of blue light could improve the brightness of LCD/LED white light.

Berson et al. from Brown University found that the third type of photoreceptor cells (ipRGC) of eye retina could transmit signal to the brain's biological clock regulator - the suprachiasmatic nucleus (SCN), thus helping the human regulate circadian rhythms and other biological effects, including the change of human vital signs (blood pressure, pulse, blood oxygen, temperature, etc.), hormone secretion and the degree of alertness and excitement of the body [1]. YasuKouchi et al. studied that ipRGC accepted light signal and sent it to the brain's hypothalamus suprachiasmatic nucleus (SCN), which controls the pineal gland of the hypothalamus that secretes certain hormones, realizing the regulation of circadian rhythms and hormone control [2]. Hormones biologically regulated and affected by the light are mainly melatonin and cortisol. Melatonin is a hormone of sleep. The increase of its content in the blood makes people tired and drowsy and the decrease makes people mental excited. Cortisol was stress hormone, which could provide people with energy, make people focused and enhance immunity. People would be tired and less efficient when cortisol was at a high level for a long time. It was generally recognized that the wavelength range of light which had a significant impact on the human body rhythms was between 420–420 nm [1]. Cajochen et al. from the University of Basel, Switzerland, showed that the high-energy visible light with wavelength of 400–500 nm from the display was the main light source that harmed the human eye [3]. The light of 460 nm increases body temperature and speeded up heart rate. In 2012, Yan-dan Lin et al. from Fudan University found low chroma LED had an obvious effect on heart rate [4]. A study suggested that blocking short-wavelength light could reduce eye fatigue, and the higher the blocking effect of the lens, the lower the reduction in the critical flicker frequency [5].

There were many display products trying to reduce blue light influence, for example, the "Night Shift Mode" in iPhone, which was one of the most salient examples tries to reduce blue light exposure before going to sleep. Organic light-emitting diode (OLED) was another more and more popular display technology which could self-luminous. Its peak blue light wavelength often falled out the harmful blue light range. But it had not been studied with low blue light LCD products simultaneously on the harm to human vision.

In this study, the comparison of visual fatigue and comfort between general LCD, low blue light LCD and OLED was investigated. We expected that the low blue light LCD and OLED could make participants more comfortable and reduce visual fatigue after long time exposing under phone screen.

2 Method

2.1 Design

This experiment was to test the subjective and objective differences of visual search task performance using mobile phone displays controlled by different blue light (general LCD, low blue light LCD and OLED displays) under indoor light uniform conditions, and to examine the sensitivity of the test index. This experiment was a two-factor within-subject design. Two factors were different screen samples and time factors. Samples contain three different conditions: general LCD, low blue light LCD and OLED displays, respectively. Time factor was before and after watching. Dependent variables were complete performance, visual fatigue, emotional response and comfort under different conditions of the task.

2.2 Participants

Twenty ordinary adults from 21 to 26 years old (10 male and 10 female, mean age = 24, standard deviation of age = 1.2) were recruited and paid to participate in the experiment. All of them had more than 4.8 normal or corrected-to-normal visual acuities and healthy physical conditions. They did not have any history of neurological and mental or ophthalmic diseases.

2.3 Materials

The samples were a Normal Blue display phone, a Low Blue display phone for a company and a commercially available OLED display phone. The brightness of the three screens was controlled at 68 nit.

2.4 Apparatus

The research used Standard logarithmic visual acuity chart developed by the eye hospital of WMU [6]. Tests carried out in accordance with the standard logarithmic far end requirements, the BD-II-118 type critical fusion frequency meter (test light intensity was 1, bright black ratio was 3:1, background light was 1:16, background light was 1:16) independently developed by APTECH, single lead electrode EEG equipment to record subjects EEG indicators in the experimental task process and Nuerosky's chip and relaxation and tension algorithm to reflect the degree of brain stress and emotional stress. The visual fatigue scale, which was developed by internationally recognized James E. Sheedy, was used to test the visual fatigue, including eye fatigue (such as eye burning sensation, eye pain, eye tightness, eye irritation, tearing eyes, blurred vision, ghosting and eye dryness, etc.), self-rating test, overall comfort, color comfort and satisfaction of the overall display. The overall situation score of interviewed questions in the subjective comfort, color comfort, clarity was recorded as the best 1, and the worst 3.

2.5 Procedures

Experiments were conducted in a quiet laboratory environment that low light indoor mobile phone could be used (experimental environment illumination value was 50 lx). First, participants signed the informed consent and completed a general demographic information survey. The experimenter introduced the experimental task including the critical fusion frequency test and visual search task. The experimenter asked the subjects to practice these tasks until they were very skilled to do these tasks. The visual search task was to search for the open ring in the closed rings, which the opening direction includes the upper, lower, left and right directions. The task required subjects to find the opening target as fast as possible and select its opening direction. The critical fusion frequency test was applied before and after the search task. And the entire search task need participants to wear EEG to record the EEG changes. The task was done at the Tobbi Eye Movement Stent. The visual search task lasted for 50 min for each sample. The test sequence of samples in each subject was balanced by Latin square method. After completing the test of each sample, the subjects were asked to fill in the visual fatigue subjective perception scale and the comfort survey scale. The subjective interviewed on the comfort and other aspects of all the samples were applied in the end of the whole test. There was 20 to 30 min rest between each two tests to avoid visual fatigue. After completing the experiment, the subjects would get a certain reward.

2.6 Data Analysis

The changes value of visual search and visual fatigue data were analyzed by IBM SPSS 20 Statistics software (IBM-SPSS Inc. Chicago, IL). The method of repeated-measure ANOVA analysis and test were applied to the experiment data. The experiment compared the differences of 3 factors from 3 quantitative indicators, namely: search time, response accuracy and subjective report data. In the subsequent sections, P was test significance. When the significance standard was 0.05, P was less than 0.05, the results were obvious, if not they were not.

3 Results and Analysis

3.1 Comparison of Visual Search Results

The results showed that the performance of the task was the best under the condition of OLED, and the performance of the task was the worst under the condition of NB. Visual search behavior data showed that the subjects had the highest number of completed under the condition of OLED ($M = 147.40$, $SD = 46.39$), and the time required for each search task was the least ($M = 19696.93$, $SD = 6691.91$). In the LB condition, the number of completed ($M = 132.90$, $SD = 38.62$) and the time required for each task was the second ($M = 21380.50$, $SD = 7776.95$); the number of completed was the lowest ($M = 130.30$, $SD = 34.63$) under the NB condition, which was the most time required for each search task ($M = 21650.78$, $SD = 7031.89$). The correct rate of the three samples was all more than 97%. A repeated-measure ANOVA was applied to task completion rate of different conditions, a significant main effect of completion number under different conditions (F = 3.668,P = 0.035 < 0.05) was found. The post hoc test showed that completion number of search task under the OLED condition was significantly higher than that under LB condition (P = 0.023 < 0.05). The completion number of search task under the OLED condition was significantly higher than that under NB condition (P = 0.034 < 0.05). A repeated-measure ANOVA was applied to the time required to complete each search task under different conditions and the error rate of the completed search task, there was no significant difference under different conditions (Ps > 0.05). The results showed that the completion rate of visual search task under OLED condition was significantly better than that of LB condition and NB condition (Tables 1 and 2).

Table 1. The repeated-measure ANOVA result of completion performance under different conditions

	SS	df	MS	F value	Sig. (2-tailed)	η^2
Task completion number of different conditions	3396.13	2	1698.07	3.668	.035*	.162
Task completion time of different conditions	44833321	2	244.000	1.023	.369	.051

Table 2. The post hoc test results of completion performance under different conditions

		Mean difference	Std. error	Sig.
OLED	LB	14.50*	5.876	.023*
	NB	17.10*	7.506	.034*

3.2 Comparison of Critical Fusion Frequency Data Results

Critical fusion frequency data showed that the subjects had different degrees of visual fatigue under different conditions, the critical fusion frequency after the task (M = 33.39, SD = 4.53) was lower than that before the task (M = 34.24, SD = 4.45). The two-factor ANOVA results showed that the difference of critical fusion frequency before and after the task was significant (F = 66.07, P = 0.00 < 0.01). The degree of visual fatigue was different under different conditions which the OLED conditions was the lightest. The decrease of critical fusion frequency was the lowest under OLED condition (M = −0.56, SD = 0.89). Under the condition of LB, the decrease rate of critical fusion frequency was the second (M = −0.83, SD = 1.03). Under NB condition, critical fusion frequency decreased the most (M = −1.14, SD = 1.82). A repeated-measure ANOVA was applied to critical fusion frequency under different conditions, a significant main effect of critical fusion frequency (F = 3.14, P = 0.047 < 0.05), The post hoc test results showed that the decrease in the critical fusing frequency under the NB condition and the OLED condition was significant (P = 0.029 < 0.05), the difference between the LB condition and the OLED condition was not significant (P = 0.135) (Tables 3 and 4).

Table 3. A repeated-measure ANOVA result of critical fusion frequency

	SS	df	MS	F vaule	Sig. (2-tailed)	η^2
Phone sample	10.158	2	5.079	3.135	.047	.050

Table 4. The post hoc test results of critical fusion frequency under different conditions

		Mean difference	Std. error	Sig.
NB	LB	−0.305	0.247	.222
	OLED	−0.582*	0.260	.029

3.3 Comparison of EEG Data

The results of emotional EEG data showed that the emotional intensity had increased under three conditions. The emotional index of the subjects after the task (M = 47.08, SD = 0.972) was higher than the emotional index before the task (M = 45.11, SD = 1.975). The degree of rise of emotional index was different in different conditions, the emotional index increased the most under the NB condition (M = 3.24, SD = 10.956), the emotional index increased the second under OLED condition (M = 1.92, SD = 12.690), the smallest increase in emotional index under LB condition (M = 0.73, SD = 11.06). A repeated-measure ANOVA was applied to the emotional index on the same time period under different conditions, the results showed that there was no significant difference in the emotional index between the subjects in the beginning of the task (F = 0.241, P = 0.787), indicating that the basic test results of three samples were consistent. A repeated-measure ANOVA was applied to the emotional index of different time periods under different conditions. The results showed that the difference of phone sample was significant (F = 3.253, P = 0.05) and other factors were not significant. The results of post hoc test showed that the increase of emotional index under NB condition was significantly higher than that under LB condition (P = 0.031 < 0.05), which was also higher than that under OLED condition (P = 0.03 < 0.05). The difference between LB and OLED was not significant. The results of the general data showed that the degree of emotional stress was increased in the course of the task, and the increase of emotional stress was the highest under NB condition, was the lowest under LB condition (Tables 5 and 6).

Table 5. A repeated-measure ANOVA result of emotional index under different conditions

	SS	df	MS	F value	Sig. (2-tailed)	η^2
Phone samples	410.773	2	205.387	3.253	.050	.146
Time	115.672	1	115.682	1.618	.219	.146
Phone samples *time	31.409	2	15.705	.241	.787	.146

Table 6. The post hoc test results of emotional index under different conditions

		Mean difference	Std. error	Sig.
NB	LB	−3.926*	1.683	.031
	OLED	−3.924*	1.669	.030

3.4 Subjective Evaluation Results of Phone Screen

The subjective perception evaluation data of visual fatigue showed that eye fatigue degree (M = 32.3, SD = 21.07) and brain fatigue degree (M = 33.05, SD = 6.047) of subjects was the highest under the NB condition; the eye fatigue degree (M = 32.68, SD = 34.93) and brain fatigue degree (M = 27.15, SD = 27.67) of subjects was the second under OLED condition, and eye fatigue degree (M = 27.43, SD = 23.67) and brain fatigue degree (M = 26.35, SD = 24.396) of subjects was the lowest under LB condition. But the repeated-measure ANOVA results showed that the difference between the samples was not significant.

The results of the subjective perception evaluation scale of the mobile phone screen showed that in terms of the comfort of the display color, the LB condition was the most comfortable (M = 72.35, SD = 18.446), the OLED condition was the second (M = 68.70, SD = 14.261), the NB condition was the most uncomfortable (M = 62.45, SD = 15.244). A repeated-measure ANOVA was applied to data of subjective perception evaluation scale, there was a marginal significant difference in phone samples (F = 3.069, P = 0.058), and the results of the post hoc tests showed that the color comfortable of the LB condition was significantly better than that of the NB condition (P = 0.027 < 0.05). In terms of overall performance of satisfaction, LB conditions were the most satisfied (M = 73.50, SD = 13.543); sample OLED followed (M = 70.10, SD = 18.102); NB conditions were most dissatisfied (M = 60, SD = 19.548). A repeated-measure ANOVA was applied to further study, the results found that the difference between the samples was significant (F = 4.542, P = 0.017 < 0.05), the post hoc test results show that the satisfaction of LB conditions was significantly better than that of NB condition (P = 0.06 < 0.05) (Tables 7 and 8).

Table 7. A repeated-measure ANOVA result of subjective questionnaire under different mobile phone screen

	SS	df	MS	F value	Sig. (2-tailed)
Color comfort	1002.633	2	544.661	3.069	.058
Overall comfort	1972.133	2	986.067	4.542	.017

Table 8. The post hoc test results of color comfort and overall comfort

		Mean difference	Std. error	Sig.
NB	LB	−9.90	4.117	0.27
	OLED	−6.25	4.489	0.18
NB	LB	−13.50	4.362	0.06
	OLED	−10.10	5.587	0.87

The subjects were interviewed on the subjective comfort, color comfort, clarity and general situation of the mobile phone, and the subjects sorted the three samples in the course of the experiment (the best was 1, the worst was 3). The results showed that the subjective comfort of the LB condition was the best (M = 2.25, SD = 0.864), OLED condition was second (M = 2.05, SD = 0.759), NB condition was worst (M = 1.70, SD = 0.865). The score of screen color comfort under LB and OLED conditions was the same (M = 2.05), NB was the most uncomfortable (M = 1.60, SD = 1.005). In terms of screen clarity, OLED condition was the most clear (M = 2.25, SD = 1.02), LB condition was second (M = 1.65, SD = 0.81), NB condition was the worst (M = 1.50, SD = 1.00). Overall, LB condition was the best situation (M = 2.25, SD = 0.79); OLED followed (M = 2.20, SD = 0.696), NB condition was the worst situation (M = 1.55, SD = 0.826). A repeated-measure ANOVA was applied to the interviews of different conditions revealed that the difference between the definition and the comprehensive satisfaction data was significant, but the other difference was not significant. The results showed that the definition between the different samples was significant ($F = 4.03$, $P = 0.026 < 0.05$), and the post hoc test results showed that the OLED condition was significantly better than the NB and LB conditions ($P < 0.05$). The comprehensive satisfaction between different samples was significant ($F = 3.419$, $P = 0.043 < 0.05$), LB and OLED conditions were significantly better than NB ($P < 0.05$) (Tables 9 and 10).

Table 9. A repeated-measure ANOVA result of different mobile phone interview

	SS	df	MS	F value	Sig. (2-tailed)
Clarity	6.30	2	1.655	4.030	.026
Overall satisfaction	6.10	2	3.05	3.419	.043

Table 10. The post hoc test results of clarify and overall satisfaction

		Mean difference	Std. error	Sig.
OLED	NB	−.75	.323	.032*
	LB	−.60	.210	.010**
NB	LB	.70	.325	.044*
	OLED	.65	.293	.039*

4 Conclusions and Discussion

The test results showed that OLED condition and low blue light condition had advantages in the reduction of visual fatigue, and can enhance the user's comfort to a certain extent. Although there were no significant differences in behavioral performance between the three conditions, the critical fusion frequency data showed visual fatigue was induced after the task, and the critical fusion frequency was significantly different between different samples. In the case of OLED, the decrease rate of critical fusion frequency was the lowest, and the degree of visual fatigue was the lightest. Under the condition of LB, the decrease rate of critical fusion frequency was the second, and the decrease of critical fusion frequency was the largest under NB condition, the degree of visual fatigue was the highest. With the progress of the task, the mood of the subjects became more and more irritable, especially in the NB condition, the emotional index increased significantly, while in the LB condition, the emotional index increased the smallest, OLED condition second. Combined with the results of the interview analysis, the integrated situation of subjects in the LB conditions to complete the Orchid Ring search task screen was the best, the subjects felt more comfortable, not easy to cause eye fatigue, and emotional stability; and under NB condition, the subjective scale was the worst, it is easy to make the subjects eye and brain fatigue causing irritability, so the emotional index was very unstable. Subjective evaluation results showed that LB and OLED condition comfort was the best, causing the lowest sense of fatigue; and NB condition comfort was the worst, causing the strongest fatigue. There was not significantly different between LB condition and OLED condition in overall subjective feelings. The results showed that the behavioral performance was not sensitive to the distinction between the blue light control and the subjective evaluation results were not exactly the same, and the critical fusion frequency and the emotional index result were consistent, which was an objective index to evaluate the visual fatigue of the screen that blue light controlled.

Acknowledgments. The authors would like to gratefully acknowledge the support of the National Key R&D Program of China (2016YFB0401203), and China National Institute of Standardization through the “special funds for the basic R&D undertakings by welfare research institutions” (712016Y-4940, 522015Y-3991, 522016Y-4483).

References

1. Berson, D.M., Dunn, F.A., Takao, M.: Phototransduction by retinal ganglion cells that set the circadian clock. Science **295**(5557), 1070–1073 (2002)
2. Yasukouchi, A., Ishibashi, K.: Non-visual effects of the color temperature of fluorescent lamps on physiological aspects in humans. J. Physiol. Anthropol. Appl. Human Sci. **4**(1), 41–43 (2005)
3. Cajochen, C., Mnnch, M., Kobialka, S., et al.: High sensitivity of human melatonin, alertness, thermoregulation and heart rate to short wavelength light. J. Clin. Endocrinol. amp Metabolism **90**(3), 1311–1316 (2005)
4. Ying-bing, C., Yao-jie, S., Yan-dan, L.: Non- visual biological effects of different color light under low illuminance conditions. China Illum. Eng. J. **23**(6), 19–21, 77 (2012)
5. Ide, T., Toda, I., Miki, E., Tsubota, K.: Effect of blue light reducing eye glasses on critical flicker frequency. Asia Pac. Acad. Ophthalmol., 1–6 (2014)
6. Standard logarithmic visual acuity chart developed by the eye hospital of WMU. People’s Medical Publishing House, July 2012

Comparison of Visual Comfort and Fatigue Between Watching Different Types of 3D TVs as Measured by Eye Tracking

Yunhong Zhang[1(✉)], Haibo Yang[2], Yufeng Xu[2], and Lei Feng[1]

[1] AQSIQ Key Laboratory of Human Factor and Ergonomics, China National Institute of Standardization, Beijing, China
zhangyh@cnis.gov.cn

[2] Academy of Psychology and Behavior, Tianjin Normal University, Tianjin, China

Abstract. An eye movement study was conducted to make clear whether different types of 3D TVs would help to relieve visual fatigue after watching films for a long time. 64 undergraduates and ordinary researchers were measured to compare the difference of watching different 3D TVs by Eye-tracking. 64 participants were divided into four groups after being matched, and the four matched groups were separately arranged to watch Switched 3D TV, polarized 3D TV, naked 3D TV and 2D TV. They watched the same video contents which were scenery video and a film, while eye movement data were recorded. The results showed that: (1) with the increase of watching time, participants' fatigue also increased. The blink frequency, blink counts, blink total duration, saccade angle and saccade velocity of the four group participants who watched different types of 3D TVs remarkably increased in the overall trend with time; (2) There was a remarkable difference between the participants watching polarized 3D TV and others watching 3D TV and 2D TV in average saccade duration, saccade duration peak, average saccade angle and saccade total angle, which might indicate that the principle of polarized 3D TV would affect users' saccade duration and saccade distance during watching videos. (3) The saccade amplitude of switched 3D TV was significantly higher than that of other three conditions, which might indicate a greater influence of saccade amplitude in switched 3D TV.

Keywords: Visual fatigue · Visual comfort · Eye-tracking · 3D TV

1 Introduction

The human brain gets most of the information from the visual system. With the continuous development of the Internet and the advance of 3D display technology, the spread of information and the way people access to information are constantly changing. More and more people obtain information from the Visual Display Terminal (VDT). With the expansion of 3D display in ordinary families, visual fatigue problem associated with watching 3D images has aroused the concern of all sectors of society. Visual fatigue shows a series of visual uncomfortable symptoms in physical, such as eye fatigue, dryness, tearing, strain and double vision [1].

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_16

At present, the visual fatigue assessment methods are divided into two categories, subjective evaluation and objective evaluation. Subjective evaluation is mainly a questionnaire, the use of which is simple. But there is no unified standard of quantitative and it is difficult to quantify the level and degree of fatigue. But the development of subjective measurement is more mature. SSQ rating scale is more commonly used, which has 5 point scale, 7 point scale and 13 point scale to choose. There are many subjective evaluation indexes, which can be filtered according to the experiments. Parameters of objective measurement are the eye parameters, event related potential (ERP) and physiological signals, etc. The eye parameters include critical flicker frequency, blinking frequency, the diameter of the pupil and reaction speed, etc. The drop of critical flicker frequency (CFF) reflects the weakening of retinal function. So the change of CFF value can reflect the human eye fatigue to some extent. Objective evaluation usually detects visual fatigue by physiological indexes to determine the level of fatigue, such as pulse waves, brain waves, heart waves, etc. Zou et al. study the condition of eye fatigue after viewing random stereoscopic point figure using brain waves. They found alpha wave is a sensitive index to represent visual fatigue [2]. Although the data from this method is more scientific and accurate, it needs participants to contact the measuring instrument probe and the test process is complex. Comparatively speaking, using non-contact way is easier to evaluate visual fatigue for the researchers.

In recent years, with the development of computer technology, more and more people use the technology of image processing and eye movement parameters to evaluate fatigue of eyes. Eriksson monitored fatigue state through tracking, positioning and recognizing the state of eyes [3]. Grace developed a system to monitor the drowsiness of heavy duty truck driver combined the method of percentage of eyelid closure (PERCLOS) [4]. Guo established a set of fatigue test system through PERCLOS value of fatigue judgment [5]. At present, the mainstream eye-movement apparatus uses the high speed infrared camera to record eye movement trajectory. Represent visual fatigue using the blink of eyes, gaze, pupil size and saccadic indexes [6–8]. Studies have shown that the blink of eyes, pupil size and other indexes can reflect the fatigue status of humans. A number of researchers have shown the relationship between the change of blink frequency and saccadic parameters and the fatigue of eyes. Zhang, Lee, Sakamoto, Kim all use blinking frequency to reflect the degree of visual fatigue caused by watching different VDT. Frequency increased with the extension of time [9–12]. Schleicher and other scholars studied the subjects' behavior of blinking eyes with eye movement apparatus in a simulated driving process. The blink frequency increased with the evolution of the four state of consciousness, awakening, alert reducing, tired and sleepy [8]. At the same time, with the transition to the serious drowsiness, duration of blinking eyes extended. Similarly, some researchers also found the relationship between saccadic parameters and visual fatigue. When Zou studied visual fatigue of 3D display, he found the ratio of peak velocity and twitch amplitude is an effective visual fatigue index produced by watching the 3D display screen. The larger twitch amplitude is and the greater the distance of moving the eyes is, the more difficult of information processing will be. Changes of twitch amplitude shows the humans have fatigue [2]. Not only the eye movement parameter can be used as a visual fatigue index, the parameter of the eye itself can also show the eye fatigue.

As some researchers put forward the method of judging eye fatigue degree of watching free stereoscopic display by measuring the pupil diameter, and the pupil diameter measuring device is developed. The device can record the viewer's pupil diameter when continuously playing stereo video. According to the subjective evaluation of the viewer, the degree of visual fatigue was increased with the increase of viewing time. The pupil diameter of the viewer also increased with the increase of viewing time. There was a positive correlation between visual fatigue and pupil diameter, pupil diameter could be regarded as a simple index of visual fatigue measurement [13, 14]. The study of Zhang et al. found that watching 3D video will make blink frequency decrease, saccade speed, saccade frequency and amplitude increase. The pupil diameter in watching 2D and 3D video showed no significant difference [15].

The above studies have shown that the eye tracking method can be used to test the visual fatigue of the user, but there are inconsistencies between the studies, which need further verification. The purpose of this study is to evaluate the difference of visual fatigue caused by three different production principles of 3D TV monitor to the users by eye tracking method. Compare the effects of different types of 3D TV monitors on visual fatigue and comfort.

2 Method

2.1 Experimental Design

This study is 4 * 2 mixed design, which the between-group factor was three different types of 3D and 2D TV. The within-group factor was before and after watching the film. The evaluation indexes included blink frequency, saccade velocity, saccade amplitude, average saccade time, and average saccade angles etc. Use the subjective fatigue assessment scale to assess the fatigue degree. Subjective evaluation was used to evaluate the comfort, fatigue and satisfaction of the subjects to different types of 3D and 2D TV.

2.2 Subjects

There were 64 college students or researchers participating in the experiment in 18–35 years old. The average age was 24.48 years old, male 33, female 31. All subjects were normal or corrected visual acuity above 0.8 and physical and mental health, no stereopsis blindness, no eye disease, neurological or psychological disease. The four groups were matched for demographic variables such as gender, age, and the time to use the visual display terminal (VDT). All the subjects had not seen the film recently.

2.3 Experimental Material

The test samples were different brands of 47 or 48 Inch 3D LCD TV. Each sample was set to the standard state of its own. Switched 3D TV sample was tcll48f3500a-3D; polarized 3D TV sample was lg47la6200-cn TV; naked 3D TV sample is TCL self-research as3Dp, 2D TV was mainly based on lg2d (14:16).

2.4 Experimental Instruments and Environment

The experimental instrument was SMI RED eye tracker. The sampling rate was 60 Hz. Experimental environment was the normal lighting environment (about 100lx).

2.5 Experimental Procedure

The experiment was carried out in a soundproof laboratory simulated home environment. Subjects were asked to relax for 10 min or more before the experiment, then read the informed consent of the experiment and filled in the demographic information. The object explain the test process considerations and requirements to the subjects. The subjects were divided into four groups, each group respectively watched switched 3D TV, polarized 3D TV, naked 3D TV and 2D TV. Before the experiment, adjusted the height of the table and the position of the display, so that the eyes of the subjects were flush with the center of the display. The distance of the test was about 250 cm. The distance between the eye and the eye tracker was about 70 cm, and the range of the head movement was 40 * 40 cm. The experiment was first calibrated, and then began a formal experiment. Require the subjects to watch scenery video in their own natural state for 10 min. Then watch the 90 min movie “the lightning dog” and watch the same scenery video again for the next 10 min.

2.6 Data Statistical Analysis

The eye movement data of watching the same scenery video before and after seeing the film and subjective evaluation data with different TVs were analyzed by IBM SPSS 22 Statistics software (IBM-SPSS Inc. Chicago, IL) and other professional software. The method of repeated-measure ANOVA analysis were applied to the experiment data.

3 Result Analysis

3.1 Results of Subjective Evaluation of Visual Fatigue

Comfort, Fatigue and Satisfaction Evaluation of Watching Different Types of TV. The results showed that there was no significant difference among users’ comfort and satisfaction by viewing different TVs ($F = 0.01$, $p = 0.96$), but there was a significant difference in the subjective fatigue ($F = 3.88$, $p = 0.01$). The post hoc test showed that the fatigue degree of naked 3D was the most serious, followed by polarized 3D. Switched 3D and 2D had no significant difference. The fatigue feeling of switched 3D, polarized 3D and 2D was significantly less than naked 3D ($ps < 0.05$) (Table 1).

Subjective Fatigue Assessment of Watching Different Types of TV. The subjective fatigue evaluation results of each subscales were shown in Table 2. The data were analyzed by repeated measures ANOVA. The statistical results showed that there was

Table 1. Evaluation of comfort, fatigue and satisfaction of different types of TV

	2D (SD)	Polarized 3D (SD)	Naked 3D (SD)	Switched 3D (SD)
Comfort	51.73 (4.45)	51.88 (4.67)	49.88 (3.98)	49.31 (4.52)
Fatigue	50.13 (3.91)	55.35 (3.81)	66.94 (3.00)	51.19 (4.64)
Satisfaction	65.13 (2.46)	71.71 (2.22)	68.13 (2.57)	64.88 (4.90)

Table 2. Subjective fatigue perception evaluation results of watching different types of TV

	2D	Polarized 3D	Naked 3D	Switched 3D	F	Sig.
Burning	30.27 (4.81)	22.94 (3.22)	34.00 (4.03)	20.63 (2.57)	2.85	0.04
Strain	36.20 (3.88)	34.71 (3.34)	42.63 (3.87)	25.69 (3.05)	0.62	0.01
Tearing	22.40 (4.50)	19.18 (2.90)	28.06 (3.69)	12.63 (2.28)	3.60	0.02
Double vision	18.80 (3.38)	19.82 (3.27)	24.06 (4.78)	11.44 (1.72)	2.30	0.08
Dryness	38.73 (3.94)	32.53 (3.02)	45.31 (4.87)	26.31 (3.14)	4.63	0.00
Headache	36.87 (4.45)	43.76 (4.67)	29.00 (3.98)	29.06 (4.52)	2.66	0.05

significant differences among the subscales of subjective fatigue ($ps < 0.05$). The post hoc test showed that the burning, strain, tearing, double vision, dryness and other problems caused by switched 3D were significantly lighter than those of naked 3D ($ps < 0.05$). The blur degree of polarized 3D was significantly lighter than that of naked 3D ($ps < 0.05$). The headache and dryness degree caused by switched 3D were significantly lighter than those of polarized 3D ($ps < 0.05$). The dryness degree caused by polarized 3D was significantly lighter than that of the naked 3D, but the headache caused by polarized 3D was greater than that of naked 3D and switched 3D.

3.2 Objective Measurement Result of Visual Fatigue

The Saccade Amplitude Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The saccade amplitude data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the time effect was not significant ($F = 0.329$, $p > 0.05$), the differences among the four types of TV were significant ($F = 3.009$, $p < 0.05$), see Table 3. The post hoc test showed that the saccade amplitude of switched 3D was significantly greater than that of three other types of TV ($ps < 0.05$).

Table 3. The variance analysis results of saccade amplitude of watching different types of TV

	SS	df	MS	F	Sig.
TV types	35168189352.29	3	11722729784.10	3.009	0.038
Time	235898954.34	1	235898954.34	0.329	0.568
TV types *time	36867625.94	3	12289208.65	0.017	0.997

The Average Saccade Time Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The average saccade time data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the time effect was significant ($F = 7.510$, $p < 0.01$), the TV types effect was significant ($F = 4.426$, $p < 0.01$), see Table 5. The results of multiple comparison showed that the average saccade time of polarized 3D was significantly greater than that of 2D and switched 3D (ps < 0.01). Afterwards, The post hoc test showed that the average saccade time of viewing the scenery video after the film with polarized 3D was significantly greater than that of viewing the scenery video before the film ($p < 0.01$) (Table 4).

Table 4. The variance analysis results of average saccade time of watching different types of TV

	SS	df	MS	F	Sig.
TV types	61130.38	3	20376.79	4.426	0.007
Time	6321.33	1	6321.33	7.510	0.008
TV types *time	2486.85	3	828.95	0.985	0.407

The Saccade Peak Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The saccade peak data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. Results showed (see Table 5) that the time effect was significant ($F = 4.821$, $p < 0.05$), the difference of saccade peaks among the four types of TV was significant ($F = 3.219$, $p < 0.05$). The multiple comparison results showed that the saccade peak of polarized 3D was significantly greater than that of 2D and switched 3D (ps < 0.05). Afterwards, The post hoc test showed that the saccade peak of viewing the scenery video after watching the movie had great changes. The saccade peaks of viewing the scenery video after watching the movie with polarized 3D and naked 3D were significantly greater than that of viewing the scenery video before watching the movie (ps < 0.05).

Table 5. The variance analysis results of saccade peak of watching different types of TV

	SS	df	MS	F	Sig.
TV types	7146106.31	3	2382035.44	3.219	0.029
Time	1547040.62	1	1547040.62	4.821	0.032
TV types *time	1760723.48	3	586907.83	1.829	0.152

The Average Saccade Angle Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The average saccade angle data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the time effect was significant ($F = 17.933$, $p < 0.05$), the difference of saccade average angle among the four TVs ($F = 2.676$, $p = 0.056$) was critical significant (see Table 6). Multiple comparison showed that the average saccade angle of polarized 3D was significantly greater than that of 2D and switched 3D ($ps < 0.01$), the average saccade average of naked 3D was significantly greater than that of 2D ($p < 0.05$). The post hoc test showed that the average saccade angles of viewing the same scenery video after watching the film with polarized 3D and naked 3D was significantly larger than that of viewing the same scenery video before watching film ($ps < 0.05$).

Table 6. The variance analysis results of average saccade angle of watching different TVs

	SS	df	MS	F	Sig.
TV types	1005.93	3	335.31	2.676	0.056
Time	389.17	1	389.17	17.933	0.000
TV types *time	79.16	3	26.39	1.216	0.312

The Total Saccade Angle Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The total saccade angle data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the time effect was significant ($F = 15.879$, $p < 0.001$), the total saccade angle differences between the four TV types were significant ($F = 2.783$, $p < 0.05$), see Table 7. The post hoc test showed that the total saccade angle of polarized 3D was significantly greater than that of 2D and switched 3D ($ps < 0.05$). Afterwards, the comparison showed that total saccade angle of polarized 3D and naked 3D after watching the film was significantly greater than that of before watching the film ($ps < 0.05$).

The Blink Counts Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The blink response data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results show that the time effect was significant ($F = 22.298$, $Ps < 0.001$), but the difference among the four types of TV was not significant, see Table 8. The results showed that the blink counts of the four types of TV during watching the movie has an increasing trend, but there was no significant difference among the four types of TV.

Table 7. The variance analysis results of saccade total angle of watching different types of TV

	SS	df	MS	F	Sig.
TV types	349839306.50	3	116613102.17	2.783	0.049
Time	100522228.54	1	100522228.54	15.879	0.000
TV types *time	6164822.55	3	2054940.85	0.325	0.808

The Blink Frequency Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The blink frequency data of viewing the same scenery video before and after seeing the film with different TVs was

Table 8. The variance analysis results of blink response of watching different types of TV

	SS	df	MS	F	Sig.
TV types	14344.24	3	4781.41	0.815	0.491
Time	22165.35	1	22165.35	22.298	0.000
TV types *time	77.75	3	25.92	0.026	0.994

analyzed by the repeated measurement ANOVA. The results showed that the time effect of blink frequency was significant ($F = 20.527$, $P < 0.001$), but there was no significant difference of the blink frequency among the four types of TV, as shown in Table 9. The results showed that there was an increasing trend of blink frequency in the four types of TV during watching the movie, but there was no significant difference among the four types of TV.

Table 9. The variance analysis results of blink frequency of watching different types of TV

	SS	df	MS	F	Sig.
TV types	0.176	3	0.059	0.830	0.483
Time	0.265	1	0.265	20.527	0.000
TV types *time	0.005	3	0.002	0.120	0.948

The Average Saccade Speed Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The average saccade speed data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the time effect of saccade average speed among the four types of TV was significant ($F = 32.878$, $p < 0.001$). The post hoc test showed that the saccade speed of viewing the same scenery video after watching the film with 2D, polarized 3D, the naked 3D and switched 3D was significantly greater than that of before watching the film ($ps < 0.05$), see Table 10.

Table 10. The variance analysis results of average saccade speed of watching different types of TV

	SS	df	MS	F	Sig.
TV types	2993.53	3	997.84	0.735	0.535
Time	5344.95	1	5344.95	32.878	0.000
TV types *time	257.56	3	85.85	0.528	0.665

The Blink Time Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The blink time data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the time effect of total blink time was significant ($F = 25.507$, $p < 0.001$). The results showed that the blink time of the four types of television during watching the movie was increased, the total blink time of viewing the same scenery video before and after watching 2D TV had greatly difference ($p = 0.057$), the degree of fatigue was the most light. See Table 11.

Table 11. The variance analysis results of total blink time of watching different types of TV

	SS	df	MS	F	Sig.
TV types	890417248.90	3	296805749.63	0.872	0.461
Time	1489926846.25	1	1489926846.25	25.507	0.000
TV types *time	110786433.96	3	36928811.32	0.632	0.597

The Saccade Count Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The saccade count data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the saccade count before and after watching movies with four types of TV was significantly different ($F = 15.262$, $p < 0.001$). The post hoc test showed that in addition to the 2D effect, the saccade times of viewing the same scenery video after watching movies polarized 3D, naked 3D and switched 3D were significantly less than those of viewing the same scenery video before watching the film ($ps < 0.05$) (Table 12).

Table 12. The variance analysis results of saccade count of watching different types of TV

	SS	df	MS	F	Sig.
TV types	140177.47	3	46725.82	0.899	0.448
Time	174122.25	1	174122.25	15.262	0.000
TV types *time	7562.29	3	2520.76	0.221	0.881

The Saccade Frequency Changes of Viewing the Same Scenery Video Before and After Seeing the Film with Different Types of TV. The saccade frequency data of viewing the same scenery video before and after seeing the film with different TVs was analyzed by the repeated measurement ANOVA. The results showed that the time effect of the saccade frequency was significant ($F = 13.206$, $p < 0.01$). Afterwards, the post hoc test showed that saccade frequency of viewing the same scenery video after watching the film with 3D polarization and naked 3D was significantly less than that of before watching the film ($p < 0.05$), the saccade frequency of viewing the same scenery video with switched 3D before and after watching the movie had critical significant difference ($p = 0.075$) (Table 13).

Table 13. The variance analysis results of saccade frequency of watching different types of TV

	SS	df	MS	F	Sig.
TV types	1.75	3	0.58	0.956	0.420
Time	1.81	1	1.81	13.207	0.001
TV types *time	0.11	3	0.04	0.268	0.848

3.3 Discussion

The subjective evaluation results showed that there was no significant difference between the 4 conditions in terms of comfort and satisfaction. In terms of fatigue, naked 3D was the most serious, followed by polarized 3D, switched 3D and 2D. The results showed that the degree of comfort and satisfaction was difficult to be distinguished by subjective evaluation. But the fatigue could be quantified by subjective evaluation. The fatigue of different TVs watching movies was mainly manifested in eyes dryness, eyes strain, tearing, burning, headache and double vision. The difference was not the same pattern. Naked 3D in dryness, tearing, strain, burning and double vision is significantly greater than the switched 3D. The dryness and blur of naked 3D was significantly greater than polarized 3D. The headache and dryness by Polarized 3D was significantly greater than the switched 3D. The degree of headache caused by polarized 3D was significantly higher than that of switched 3D. Therefore, naked 3D was more likely to lead to the user's tearing, dryness, strain, burning, blur and double vision, polarized 3D was more likely to cause a headache. Because the other parameters of the sample were not strictly controlled, these differences were influenced by many factors of the sample itself. In this study, the technique of naked 3D sample was not mature, the difference between polarized 3D and switched 3D was not significant in many aspects. Therefore, the advantages and disadvantages of polarized 3D and switched 3D could not be determined.

Previous studies showed that the extending of blink duration reflects the excitability of the nervous system decreased, physiological process of nervous system inactivation or slow down. The faster of saccade frequency may indicate that the information processing was difficult. The study found that, with the increase of watching time, blink times, blink frequency and total blink time all showed the overall upward trend, the data of blink times, blink frequency and total blink time after watching movies was significantly higher than that of watching movies before, indicating that these indicators can be characterized as fatigue. Under the four conditions, the blink frequency and blink times showed a trend of decreasing and then increasing. It is probably because the video has not been seen before the subject, so that the subjects focused attention by the impact of the story, leading to blink frequency and the number of blinks first showed a downward trend. With the extension of time, the increase of information acceptance caused the increase of visual fatigue, which led to the increase of blink frequency and blink time and had visual fatigue. This was consistent with the Zhang, stern, Lee, Sakamoto, kim [8–11], etc., which reflect the visual fatigue caused by watching different VDT by blink frequency and blink time. Blink frequency was associated with the individual fatigue degree. Blink frequency decreased under the condition of high concentration. Prolonging blink duration reflects the decreasing excitability of the nervous system, physiological process of nervous system inactivation or slow down.

Saccade parameters of eye movements revealed the mental and physical fatigue of subjects. When using the four kinds of TV to watch the movie, the saccade indexes of subjects increase. Watching 3D showed a higher frequency of saccades, greater saccade amplitude, shorter saccade time and faster saccadic speed. This might imply that the subjects were more alert when watching 3D videos because they need to deal with more visual information. Along with the prolonging of viewing time, saccade parameters decreased in end-stage of watching 3D videos, which shows the input of more information to the brain leads to decrease alertness. Previous studies showed that longer saccade time, bigger saccade angle and greater distance of the eye movement could make information processing more difficult and cause more serious visual fatigue. The saccade speed reflected the excitability level of the physiological system. The study found that the saccade average time, saccade peak, average saccade angle, total saccade angle, average saccade speed, number of saccade and saccade frequency of viewing the same scenery video were significantly different before and after watching the films. The number of saccade and saccade frequency decreased significantly and the other saccade indexes significantly increased. These indicators could be characterized as fatigue phenomenon.

The study found that the eye movement indexes like the average saccade time, saccade peak, saccade average angle, total saccade angle and saccade amplitude between different types of TV had significant differences. The differences between the four kinds of TV mainly manifested that the average saccade time, saccade peak time, saccade average angle and total saccade angle of polarized 3D were significantly greater than that of 2D and switched 3D. The saccade average angle of naked 3D was significantly greater than that of 2D. The results of this study showed that saccade time and distance of the users watching video would be affected by the principle of polarized 3D television. Saccade amplitude of switched 3D was significantly greater than that of the other three kinds of TV. This might indicate that the principle of switched 3D was more sensitive to the saccade amplitude. Saccade amplitude of 2D TV was larger than that of the other two types TV, indicating the saccade amplitude could distinguish viewing effect of various televisions. This was consistent with Zou's [2] visual fatigue study. This might indicate that the average saccade time, saccade duration peak, average saccade angles, total saccade angles and saccade amplitude indicators might be sensitive to the differences of different types of TV.

4 Conclusion

The study found that, with the increase of viewing time, the degree of fatigue that subjects feel had increased. The data of the blink times, blink frequency, saccade average speed, blink time, the average saccade time, saccade peak, saccade average angle and total saccade angle was significantly higher than that of viewing the secrecy video before watching movies, which has an upward trend. The data of the saccade number and saccade frequency was significantly lower than that of before watching movies, showing a downward trend. It indicates that these eye movement indicators may reflect the fatigue of eyes, providing more evidence for the study of visual fatigue.

On the visual fatigue evaluation of different types of display TV, in the indexes of average saccade time, saccade peak, average saccade angle and total saccade angle, that of polarized 3D was significantly higher than that of other 2D and 3D television switched TV. It may indicate the principle of 3D polarization television may affect the users' saccade time and distance of watching video. The saccade amplitude of switched 3D TV was significantly larger than that of other three conditions, which may indicate switched 3D TV had a greater impact on the saccade amplitude. From the evaluation method, eye tracking analysis method can be used as a reliable method of visual fatigue evaluation. But for display terminals with different production principles, the sensitive indicators will be different. As for the subjective evaluation method, it is difficult to distinguish the comfort and satisfaction of the monitor by subjective evaluation. But the fatigue could be quantified by subjective evaluation.

Acknowledgments. The authors would like to gratefully acknowledge the support of the National Key R&D Program of China (2016YFB0401203), and China National Institute of Standardization through the “special funds for the basic R&D undertakings by welfare research institutions” (712016Y-4940, 522016Y-4483, 522015Y-3991). The authors also acknowledge the support of TCL Corporation, too.

References

1. Yanyan, C., Lili, W., Lanlan, Y.: Research on visual fatigue of 2D/3D display under different lighting conditions. Electron Device **39**(2), 242–247 (2016)
2. Zou, B., Liu, Y., Wang, Y.: 11.4: Study on the saccadic eye movement metric of visual fatigue induced by 3D displays. In: SID Symposium Digest of Technical Papers, pp. 134–137 (2015)
3. Eriksson, M., Papanikotopoulos, N.P.: Eye-tracking for detection of driver fatigue. In: IEEE Conference on Intelligent Transportation System ITSC 1997, Boston, Massachusetts (1997)
4. Grace, R., Byrne, V.E., Bierman, D.M., et al.: A drowsy driver detection system for heavy vehicles. In: Digital Avionics Systems Conference, Seattle (1998)
5. Guo, B., Fang, W.: Fatigue detection method based on eye tracker. Aerosp. Med. Med. Eng. **7**(4), 256–260 (2004)
6. Morris, T.C., Miller, J.C.: Electrooculographic and performance indices of fatigue during simulated flight. Biol. Psychol. **42**, 343–360 (1996)
7. Summala, H., Hakkanen, H., Mikkola, T.: Task effects on fatigue symptoms in overnight driving. Ergonomics **42**, 798–806 (1996)
8. Schleicher, R., Galley, N., Briest, S., et al.: Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? Ergonomics **51**, 982–1010 (2008)
9. Zhang, Y., Zhao, C., Zhang, X.: Visual comfort and fatigue between watching linear polarized and circular polarized LCD TVs as measured by eye tracking. In: Harris, D. (ed.) EPCE 2015. Springer, Cham (2015)
10. Lee, E.C., Park, K.R., Whang, M., et al.: Measuring the degree of eyestrain caused by watching LCD and PDP devices. Int. J. Ind. Ergon. **39**(5), 798–806 (2009)
11. Sakamoto, K., Aoyama, S., Asahara, S., et al.: Measurement of visual fatigue for large-sized TVs in a home viewing environment. In: The 13th IEEE International Symposium on Consumer Electronics, pp. 738–742 (2009)
12. Kim, D., Choi, S., Park S, et al.: Stereoscopic visual fatigue measurement based on fusional response curve and eye blinks. In: Digital Signal Processing (DSP), pp. 1–6 (2011)
13. Feng, Z., Weijie, F., Guiyang, Z.: The relationship between pupil diameter and visual fatigue of 3D LCD using. J. Clin. Ophthalmol. **21**(2), 153–155 (2013)
14. Qionghua, W., Dongdong, P., Xiaofang, L., et al.: The relationship between pupil diameter and visual fatigue in autostereoscopic display. J. Photonics **40**(3), 350–353 (2011)
15. Zhang L., Ren J., Xu L., et al.: An experimental study on the evaluation of visual fatigue induced by stereoscopic video display by eye tracker. Ophthalmology **1**, 37–42 (2014)

An Eye-Movement Tracking Study on Influence of Circularly Polarized Light LCD and Linearly Polarized Light LCD on Visual Perception Processing Ability for Long-Term Continuous Use

Yunhong Zhang[1(✉)], Jinghong Ding[2], Haibo Yang[3], Fan Zhang[1], Haitao Wang[1], and Yu Chao[3]

[1] AQSIQ Key Laboratory of Human Factor and Ergonomics, China National Institute of Standardization, Beijing, China
zhangyh@cnis.gov.cn
[2] Education Academy, Capital Normal University, Beijing, China
[3] Academy of Psychology and Behavior, Tianjin Normal University, Tianjin, China

Abstract. The study was conducted to evaluate the changes of visual perception processing ability after using circularly polarized light LCD and linearly polarized light LCD for a continuous month by eye-movement tracking method. 47 subjects were randomly divided into circularly polarized light group (experimental group) and linearly polarized light group (controlled group). A 2 (times of measurement: before and after) * 2 (display equipment: circularly polarized light LCD and linearly polarized light LCD) experiment was designed. There was 1 month interval between the two times of measurement. The dependent variables were corrective eye frequency and tracking gain. The results showed that there was critically significant difference of eye-movement tracking gain between the circularly polarized light condition and the linearly polarized light condition. The post hoc comparison showed that the tracking gain after 1 month was significantly more than that of before 1 month. Therefore, the influence of circularly polarized light LCD on visual perception processing ability was slighter than that of linearly polarized light LCD for a long term.

Keywords: Eye-movement tracking · Circularly polarized light · Linearly polarized light · Visual perception processing ability · Visual tracking gain

1 Introduction

With the popularization of video display terminals (VDT) like computers, television sets and mobiles, the population of VDT users grow rapidly. All kinds of eye discomfort caused by reading computer close for a long time or working related to computers increase, resulting Video display terminal syndrome (VDTS) becomes more and more common, known as eye fatigue syndrome clinically. About 60% of Americans have

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_17

some symptoms of asthenopia after operating computer more than 3 h according to a latest report released in the Vision Council [1]. A series of symptoms caused by VDT have harm to the user's physical and mental health. Therefore, how to reduce eye fatigue and improve work efficiency when people use VDT becomes key link of VDT product design. The influence of different types of displays on visual fatigue also has differences in addition to the environment and personal factors.

LCD has attracted the interest of general research workers and device engineers with its peculiar physical and optical properties from the date of discovery. LCD realizes the display of monitor through the liquid crystal molecules as light control switch. The brightness of the display, screen flicker and reflective can all cause visual fatigue. Most of the liquid crystal displays achieve display by the modulation of polarized light. Polarized light includes circularly polarized light and linearly polarized light. Polarization is a basic characteristic of light like the color and intensity. Difference exists between linearly polarized and circularly polarized light during the process of forming vision. Compared to the linear polarized light, circularly polarized light is more close to natural light, which is the most comfortable light for human. Shixian found that blink frequency and fatigue of eyes caused by the linearly polarized light was higher and easier during the study of polarized light television compared with circularly polarized light [2]. Yan Xiaolin discovered linearly polarized light was easier than circularly polarized light to cause visual fatigue through the eye blink frequency recorded by the electro-oculogram (EOG) during the study of influence of circularly polarized light liquid crystal TV on eye fatigue [3]. Zhang Yunhong, Zhang Li et al. also got the same result in the eye tracking study [4, 5]. It is found that circularly polarized light as light source was less likely to produce visual fatigue by summarizing predecessors' study.

Visual fatigue belongs to functional disease which is related to the habits of using eyes, one's own physique and social environment. Most research shows that visual fatigue and mental health influence each other. It belongs to the physical and mental combined syndrome. There are different evaluative methods to assess visual fatigue according to different research purposes and different environment, which mainly include ophthalmic physiological aspects and eye- movement tracking.

Critical flicker frequency (CFF) is the most applied indicator among the physical measurements of eye fatigue. Iwasaki and Akiya found the decline of CFF reflected the degree of retinal degeneration caused by visual fatigue and the wanes of retinal nerve or the optic nerve activity [6]. Many research of VDT operation visual fatigue takes flicker fusion frequency as evaluation index of eye fatigue [7]. Flicker fusion frequency was used for evaluating visual fatigue level in many researchers' experiments, which include Kumashiro's VDT visual fatigue experiment in last century [8], Murata's long-term VDT operation experiment [9], Hsu's experiment comparing visual fatigue caused by different screens [10], experiment comparing mobile screen brightness completed by Li et al. [11] and visual fatigue evaluation experiment of three-dimensional display system by Zou et al. [12]. But flicker fusion frequency can only be used for evaluation before and after the test, which cannot continuously monitor the status of fatigue of subjects. Besides, its test is influenced by many factors and has high requirement to the control of the experiment.

Eye-movement tracking method gets the state of fatigue by monitoring the subjects' eyes. The percentage of eyelid closure (PERCLOS) determines the level of fatigue with ratio of your eyes closed time accounts for a specific time. It has three kinds of decision criteria, which are P70, P80 and EM. The P80 has the best correlation with fatigue [13]. Hence, it was used as a criterion based on eye fatigue. But this method is more difficult for researchers without the basis of image recognition as it needs identification and analysis on image data of the eyes. Blinking frequency is also used as the severity index of visual fatigue, but it is vulnerable to the types of stimulus and different psychological state of the subjects. The results of different studies are not consistent. So, it is not stable index evaluating visual fatigue.

The motion tracking of eye tracking indicators can be a quantifiable and stable evaluation index, which reflects the attentional ability of eye to follow the target movement. When the individual is tired, visual tracking ability will be markedly reduced. Obtaining tracking moving target is the basic premise of visual tracking, which is to focus on track targets on the basis of the background and target mutual separation. In order to be able to effectively track, the visual system must set up object characterization of continuity based on processing of target characteristics of space-time and objective and identity to overcome the change of trajectory and temporary shelter. Visual tracking motion is the results of integration of target movement information and identity semantic information by vision system. In the process of target acquisition of visual tracking, visual system will face a lot of objects within the same time. But people cannot process all the objects into visual system at the same level of priority. Only the object through selective attention screening can be track target and got further processed. Effective clue hint can guide the observer to view specific position, then, search stimulus related to behavioral target [14]. Under the condition of same clue hint, the selective attention to location will take precedence over objects [15]. This selective attention based on space position is usually in top-down processing way promoting target detection. There are two kinds commonly used indexes in visual tracking, correction saccadic and tracking gain. Prediction strategy of correction saccadic can prevent larger deviation in position and velocity between the eye and target. When the target movement track changes or moves faster, catch-up saccadic can react and correct the existing deviation [16]. Tracking gain is the ratio between eye speed and target speed. When the tracking target's movement speed accelerates or target trajectory changes suddenly, eye speed will significantly slower than the target speed. Then tracking gain reduces and position error accumulates. As the eye movement speed drops, however, the differences between the eyes and the target becomes greater, which can the often increase tracking gain and correct the deviation. Therefore, this research intends to use eye tracking index to track study visual tracking ability of users using circularly and linearly polarized display for a long time and verify whether using computer monitor with circularly polarized light for a long time is better than that of linearly polarized light in reducing visual fatigue.

2 Method

2.1 Experiment Design

A 2 (times of measurement: before and after) * 2 (display equipment: circularly polarized light LCD and linearly polarized light LCD) experiment design was used in the study. There were 2 times of measurement, before and after the experiment treatment, the interval of which was 1 month. The subjects were divided to experimental group and controlled group after being matched. The experimental group was arranged to use circularly polarized light computer monitor group and the controlled group was arranged to use linearly polarized light computer monitor group.

2.2 Subject

The subjects comes from the graduate students and researchers under 35 years old who used the computer frequently. There are 42 people for the first test. The experimental group has 21 people (M = 25.24, SD = 25.24), including 11 men and 10 women. The controlled group has 21 people (M = 25.71, SD = 4.03), including 11 men and 10 women. There are 38 subjects in the second test after 1 month, as partial subjects' loss due to job changes. The experimental group has 20 people (M = 24.90, SD = 3.55), including 10 men and 10 women. The controlled group has 18 people, including 8 men and 10 women. All the subjects are with uncorrected or corrected visual acuity of 1.0 plus and without any ocular and mental diseases.

2.3 Apparatus

This study uses Germany SMI iView X series Hi-speed eye movement instrument to record and analyze eye movement data. Experimental program is programmed by E-Prime. The sampling frequency is 500 Hz. The stimulus presentation host is M6900 of Lenovo apocalypse with faster 2.60 GHz, memory 2.00 G and independent video card. Data collection host is faster 3.00 GHz, memory 3.00 G and independent video card. Stimulus presentation display is CRT display of Viewsonic. The display screen resolution for stimulus presentation and gathering is 1024 $\times$ 768. The refresh rate of stimulus presentation display is 75 Hz.

2.4 Procedure

There is two tests in the experiment. The first test is the former. After the former test, Paste on the computer screens of the subjects with linearly polarized light membrane and circularly polarized light membrane after tests. The second test is conducted after the subjects continuously use computer monitors for a month. The experimental location and environmental layout of the two experiments are exactly the same. Each experiment is conducted in quiet environment. Firstly, experimenter introduce experiment content to subjects. Subjects fill out basic information registration form and sign the informed consent. Eye movement instrument is placed in front of stimulus presentation display. Subjects sit in the middle of the stimulus presentation screen. The distance of eyes and monitor is about 62 cm and eye height is to screen center. When gathering the eye movement data, use the head fixed bracket to prevent the subjects too much head movement that can affect the quality of data. Experiment requires the subjects to continuously follow the beating dot and keep the movement with the dot synchronization as far as possible. The eye movement data is captured and recorded by the eye movement instrument. Stimulus presentation uses the E-prime software, including calibration, practice and formal experiment. At the beginning of each task, the experimenter let the subject sit up before the eye-movement apparatus, fix his/her head and start the calibration procedure. After calibration, experimental instruction and experiment practice was presented. Formal experiment cannot be performed until the subjects are fully familiar with the whole experimental process and operation. The experimenter has real-time monitoring eye movement trajectory of subjects through the display during the experimental process. Each task includes three groups. Each subject need to complete the visual tracking task 100 times in each test. The experimenter reminds the subjects to keep body and head rest before each group experiment. Close their eyes to have a 1 min rest to void fatigue between each group. After the experiment, thank and pay for the subjects.

3 Results and Analysis

3.1 Comparative Analysis of the Correction of Saccadic Frequency Results of Two Groups

The repeated measures ANOVA analysis was used to analyze visual tracking frequency of saccadic results of the two groups of subjects in two successive tests. The results show there was significant difference between test times ($p < 0.001$), no significant differences between groups and the interaction is not significant. The post hoc analysis results show that the correction of saccadic frequency of the second test was significantly higher than that of the first test ($F(1, 72) = 75.22$, $p < 0.001$) (Table 1).

Table 1. The variance analysis of visual tracking saccadic frequency under different test conditions

	SS	df	MS	F	Sig
Group	0.005	1	0.005	0.015	0.903
Test	24.257	1	24.257	75.218	0.000
Group*test	0.032	1	0.032	0.098	0.755

3.2 Comparative Analysis of Visual Tracking Gain Results of the Two Groups in Two Successive Tests

The repeated measures ANOVA analysis was applied to visual tracking gain results of the two groups of subjects in two successive tests. The results show that the interaction of them is significant ($F(1,72) = 5.04$, $p < 0.05$), the main effect of the two successive tests is significant ($F(1,72) = 304.73$, $p < 0.001$), and the main effect of the groups is critical significant ($F(1,72) = 3.92$, $p = 0.052$). Simple effect analysis shows that tracking gain of the experimental group is significantly higher than that of the controlled group in the second test, ($t(33) = 2.16$, $p < 0.05$). And, the track gain of the second test is significantly higher than that of the first test ($p < 0.001$) (Table 2).

Table 2. The variance analysis of visual tracking gain under different test conditions

	SS	df	MS	F	Sig
Group	0.513	1	0.513	3.915	0.052
Test	39.939	1	39.939	304.733	0.000
Group*test	0.660	1	0.660	5.038	0.028

4 Discussion

Light is a transverse wave, the plane composed of the electric vector vibration direction and heading direction of light is called the vibration plane. If the vibration of the light plane is not evenly distributed, it is called polarized light. If the vibration plane of light is limited to a fixed direction, it is called a linearly polarized light. If the vibration plane of the light turns around the direction of light and the electric vector direction has a uniform circular distribution in the cross-section of propagation direction, the polarized light is called circularly polarized light. The vibration plane of most ambient light in nature environment (such as the sun light) is evenly distributed in all directions, which is called natural light. Therefore, from the perspective of the principle of polarized light, linearly polarized light is more close to natural light in optical properties compared with circularly polarized light. Circularly polarized light display really has advantage over linearly polarized light display from the comparative test results before of continuously using circularly polarized light monitor and linearly polarized light monitor at the same time [4, 5]. The results further confirm that from a long-term continuously using 1 month linearly polarized and circularly polarized light displays, visual tracking ability of the individual changes. The correction saccade of experimental group in the second test is significantly higher than that of controlled group in the second test. Its visual tracking gain is also significantly higher than that of controlled group.

When individuals get fatigue, trouble concentrating is the first character and the visual tracking movement parameters such as saccade and tracking gain will also change. Obtaining tracking target is the basic premise of visual tracking. That is focusing on tracking targets on the basis of mutual separation of background and target. That is to say, it begins to track once the visual system identifies the target object. Visual tracking takes up plenty of cognitive resources of the subjects. It requires the subjects to focus on the visual tracking task. But now it is believed the causes of visual fatigue caused by VDT are the decrease unconsciously blink and bad blink when concentrating on looking at the screen. It makes the cornea expose to the air longer, which is unfavorable to the distribution grease in tears. It can lead to inspect table environment variation and cause asthenopia. That means the visual tracking task is a better index of testing the fatigue of eye movement.

Polarization is a basic characteristic of light like the color and intensity. Difference exists between linearly polarized and circularly polarized light during the process of forming vision. The vibration of the natural light in all directions distributes uniformly. The stimulation to the human eye's light-sensing cells is isotropic. The human eye is used to this kind of natural light. The vibration plane of circularly polarized light spins fast. The comprehensive effect to the human eye is also isotropic. It is most close to natural light. The vibration plane of the linearly polarized light is fixed. The stimulation to the visual system is not balanced. So, compared with the linearly polarized light, the human eye is easier to adapt to the image presented by circularly polarized light display. That means it is less likely to produce visual fatigue when using circularly polarized light as display light source.

5 Conclusions

This research records the changes of eye movement indexes of the subjects finishing the visual tracking task with the eye movement instrument before and after the long-term use of circularly polarized light and linearly polarized light computer monitors. Compare the change difference of the eye movement indexes of subjects in the visual tracking task before and after 1 month continuously using different types of displays. It further verifies irrelevant factors of interference in the process of task can be reduced with long-term use of circularly polarized light computer monitor compared with the long-term use of linearly polarized computer monitors. Embody in the changes of indexes like the correction saccade frequency and visual tracking gain. In the visual tracking gain indicator, the effect of aureole display shows in the second test. Results confirm that visual function damage caused by the long-term continuous use of circularly polarized light is lighter than the long-term use of linearly polarized display, which means using circularly polarized light computer monitors can reduce damage of visual function. This study also confirms the visual tracking task can be the measurement indicator of visual fatigue of visual display terminal.

Acknowledgments. The authors would like to gratefully acknowledge the support of the National Key R&D Program of China (2016YFB0401203), and China National Institute of Standardization through the "special funds for the basic R&D undertakings by welfare research institutions" (712016Y-4940, 522016Y-3991).

References

1. African vision and eye health. African Vision Research Institute African Ophthalmology Council SD Mathebula. African Vision and Eye Health, **74**(1) (2015)
2. Long, S., Wu, D., et al.: Evaluate the influence of the polarized light characteristic of LCD TV on visual fatigue using blink frequency. Chin. Ophthalmic Res. **26**(5), 380–383 (2008)

3. Lin X., Fu D., Zhu C.: The influence of circularly polarized light liquid crystal TV on eye fatigue. J. China Prev. Med. **10**(8), 798–801 (2009)
4. Zhang, L., Xu, L., et al.: Research of asthenopia caused by watching linearly polarized light liquid crystal TV and circularly polarized light liquid crystal evaluated by eye tracker **25**(1), 63–67 (2006)
5. Zhang, Y., Zhao, C., Zhang, X.: Visual comfort and fatigue between watching linear polarized and circular polarized LCD TVs as measured by eye tracking. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics, pp. 84–91. Springer, Cham (2015)
6. Iwasaki, T., Akiya, S.: The significance of changes in CFF values during performance on a VDT-based visual task. In: Kumashiro, M., Megaw, E.D. (eds.) Towards Human Work: Solutions to Problems in Occupational Health and Safety. Taylor & Francis, London (1991)
7. Chengming, C., Hui, W., et al.: Review of VDT visual fatigue and its measurement method. Chin. J. Ergon. **19**(2), 92–95 (2013)
8. Kumashiro, M., Mikami, K., Hasegawa, T.: Effects of visual and mental strain on VDT (visual display terminal) performance. Jpn. J. Ind. Health **26**(2), 105–111 (1984)
9. Murata, K., Araki, S.: Accumulation of VDT work-related visual fatigue assessed by visual evoked potential, near point distance and critical flicker fusion. Ind. Health **34**(2), 61–69 (1996)
10. Hsu, B.W., Wang, M.J.: Visual fatigue evaluation of PDP, LCD and CRT TV. In: Proceedings of 17th World Congress on Ergonomics (2009)
11. Li, H., Zhang, Y., et al.: Study of the optimal parameters of mobile phone screen brightness under different environment illumination-take Samsung mobile phone as an example. J. Psychol. Sci. **36**(5), 1110–1116 (2013)
12. Zou, B., Liu, Y., Wang, Y.: 11.4: study on the saccadic eye movement metric of visual fatigue induced by 3D displays. In: SID Symposium Digest of Technical Papers, pp. 134–137 (2015)
13. Beiyuan, G., Weining, F.: Fatigue detection method based on eye movement apparatus. Space Med. Med. Eng. **17**(4), 256–260 (2004)
14. Franconeri, S.L., Simons, D.J.: Moving and looming stimuli capture attention. Percept. Psychophys. **65**(7), 999–1010 (2003)
15. Soto, D., Blanco, M.J.: Spatial attention and object-based attention: a comparison within a single task. Vis. Res. **44**(1), 69–81 (2004)
16. de Brouwer, S., Missal, M., Barnes, G., Lefevre, P.: Quantitative analysis of catch-up saccades during sustained pursuit. J. Neurophysiol. **87**(4), 1772–1780 (2002)

Visual Pattern Complexity Determination for Enhanced Usability in Cognitive Testing

Kanaka Babshet[1(✉)], Catherine Honegger[1], Ashley Gritzman[2], and Vered Aharonson[1]

[1] School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa
kbabshet@hotmail.com, catherineannhonegger@gmail.com, vered.aharonson@wits.ac.za
[2] IBM Research, 47 Juta Street, Braamfontein, Johannesburg, South Africa
ashley.gritzman@za.ibm.com

Abstract. Computerized cognitive tests often entail tasks related to visual stimuli. An efficient complexity measure for these tasks can enhance their cognitive evaluation accuracy, specifically for elderly and cognitively impaired subjects. This paper details the design, implementation, and testing of a visual pattern complexity determination algorithm. The patterns used for the study are sixteen-bit binary patterns taken from computerized cognitive assessments. Three complexity levels were defined based on the visual perception of human subjects: *easy*, *medium*, and *hard*. The algorithm was tested on three hundred patterns and the results were compared to the parallel complexities perceived by human judges. Correlations of 72%, 74%, and 61% between human perception and the algorithm's predictions were obtained for the *easy*, *medium*, and *hard* patterns, respectively. The algorithm has potential to become an accurate measure of visual pattern complexity in computerized assessment, and could improve the usability of these tests for psychometric and cognitive evaluations.

Keywords: Binary images · Cognitive assessments usability · Human perceived complexity · Visual pattern complexity

1 Introduction

Neuropsychological tests are used for the detection of cognitive decline, which may indicate neurodegenerative diseases like Alzheimer's disease and dementia. Some tests make use of visual stimuli in the form of four-by-four, black-and-white patterns, which subjects are asked to recognize and recall [1–3]. Neurocognitive decline can be evaluated using the subject responses to these tests [4]. To accurately assess the subjects' performance in the cognitive tests, as well as to enhance the usability of these tests, the complexity of the tasks involved should be defined.

The perceived complexity of the image directly influences the difficulty of the task; thus, it is important to determine the pattern complexity of the visual stimuli. This could enable discrimination between subjects who pass a test involving more complicated tasks (which might indicate that they are less susceptible to cognitive decline),

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_18

and those who can only pass tests involving simpler tasks [5]. Previous neuropsychological tests have shown that patients perform much better under less stressful circumstances. If a method for determining the complexity of a task is developed, it would enhance the usability of computerized cognitive testing by organising tests according to a patient's abilities. The testing could begin with an intermediate task; if the patient gets the answers right, the task difficulty would increase, and if the patient gets the answers wrong, the difficulty would decrease. This allows a quantitative measure of the patient's competency. In addition, if the stimuli are randomly introduced with no adjustment of the complexity to the user capabilities, the process might be non-effective and frustrating (e.g. introducing high complexity tasks to a cognitively impaired subject) [6].

Currently, if the visual complexity of the tasks is assessed, it is done primarily by human judges, thus providing a subjective analysis and precluding the usage of a large quantity of patterns [2]. In addition, existing mathematical complexity algorithms (e.g. Shannon Information Theory and Kolmogorov Complexity) do not incorporate the human factors which define complexity from a human's perspective.

To ensure an objective assessment of a subject's task performance, an optimal visual pattern complexity algorithm is developed, with a close correlation to the human perceived pattern complexity. This algorithm is developed based on an analysis of human visual perception, translating the observations into mathematical formulations, and testing whether the algorithm can produce similar results. The algorithm was implemented in Java.

The performance of the algorithm is evaluated by measuring the correlation between its computed complexity rating, and the respective human complexity rating.

Section 2 details the project background, including the underlying rationale of the algorithm, visual stimuli patterns and tasks, and the testing database. Section 3 describes the algorithm development, and Sect. 4 presents and analyses the test results.

2 Background

2.1 Underlying Rationale of the Algorithm

The algorithm aims to capture the human visual perception of complexity. The concept is based on Attneave, who describes visual perception as an information-handling technique, where a considerable amount of the information communicated is likely to be redundant [7]. Attneave states that humans use redundant visual stimuli to see patterns and simplify the information that the brain perceives. Any visual invariance constitutes a source of redundancy because the subject automatically begins to extrapolate the rest of the image. The techniques that Attneave believes the brain employs include the concentration of information at certain points, image symmetry, and adjoining pixels of the same colours, shapes, lines or even angles.

As the redundancy of an image increases, a subject is more likely to approximate the rest of the image correctly, thus decreasing its perceived complexity. The extent to which a human can extrapolate this information is still unknown [7]. The information distribution of an image according to Attneave's hypothesis is the trait that allows humans to identify objects easily [7].

2.2 Visual Stimuli Patterns

The tests in standard neuropsychological assessments of cognitive decline (used for the early detection of Alzheimer's and other neurodegenerative diseases) include abstract patterns from a dataset [2–4]. These patterned images are four-by-four square matrices where each element is either black or white in colour. Three randomly chosen images from the dataset are presented in Fig. 1 [2].

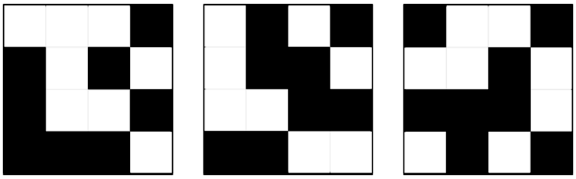


Fig. 1. Examples of the patterned images used in standard neuropsychological assessments.

2.3 Visual Stimuli Based Tasks

The first task is a recall test: the subject is shown a single pattern for a short time, after which the pattern disappears. The subject is then presented with three patterns of which one was previously shown. The subject is required to remember this pattern and point it out amongst the three displayed. The second task is a recognition test: the subject is presented with three patterns where two are the same and one is different. The subject is then required to recognise the different pattern and select it.

In order to distinctively measure a subject's performance in the cognitive test, the complexity of the task where two or three different patterns are displayed, must be assessed. In this study, however, only the perceived complexity of a single pattern is evaluated, which directly influences the difficulty of the task.

2.4 Database

The preliminary dataset included all 65 536 (2^{16}) possible sixteen-bit binary patterns. These patterns were graphically generated for visual inspection by two human judges. An image was randomly selected, and each judge independently rated it as either *easy*, *medium*, or *hard*. If the judges concurred, then the image was assigned the corresponding rating. If the judges differed, then a third judge was asked to cast the deciding vote. The process was repeated until one hundred images from each of the three categories were accumulated. The dataset was then constructed with these three hundred images, along with their human perceived complexities, and was stored in a text file for testing purposes.

3 Algorithm Development

3.1 Algorithm Overview

Based on Attneave's study on human visual perceptions [7], an algorithm was designed and developed to classify cognitive decline test patterns into three complexity levels: *easy*, *medium*, and *hard*. The aim of the algorithm is to mirror the human observations, and to yield similar results to human perceived complexities. This algorithm determines the number and length of adjacent black paths, adjacent white paths, diagonal black paths, diagonal white paths, and single black and white blocks. Once the components within the image have been stored, a weighted addition of the components is calculated and used to determine the visual pattern complexity of the image. The algorithm then evaluates an overall accuracy of the results against the complexities measured by the human judges, and outputs each computed complexity level with detailed percentage correlations.

Figure 2 illustrates the overview of the optimized visual pattern complexity algorithm.

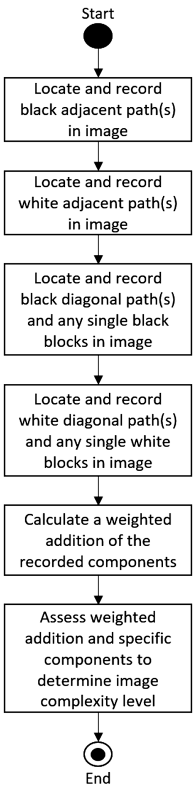


Fig. 2. Logical flowchart of the algorithm implemented.

3.2 Algorithm Design

This algorithm is run on a sixteen-bit binary array. The array is first transformed into a four-by-four array to simulate the visual representation of the image. This allows for better navigation of the image from the top left square to the bottom right square when conducting the appropriate analyses.

Adjacent Path Analysis. The blocks in the two-dimensional array are given position numbers from one to sixteen starting at the top left corner, and proceeding row by row, left to right. Each element in this array holds an empty adjacency list that will be populated based on its adjacent members. The algorithm checks if the current block is black as well as if it has any other black blocks adjacent to it. The position numbers of this block as well as any adjacent black blocks are then appended to the original element's adjacency list. Should the block not be black and/or it does not form an adjacent line, the list will remain empty. The algorithm then goes through the adjacency list array merging all paths with at least one block in common to create a list of black adjacent paths. The results are counted to return the number of adjacent paths in a pattern.

The longest path analysis is illustrated using an example of pattern 1011000 101001011 in Fig. 3. In this instance, the algorithm will calculate that there are two adjacent paths of black blocks. These adjacent paths will then be removed from the array to avoid conflicting measurements.

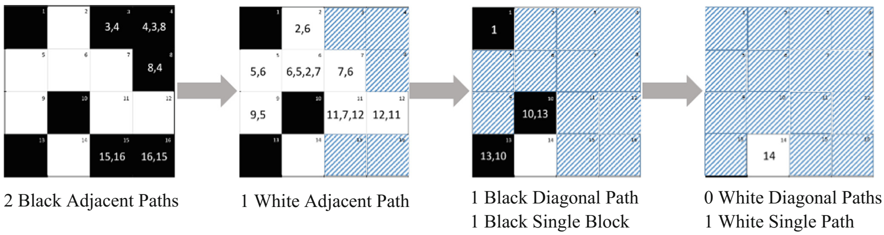


Fig. 3. Step-by-step example of the algorithm on pattern 1011000101001011

Diagonal Path Analysis. A similar procedure is then conducted to detect the diagonal paths in a pattern; however, the algorithm now checks the diagonal neighbours of each block. Therefore, if a block itself is black, the list in that element will contain the position numbers of any corresponding black diagonal neighbours as well. The algorithm then goes through the diagonal list array merging all paths with at least one block in common to create a list of the black diagonal paths contained in the pattern. This procedure is shown in the third step in Fig. 3.

Single Block Analysis. After the adjacent and diagonal black paths have been eliminated from the pattern, any remaining black blocks are classified as 'single blocks'. These have also been identified and incorporated into the complexity measure since they add an element of complexity to the pattern. This is illustrated in step three of Fig. 3, where the example now has one single black block.

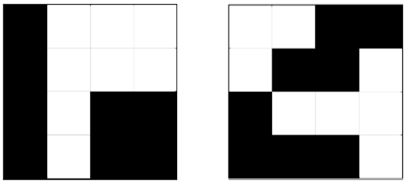


Fig. 4. Comparison of two patterns with the same measure of black adjacent paths.

White Component Analysis. Not only does the presence of the black components affect the pattern complexity, but their positions and shapes alter the measure as well. The white space in a pattern gives an indication of how the black components are dispersed. For example, more concentrated white space indicates that there are fewer black components, and/or the existing black components in the pattern are more clustered together, which could make a pattern simpler. This principle is illustrated in Fig. 4 below where both patterns have the same black component measures; however, one is classified as more complex than the other due to the positions and shapes of the black paths. Therefore, the fact that there is a single large white adjacent path through a pattern could make it simpler than another pattern that has smaller, disconnected white components.

Consequently, the number of white long paths, white diagonals, and white single blocks are calculated in a similar method for each pattern as illustrated in Fig. 3.

Component Integration. At this point, six different components must be examined to produce a single pattern complexity measure. Before combining the components, there are three specific characteristics that are assessed to automatically classify an image with complexity level 1 (*easy*):

1. If there are less than three black adjacent paths, no other black components, and only a single white adjacent path
2. If there is only one black diagonal path, and no other black components
3. If there are less than four single black blocks, and no other black components

If the pattern does not have any of these characteristics, a weighted addition of the components is calculated.

Weighted Addition. A greedy competitive algorithm is used to combine the different components in the image and give the image a value. The greater the value, the more complex the pattern. The weighted addition equation is:

$$\begin{aligned} Weighted\,Addition = {} & \alpha_1 B_{adjacent} + \alpha_2 W_{adjacent} + \alpha_3 B_{diagonal} \\ & + \alpha_4 W_{diagonal} + \alpha_5 B_{single} + \alpha_6 W_{single} \end{aligned} \tag{1}$$

The weights of the paths are based on their impact on the overall measure, and are given as follows:

- Black adjacent path (α_1) = 1.5
- White adjacent path (α_2) = −0.5
- Black diagonal path (α_3) = 1
- White diagonal path (α_4) = 0.5
- Black single block (α_5) = 1
- White single block (α_6) = 0.5

All the white components are weighted lower than the black components as it was established that the black elements are more visually prominent in an image during human observations. The white adjacent path weighting is negative as the presence of white adjacent paths makes the image appear simpler to the human eye. Lastly, the black adjacent path weighting is greater than the black diagonal path weighting as adjacent paths can be multi-directional, and show less symmetry than diagonal paths. This makes the image appear more complex.

Complexity Level. Once the weighted additions for all of the images are calculated, the total range in which they vary is divided into three smaller ranges for a simple, three-level classification. If an image's weighted sum falls within the first range with the smallest values, then the image is deemed *easy* (complexity level 1). If the sum is within the middle range, then the image is deemed *medium* (complexity level 2). Otherwise, the image is deemed *hard* (complexity level 3).

4 Results

The algorithm was tested by comparing each pattern's computationally classified complexity against its corresponding human perceived complexity. The overall accuracy of the algorithm is determined as well as the success rate for each specific complexity level.

The final optimized visual pattern complexity algorithm yields an accuracy of 69%. A confusion matrix of the algorithm performance with respect to human perception of complexity has been drawn for a more thorough performance illustration and is shown in Table 1. The diagonal should contain the highest percentages given that it represents

Table 1. Confusion matrix of the results yielded by the algorithm against human perceived complexities.

		Human Complexity		
		1	2	3
Computational Complexity	1	72 %	14 %	4 %
	2	25 %	74 %	35 %
	3	3 %	12 %	61 %

the correct answers, and the other values should decrease as they move away from the diagonal since these represent the errors in the algorithm.

The results in this matrix illustrate that the desired trend has been achieved. The final algorithm achieves success rates of 72%, 74% and 61%, for *easy*, *medium* and *hard* patterns, respectively as compared to human complexity perception. The success rate of *hard* images is more than 10% points lower than each of the other complexity levels. The results suggest that a further investigation of *hard* images is required, and that the complexity level scale might need to be refined.

The lower success rate could be attributed to the limited factors analysed in this algorithm. The major focus of the current algorithm has been on the number of components in an image and its symmetry; however there are several additional factors that impact a human's perception of image complexity, specifically with regards to *hard* images. Therefore, many factors that could not be included within the scope of this preliminary work, such as the ratio of black space to white space, will be studied in further versions of the algorithm. There are also personal triggers and memories that affect human complexity perception and are individual to each subject. These indeterminable factors therefore impede the success rate, and make defining a concrete division between the three complexity ranges even more challenging.

When examining the images that were incorrectly determined by the algorithm, it was discovered that 92% of the total misclassifications were "small errors" since they were incorrectly determined by only one level. Therefore, only 8% of the misclassified images were considered "large errors" as they were incorrect by two levels. This shows that the algorithm is heading in the right direction and could be improved by adjusting certain factors that overlap between complexity levels.

It may also be worthwhile to consider integrating other pattern complexity techniques including Kolmogorov and fractal image compression. Additionally, a data driven approach using machine learning could be incorporated into the algorithm by using logistic regression to automatically determine the path weightings. Further investigations of these weightings could provide substantial insights into human visual perception. If these future improvements yield closer correlations to the human perceived complexities that need to be mirrored, the number of complexity defining levels could also be increased as this would allow for a more comprehensive, flexible and robust scale of measure.

5 Conclusion

This paper detailed the design, implementation, and testing of an optimized visual pattern complexity determination algorithm. The algorithm obtained 72%, 74%, and 61% correlation to human perception for the *easy*, *medium*, and *hard* complexity levels in the patterns, respectively. Furthermore, only 2% of the total classifications were misclassified by two levels. A critical analysis was provided to analyse the algorithm as well as to suggest future improvements that can be made. The results can still be improved by fine-tuning the algorithm and incorporating additional factors into the measurement. This can also be extended into a relative pattern complexity measure in tasks where relative complexity between two or three patterns displayed. This algorithm has the potential to become an objective measure of visual pattern complexity, which matches human perception. This capability, when implemented in computerized assessment, could improve the usability of these tests for psychometric and cognitive evaluations, specifically for elderly and cognitively impaired subjects.

References

1. Eysenck, M.W., Keane, M.T.: Cognitive Psychology: A Student's Handbook. Taylor & Francis, Abingdon (2000)
2. Aharonson, V., Korczyn, A.D.: Human–computer interaction in the administration and analysis of neuropsychological tests. Comput. Methods Programs Biomed. **73**, 43–53 (2004)
3. Korczyn, A.D., Aharonson, V.: Computerized methods in the assessment and prediction of dementia. Curr. Alzheimer's Res. **4**, 364–369 (2007)
4. Aharonson, V., Halperin, I., Korczyn, A.A.: Computerized diagnosis of mild cognitive impairment. Alzheimer's Dementia **3**, 23–27 (2007)
5. Pi, Y., Lu, J., Liu, M., Liao, W.: Theory of Cognitive Pattern Recognition. INTECH Open Access Publisher, Guangzhou, Guangdong, China (2008)
6. Aharonson, V., Korczyn, A.: Computerized methods for cognitive testing. In: Majd, A., Felder, R.A. (eds.) Eldercare Technology for Clinical Practitioners. Humana Press, New York (2008)
7. Attneave, F.: Some informational aspects of visual perception. Psychol. Rev. **61**(3), 183–193 (1954)

Effects of Background Music on Visual Lobe and Visual Search Performance

Lindong Yang(✉) and Ruifeng Yu

Department of Industrial Engineering, Tsinghua University, Beijing 100084, China
yldl3@mails.tsinghua.edu.cn

Abstract. The present study measured the visual lobe shape and visual search performance of 65 participants under different background music conditions. A 2 (tempo: 40% faster vs. 40% lower) × 2 (volume: 85 dB vs. 55 dB) between-subject factorial design was used in this experiment. Results revealed that the tempo and volume of background music had significant effects on visual lobe area, and the size of visual lobe was largest when the music tempo was fast and the volume was large. As to visual search performance, search time was significantly shorter when music tempo was fast. The obtained results indicated that the tempo, volume as well as subjects' perceived feeling of background music had an influence on visual lobe shape and visual search performance. The conclusions could provide guidelines for the design of visual tasks with background music.

Keywords: Visual search · Visual lobe · Background music · Tempo · Volume

1 Introduction

Background music is everywhere in our daily life. It accompanies us in stores, cars, online games, and even during our work. In the past decades numerous studies have been done to explore the effect of background music on advertising, retailing, manufacturing, sports and so on [1]. The influence of music on human behavior is non-negligible. Background music in stores can adjust the atmosphere to affects the shopping desire [2]. Customers' perception of time was also influenced by background music [3]. McElrea and Standing [4] found that the speed of drinking had a relationship with the tempo of background music. People drank faster with faster music. Chang et al. [5] revealed that slow music improved sleep quality of the insomnia. Background music was also found to relief the feeling of pain [6]. Having analyzed nearly one hundred studies on music's influence on human behavior, Kämpfe et al. [7] concluded that the effect of background music was not uniform, i.e., sometimes it was beneficial, sometimes detrimental, and sometimes no effect on behavior, cognition and emotion. The impact depended on the style, tempo and volume of music as well as listeners' age, personality and music preference [1, 3, 8, 9]. Among those influencing factors, music tempo and volume are two of the most common studied factors [7]. Davies et al. [10] investigated the performance of visual vigilance task with background music and they

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_19

found that performance was enhanced by music. Day et al. [1] found that only music with fast tempo could increase decision making accuracy. In addition, music with fast tempo had positive influence on the speed and accuracy of reading news on computer screen [11]. As to cognition performance, the music's effect was positive for extroverts and negative for introverts [8]. There were also studies concluding that music had negative effects on memory and attention [12–17]. For example, the memory of advertisement pushed by short message was harmed by background music [12]. Brown [18] found a decrease of driving speed with background music during traffic jams.

Although visual search tasks are often accompanied with background music such as driving with car music and playing video games with music, little is known about its influence on visual lobe and visual search performance. Phillips and Madhavan [19] found that background music had a positive effect on visual tasks. Beh and Hirst [20] investigated the effect of music on driving-related tracking and vigilance tasks. The results indicated that the simple tracking task was not affected by music. While the vigilance task performance for central signals was improved under music conditions. The response time for peripheral signals was increased with high-intensity background music. Crust et al. [21] found that background music could facilitate simple visual search task performance, and performance was better under Lyrical Music conditions than that under Instrumental Music conditions. Norbert and Katarina [22], however, found no significant effect on visual search performance.

The present study investigated the effect of background music on visual lobe and visual search performance using Visual Lobe Measurement System (VILOMS) and Visual Search Time Measurement System (VSTMS). The fourth movement of Eine Kleine Nachtmusik from Mozart was selected as the background music. A 2 (music tempo: 40% faster vs. 40% lower) $\times$ 2 (music volume: 85 dB vs. 55 dB) between-subject factorial design with a control group was used in this experiment. After the measurement, the participants finished a questionnaire with 11 adjectives to rate the extent to which each term characterized their feelings of the music on a five-point scale ranging from "little or no such feeling" to "very strong feeling". The conclusions could provide guidelines for the design of visual tasks with background music.

2 Method

2.1 Participants

65 University students aged between 18 and 23 participated in this study. All participants had normal or corrected to normal vision. None of them had any previous experience with such visual search tasks. They were randomly divided into 5 groups under different background music conditions, as shown in Table 1.

2.2 Apparatus and Stimulus

The measurement was divided into 2 sessions. The first session used Visual Lobe Measurement System (VILOMS) to measure visual lobe and the second session used Visual Search Time Measurement System (VSTMS) to measure visual search

Table 1. The experimental settings of the 5 groups.

Group	Volume	Tempo	Participant no.
Control group	No music	No music	14
Experiment group 1	Large	Slow	14
Experiment group 2	Large	Fast	12
Experiment group 3	Small	Fast	13
Experiment group 4	Small	Slow	12

performance. VILOMS could map visual lobe and calculate visual lobe shape indices. While VSTMS measured the reaction time of visual search tasks [23]. The shape indices included 17 indices which were divided into 5 categories: roundness, boundary smoothness, symmetry, elongation, and regularity in addition to visual lobe area and perimeter (For more details, see Chan and So [24]). The volume of the music was measured by an AZ Sound Level Meter (AZ8928; AZ Instrument Corporate, Shenzhen, China). The tempo of the music was controlled by GoldWave (version 5.65, GoldWave Inc., Newfoundland, Canada). The software was run on a Dell laptop with the resolution set at 1366 × 768 pixels and a refresh frequency of 60 Hz. The participants were asked to sit on a chair with adjustable height at a distance of 500 mm from the screen. Their eyes were at the same height as the center of the screen.

For session 1, the stimulus image was rectangular with 419 distractor items (character X) and 1 target item (character O), the size of which was 2 mm × 2 mm. The test field was 21° in width and 16° in height. The background was black and the items were white, as shown in Fig. 1. The target randomly located at 214 possible positions. For each position, the target appeared twice in total, and if the participant correctly identified the position at least once, then the position was considered within the visual lobe. The items were replaced by the masking characters (+) 300 ms after the appearance of the stimulus to ensure only one fixation during each trial. For session 2, the stimulus image was composed of 299 distractors and 1 target with the size of 21° (width) × 16° (height). The distractors and target were the same with experiment 1. The target randomly appeared at 96 preset locations. The items were replaced by the mask (+) until the participant found the target and clicked the mouse.

This study did not choose songs as background music to avoid the influence of lyrics. Another consideration of music selection was the degree of familiarity for participants. The fourth movement of Eine Kleine Nachtmusik from Mozart was selected as the background music. It was written in sonata form with a tempo of about 126 bpm. None of the participants was familiar with this piece of music. The tempo of the music was adjusted for two tempo conditions. For the slow tempo groups, it was 40% slower than the original; for the fast groups, it was 40% faster. The volume for experiment group 1 and 2 was 85 dB while the volume for experiment group 3 and 4 was 55 dB. After the measurement, the participants finished a questionnaire with 11 adjectives to rate the extent to which each term characterized their feelings of the music on a five-point scale ranging from “little or no such feeling” to “very strong feeling”. The selection of the adjectives referred to North et al. [2] and Fried and Berkowitz’s [25] studies.

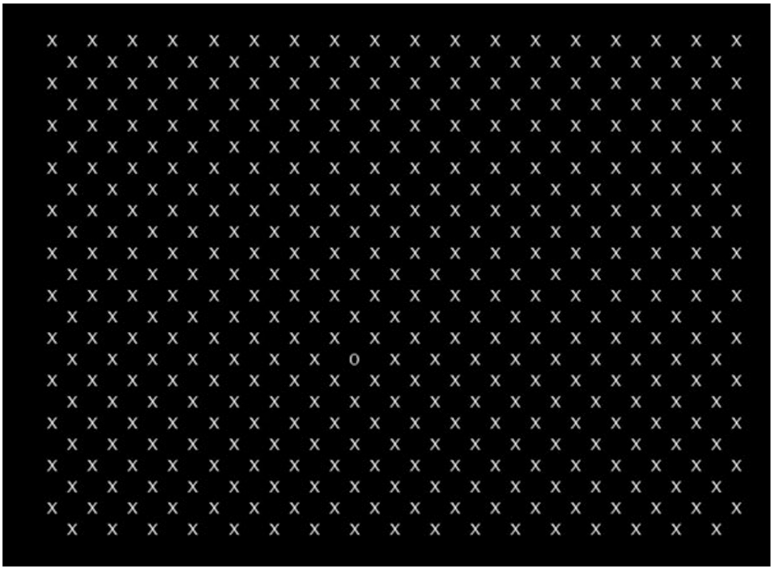


Fig. 1. An example of stimulus image used in session 1.

2.3 Procedure

For session 1, the experimenter first explained the experiment to the participants and showed how to use the software. Then the participants were provided with a 20-minute practice. The practice trials were exactly the same as that of actual experimental tasks. After the practice session was finished, the experiment trial began. For each trial, the stimulus duration was 300 ms and then it was replaced with the mask image. The participants right clicked the position where they saw the target character, after which the next trial began. If the participants did not see the target, then they right clicked any position of the mask image.

Session 2 began after session 1 finished. In trials of session 2, the stimulus image was presented until the participants found the target. The participants right clicked any position once they saw the target and then the stimulus was masked. The participants clicked the "+" where the target appeared. During the entire experiment, the participants had a 2-minute rest after 15 min of the experiment. The software used Fovea Fixation Mechanism (FFM) to ensure the participants gazed at the center of the image at the beginning of each trial (For more details about FFM, see Chan and So [24], Yu and Yang [26]).

3 Results and Discussion

The means and standard deviations (SD) of each group's lobe shape indices are presented in Table 2.

Table 2. The experimental settings of the 5 groups.

Lobe shape Index	Control group		Experimental group 1		Experimental group 2		Experimental group 3		Experimental group 4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Visual lobe perimeter	0.950	0.104	0.965	0.046	0.952	0.086	0.993	0.148	0.970	0.081
Visual lobe area	0.041	0.008	0.049	0.004	0.042	0.007	0.044	0.007	0.048	0.005
Roundness										
Form factor	0.599	0.178	0.653	0.115	0.602	0.137	0.603	0.164	0.655	0.126
Perimeter-area ratio	0.763	0.128	0.805	0.072	0.771	0.092	0.767	0.124	0.805	0.086
P-A ratio of convex Hull	0.908	0.021	0.897	0.028	0.898	0.033	0.886	0.030	0.892	0.020
Area-max. area ratio	0.600	0.102	0.627	0.072	0.621	0.068	0.615	0.074	0.629	0.060
Ratio of radii	0.772	0.069	0.790	0.045	0.786	0.045	0.787	0.052	0.792	0.038
Boundary Smoothness										
Global convex deficiency	0.159	0.097	0.129	0.054	0.145	0.056	0.181	0.130	0.125	0.064
Rugosity	1.169	0.156	1.107	0.070	1.148	0.098	1.167	0.207	1.106	0.094
Spike parameter	0.242	0.052	0.271	0.047	0.243	0.037	0.273	0.042	0.276	0.034
Symmetry										
Vertical vertices symmetry	0.119	0.056	0.094	0.035	0.119	0.073	0.127	0.064	0.089	0.037
Horizontal vertices symmetry	0.093	0.043	0.124	0.139	0.096	0.038	0.086	0.053	0.079	0.025
Left-right area symmetry	0.096	0.072	0.073	0.047	0.125	0.195	0.059	0.037	0.054	0.020
Vertical symmetry of convex hull	0.924	0.064	0.923	0.045	0.898	0.120	0.852	0.239	0.917	0.048
Horizontal symmetry of convex hull	0.917	0.080	0.922	0.029	0.915	0.078	0.920	0.093	0.901	0.035
Elongation										
Length-width ratio	1.153	0.258	1.094	0.137	1.317	0.325	1.189	0.098	1.213	0.099
Regularity										
Boycle-clark index	0.917	0.023	0.930	0.015	0.913	0.028	0.885	0.091	0.928	0.020

The standard deviations of most shape indices for different groups were less than 0.2. Thus the individual difference of each group was not large. Visual lobe perimeter of experiment group 3 was the largest while experiment group 1 had the largest visual lobe area. As to roundness indices, all the five indices had medium to large values, thus the shapes were relatively round. The boundary smoothness indices indicated a low level of boundary smoothness. The values of Vertical Vertices Symmetry, Horizontal Vertices Symmetry, and Left-Right Area Symmetry varied from 0 to 0.13, and Vertical Symmetry of Convex Hull, Horizontal Symmetry of Convex Hull had values of nearly 1, thus it can be concluded that generally the lobes had high levels of shape symmetry. The Length-Width Ratio of the five groups was larger than one, revealing that the lobes were horizontally elongated regardless of background music conditions. The values of regularity index ranged from 0.885 to 0.930, indicating that visual lobe had high level of regularity.

ANOVA results revealed that both the tempo ($F(2, 63) = 6.41$, $p = 0.003$) and the volume ($F(2, 63) = 3.36$, $p = 0.041$) of background music had significant effects on visual lobe area at 95% confidence level. The interaction effect of tempo and volume was not significant. The effects of tempo and volume on visual lobe area were shown in Fig. 2. As can be seen, visual lobe area was smaller when the tempo was faster. The effect of volume on visual lobe area depended on the tempo conditions: when the tempo was slow, visual lobe area had a slight increase when the volume increased; however, the size of visual lobe decreased as the volume increased when the tempo was fast. The effects of background music tempo and volume on other visual lobe indices were not significant.

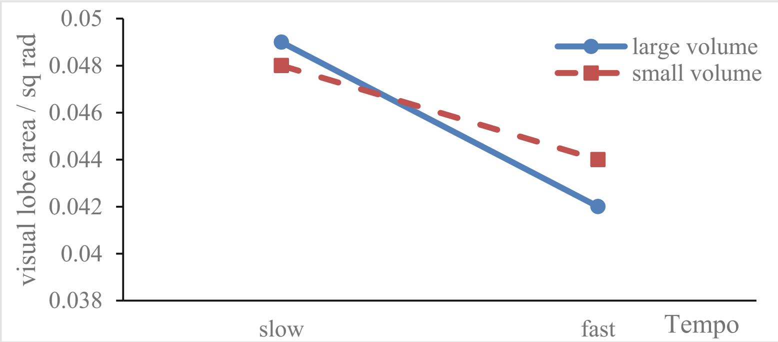


Fig. 2. Visual lobe area of experiment groups.

The five-point rating scales measured the participants' subjective feelings of the music. The rating 1 to 5 represented no such feeling to very strong feeling of the corresponding adjectives. The means and standard deviations of the subjective ratings are presented in Table 3. The highest rating was the feeling "happy", while means of feelings "scared", "violent", "sad" and "angry" was very low, indicating that the music of Mozart could induce positive feelings. ANOVA results revealed that the effects of "happy" ratings on Form Factor ($p = 0.024 < 0.05$), Perimeter-Area Ratio ($p = 0.04 < 0.05$), Area-Max. Area Ratio ($p = 0.037 < 0.05$), Ratio of Radii ($p = 0.021 < 0.05$), Spike Parameter ($p = 0.018 < 0.05$) and Vertical Vertices Symmetry ($p = 0.012 < 0.05$) were significant, while the other ratings' effects on visual lobe indices were not significant. To further analyze the effect of the rating "happy" on the indices, t-test was performed. The rating results were divided into three groups: group 1, little feeling (score 1 and 2), group 2, neutral (score 3), group 3, strong feeling (score 4 and 5) because few participants scored 1 and 5. The t-test results were presented in Table 4. Visual lobe area of group 1 was significantly smaller than that of group 3. The four roundness indices, i.e., Form Factor,P-A Ratio of Convex Hull,Area-max. Area Ratio and Ratio of Radii for group 2 were the largest, and those for group 3 was the smallest. Thus those who felt happy about the music had lower level of roundness. For Global Convex Deficiency, group 1 and group 2 had significantly smaller value than group 3. The Spike Parameter of group 2 was significantly larger than that of group 1 and 3. Thus group 3 had lower level of boundary smoothness than group 1 and 2. As to Vertical Vertices Symmetry, group 3 had significantly largest value, indicating that group 3 had significantly less symmetrical lobe shapes.

For visual search task, the reaction time under different tempo and volume conditions are shown in Table 5. It can be seen that the reaction time was shorter with background music regardless of music tempo and volume. The reaction time was shorter with faster tempo. ANOVA results revealed that the tempo had a significant effect on reaction time ($p = 0.035 < 0.05$), while the effect of volume on reaction time was not significant ($p = 0.101 > 0.05$). The interaction effect of tempo and volume was significant ($p = 0.013 < 0.05$). T-test results indicated that the difference between fast tempo groups and control group was significant, while the difference between fast tempo groups and slow tempo groups and the difference between slow tempo groups and control group were not significant. The interaction effect was shown in Figure. It can be seen that when the tempo was slow, the higher the volume was, the longer it took to find the target, while when the tempo was fast, the higher the volume was, the shorter the reaction time was. It was because tempo had a positive effect on reaction time while volume's effect was negative. When tempo was slow and the volume was large, the negative effect of volume was stronger, leading to a longer reaction time. However, when the tempo was fast and the volume was large, the positive effect of the tempo was strengthened, thus the reaction time was shorter. ANOVA results revealed that no subjective feeling ratings had significant effect on reaction time except the feeling "happy" ($p = 0.037 < 0.05$). Those who had strong feelings of happy had shorter reaction time than those who had little feelings of happy ($p = 0.021 < 0.05$), indicating that the feeling "happy" had a positive effect on search performance.

Table 3. Means and standard deviations of the subjective ratings.

Subjective feelings	Mean	SD
Peaceful	2.60	1.88
Noisy	2.68	1.94
Scared	1.50	0.58
Violent	1.56	0.78
Happy	3.58	1.23
Anxious	2.04	1.02
Depressed	1.70	0.62
Sad	1.80	0.69
Angry	1.66	0.56
Surprised	1.92	0.85
Bored	1.84	1.24

Both the tempo and the volume of background music affected visual lobe area. Lobe area was larger when there was background music. The effect of tempo and volume on visual lobe shape indices was not significant. However, tempo had a significant effect on visual search performance: reaction time was shorter when the tempo was faster. Beh and Hirst [20] found that for simple visual search task, the volume of background music did not affect search performance [20]. The task of this study was simple [27], consequently, reaction time was not affected by volume. Visual lobe area of fast tempo groups was smaller than that of slow tempo groups, while the reaction time of fast tempo groups was shorter, implying that larger visual lobe area did not always result in better search performance. On the other hand, it indicated that the effect of background music on search performance was not the result of its effect on visual lobe area. The findings also revealed that background music had an influence on search

Table 4. T-test results of "happy" rating's effects on lobe indices.

Lobe shape index	Group 1 and 2		Group 1 and 3		Group 2 and 3	
	t	p	t	p	t	p
Visual lobe area	0.30	0.769	1.88	0.049*	1.38	0.192
Form factor	−0.76	0.463	2.20	0.036*	2.65	0.016*
P-A ratio of convex hull	−1.03	0.320	1.13	0.284	2.56	0.026*
Area-max. area ratio	−1.52	0.150	2.14	0.046*	3.76	0.002**
Ratio of radii	−1.94	0.073	2.18	0.042*	4.24	0.001**
Global convex deficiency	0.73	0.480	−3.10	0.004**	−3.36	0.002**
Spike parameter	−2.86	0.013*	−0.06	0.952	3.52	0.003**
Vertical vertices symmetry	0.63	0.544	−2.27	0.035*	−3.90	0.001**

Note: * $p < 0.05$, ** $p < 0.01$

Table 5. Means and standard deviations of reaction time for each group.

Groups	Control group		Experimental group 1		Experimental group 2		Experimental group 3		Experimental group 4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
RT	1.564	0.163	1.276	0.062	1.368	0.131	1.187	0.132	1.519	0.055

performance, but the influence was small. This is consistent with previous studies [28]. The subjective feeling "happy" had significant effects on visual lobe area, roundness and symmetry. Those who felt "happy" with background music had smaller visual lobe area and lower levels of lobe roundness and symmetry, but better search performance.

4 Conclusion

The present study measured the visual lobe shape and visual search performance of 64 participants under different background music conditions using Visual Lobe Measurement System (VILOMS) and Visual Search Time Measurement System (VSTMS). Results revealed that the tempo and volume of background music had significant effects on visual lobe area. The effects of background music tempo and volume on other visual lobe indices were not significant. As to visual search performance, search time was significantly shorter when music tempo was fast. Music volume did not significantly influence search time. The highest rating of perceived feeling was "happy", and the feeling "happy" had significant effects on visual lobe area, roundness, boundary smoothness, symmetry, and search performance. The obtained results indicated that the tempo, volume as well as subjects' perceived feeling of background music had an influence on visual lobe shape and visual search performance. The conclusions could provide guidelines for the design of visual tasks with background music.

Acknowledgments. This work was supported by a grant from the Natural Science Foundation of China [project number 71471098].

References

1. Day, R.F., Lin, C.H., Huang, W.H., Chuang, S.H.: Effects of music tempo and task difficulty on multi-attribute decision-making: an eye-tracking approach. Comput. Hum. Behav. **25**(1), 130–143 (2009)
2. North, A.C., Hargreaves, D.J., McKendrick, J.: The effects of music on atmosphere in a bank and a bar. J. Appl. Soc. Psychol. **30**(7), 1504–1522 (2000)
3. Yalch, R., Spangenberg, E.: Effects of store music on shopping behavior. J. Consum. Market. **7**(2), 55–63 (1990)
4. McElrea, H., Standing, L.: Fast music causes fast drinking. Percept. Mot. Skills **75**(2), 362 (1992)
5. Chang, E.T., Lai, H.L., Chen, P.W., Hsieh, Y.M., Lee, L.H.: The effects of music on the sleep quality of adults with chronic insomnia using evidence from polysomnographic and self-reported analysis: a randomized control trial. Int. J. Nurs. Stud. **49**(8), 921–930 (2012)
6. Vaajoki, A., Pietilä, A.M., Kankkunen, P., Vehviläinen-Julkunen, K.: Effects of listening to music on pain intensity and pain distress after surgery: an intervention. J. Clin. Nurs. **21**(5–6), 708–717 (2012)
7. Kämpfe, J., Sedlmeier, P., Renkewitz, F.: The impact of background music on adult listeners: a meta-analysis. Psychol. Music **39**(4), 424–448 (2011)
8. Furnham, A., Allass, K.: The influence of musical distraction of varying complexity on the cognitive performance of extroverts and introverts. Eur. J. Personal. **13**(1), 27–38 (1999)
9. Huang, R.H., Shih, Y.N.: Effects of background music on concentration of workers. Work **38**(4), 383–387 (2011)
10. Davies, D.R., Lang, L., Shackleton, V.J.: The effects of music and task difficulty on performance at a visual vigilance task. Br. J. Psychol. **64**(3), 383–389 (1973)
11. Kallinen, K.: Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. Comput. Hum. Behav. **18**(5), 537–551 (2002)
12. Rau, P.L.P., Chen, D.: Effects of watermark and music on mobile message advertisements. Int. J. Hum.-Comput. Stud. **64**(9), 905–914 (2006)
13. De Groot, A.: Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. Lang. Learn. **56**(3), 463–506 (2006)
14. Furnham, A., Bradley, A.: Music while you work: the differential distraction of background music on the cognitive test performance of introverts and extraverts. Appl. Cogn. Psychol. **11**(5), 445–455 (1997)
15. Smith, C.A., Morris, L.W.: Differential effects of stimulative and sedative music on anxiety, concentration, and performance. Psychol. Rep. **41**(3_suppl), 1047–1053 (1977)
16. Nittono, H.: Background instrumental music and serial recall. Percept. Mot. Skills **84** (3_suppl), 1307–1313 (1997)
17. Woo, E.W., Kanachi, M.: The effects of music type and volume on short-term memory. Tohoku Psychol. Folia **64**, 68–76 (2006)
18. Brown, I.D.: Effect of a car radio on driving in traffic. Ergonomics **8**(4), 475–479 (1965)
19. Phillips, R.R., Madhavan, P.: The effect of distractor modality and processing code on human-automation interaction. Cogn. Technol. Work **13**(4), 233–244 (2011)
20. Beh, H.C., Hirst, R.: Performance on driving-related tasks during music. Ergonomics **42**(8), 1087–1098 (1999)
21. Crust, L., Clough, P.J., Robertson, C.: Influence of music and distraction on visual search performance of participants with high and low affect intensity. Percept. Mot. Skills **98**(3), 888–896 (2004)

22. Jaušovec, N., Habe, K.: The influence of auditory background stimulation (mozart's sonata K.448) on visual brain activity. Int. J. Psychophysiol. **51**(3), 261–271 (2004)
23. Chan, A.H., Tang, N.Y.: Visual lobe shape and search performance for targets of different difficulty. Ergonomics **50**(2), 289–318 (2007)
24. Chan, A.H., So, D.K.: Measurement and quantification of visual lobe shape characteristics. Int. J. Industr. Ergon. **36**(6), 541–552 (2006)
25. Fried, R., Berkowitz, L.: Music hath charms… and can influence helpfulness. J. Appl. Soc. Psychol. **9**(3), 199–208 (1979)
26. Yu, R.F., Yang, L.D.: Age-related changes in visual lobe shape characteristics and their relationship to visual search performance. Ergonomics **57**(9), 1300–1314 (2014)
27. Chan, A.H.S., So, D.K.T.: Shape characteristics of useful field of view and visual search time. Le travail Humain **70**(4), 343–367 (2007)
28. Kämpfe, J., Sedlmeier, P., Renkewitz, F.: The impact of background music on adult listeners: a meta-analysis. Psychol. Music **39**(4), 424–448 (2011)

The Sound of Violin: Quantifying and Evaluating the Impact on the Performer's and Near Audience's Ear

Pedro Salvadores Palacio(✉) and Mercedes de Barrios Carro

Applied Acoustic Laboratory,
Universidad de León, Campus de Vegazana, 24071 León, Spain
{pedro.salvadores,m.debarrios}@unileon.es

Abstract. The energy in the form of music created by a musical instrument has two subjective protagonists: the performer and the audience. Both of these interact and receive given sound energy in different ways by means of the natural transducer, which makes up our auditory system. From the perspective of sound, this study is determined by the choice of a group of selected musical pieces that set the reference of sound energy values. Given pieces are of individual, collective and learning-aimed interpretation. The aim of this study, based on the previous statements, has two focuses. Firstly, determining, quantifying and evaluating the impact on the performer's and near audience's ear, caused by sound energy generated by various sized violins, when interpreting the chosen musical pieces. Secondly and lastly, foreseeing the impact of given sound energy on the audition capacity of the subjective protagonists of this work in the medium and long term.

Keywords: Human factors · Violin · Sound · Sound level meter · Dosimeter · Measurement · Sound pressure · Acoustic pressure · Sound energy

1 Introduction

There are not many studies about the energy produced by musical instruments. In fact, most studies focus on the quality of sound of musical instruments, the aesthetic effects they generate and, in a way, the spiritual sensations they produce in us, when being heard by an audience. Other studies have looked into the performing spaces, and in other occasions on its effects on the ear. Jurgen Meyer's studies on different aspects of musical instruments are worth emphasizing, but few studies go in depth into the energy created by the instruments. This study thus hopes to give an answer to this energy matter applied to the violin, taking into account that the violin is used in different sizes (1/4, 2/4, 3/4 and 4/4). However, in this study we will only consider the following sizes: 4/4, 3/4 and 1/4. The reason why we do not take into account the 1/2 size is simply that the energy it generates is very similar to a 1/4 size violin.

All in all, in this study, we are going to attempt to determine, from a scientific point of view, the reference values for the energy generated by violins. It is necessary to keep in mind that given energy measurement is not possible without taking into account our natural sensor/transducer, the ear. As a result, we see ourselves forced to determine the

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_20

areas in which there is greater sound incidence: the performer and the near audience. This is going to permit us to determine the place we consider the most appropriate to carry out the energy analysis.

2 Rehearsals and Equipment

We carried out three rehearsals with the participation of two violinists. These took place in the Industrial Engineering School of Leon, in the Auditorium of Leon and in the Music Conservatory of Leon (Spain). The violin performers were Blanca and Carmen.

Next, we move on to outline the equipment used. In the first rehearsal, four SC310 sound level meters, a DC112 dosimeter, and two professional size (4/4) violins are included. In the second rehearsal, we used one DC112 dosimeter and all of the instruments from a Symphony Orchestra. Finally, in the third rehearsal, the equipment used included a SC310 sound level meter, a DC 112 dosimeter, a 1/4 size violin, a 3/4 size and a 4/4 size professional level violin.

The characteristics of the dosimeter and the sound level meter are the following ones: The SC310 sound level meter is a type 1 integrating sound level meter that can work either as a sound level meter or as a real-time spectral analyzer that works through third-octave and Octave Band, with class type 1 filters. The DC112 is a high benefit dosimeter that enables simultaneous measurements to take place, in order to measure all the necessary parameters needed to evaluate the sound exposure (SNR, HML and octaves). Not only does this device measure the equivalent level with A and C [L_{At}, L_{Ct}] frequency weightings, but also, it simultaneously carries out a frequency analysis in real time by octave bands from 63 Hz to 8 kHz (octaves method). Furthermore, it allows the choice of frequency's weighting applied to given analysis. Finally, it (without doubt) allows as well the peak level with C frequency weighting [L_{Cpeak}].

3 Procedure

In order for the rehearsals to take place, we approached the procedure in the following manner.

Firstly, we carried out the first rehearsal in three phases, in the assembly hall of the Industrial Engineering School of Leon. In the first block, the violinists performed individually the same piece (once). Thus, we took three measurements. 'G major scale with arpeggios' was the selected piece. The performers used given piece to warm up and improve their ability. Blanca was the first to perform in this block, in a not very fast tempo. Carmen was then next to performing this piece, in a fast tempo, and then once again a third time, even faster. In the second block, the violinists performed the same piece individually and then in a joint way (once). Thus, we took again three measurements. The piece selected was Bach's 'Violin Partita 2 Allemande'. Lastly, in the third block, the violinists performed individually and then in a joint manner (once) the piece 'Rode Caprice number 5'. Thus, we took three new measurements in this phase (Fig. 1).

Fig. 1. Carmen paying during the recording at the assembly hall of the Industrial Engineering School of Leon

Next, the second rehearsal took place with the Odon Alonso Symphony Orchestra, in the final rehearsal prior to the Christmas concert, in which both the violinists played. The pieces performed were Strauss' 'Waltzes and Polkas' (Fig. 2).

Fig. 2. Blanca paying during the recording in the Auditorium of the Professional Conservatory of music of Leon

Finally, the third rehearsal took place in the Auditorium of the Professional Conservatory of music of Leon, and was divided in three blocks, all performed by one violinist, Blanca. The first block was performed with a 1/4 violin (used to teach children of up to 9–10 years how to play the violin). The piece performed was 'Kücher Concertino op. 11', which was played three times at the tempo and with the cadence that a 1st or 2nd year student of violin of an elemental degree would play. The second block was performed with a 3/4 violin (used to teach children of up to 10–12 years how to play the violin). The piece played was Kücher Concertino op. 11’, which was played three times at the tempo and with the cadence that a 3rd or 4th year student of violin of an elemental degree would play. At last, the third block was performed with a 4/4 professional violin. The piece played was ‘Portnoff Concertino op.13’, which was played three times at the tempo and with the cadence that a violin professional degree student would play.

4 Instrument’s Calibration and Setting

Now, we are going to describe the way in which we calibrated and set the instruments.

Calibration: We calibrated each of the instruments before and after taking the measurement.

Setting of the sound level meters: the sound level meters were located at a microphone height of 1500 mm with respect to the ground and to the violin’s bridge. However, there was one exception for this rule. We positioned one of the sound level meters in the rehearsal that took place in the assembly hall of the Industrial Engineering School of Leon 16.5 m away from the violin, at one end of the Assembly hall. In the following images, we can see a clear view of the setting (Fig. 3):

Fig. 3. Location of the Sound Level Meters during the recording at the hall of the Industrial Engineering School of Leon

Dosimeter setting: The way in which we positioned the dosimeter’s microphone was by fastening it to the violinist’s shirt collar, to a distance of approx. 100 mm from the ear. This had the added difficulty of the positioning of the violin, which made its fixing a bit complicated. In the second rehearsal, with the Odon Alonso Symphony Orchestra the violinists were positioned in the center of the orchestra (with prior authorization of the conductor), as can be seen in the image (Fig. 4).

Fig. 4. Location of the Dosimeter microphone during while recording at the hall of the Industrial Engineering School of Leon

5 Dosimeter. Measurements in the First Rehearsal

The first rehearsal took place in three blocks, in the assembly hall of the Industrial Engineering School of Leon. The musical pieces performed were the following: ‘G major scale with arpeggios’, Bach’s ‘Violin Partita 2 Allemande’ and Rode’s ‘Caprice n5’. The performance of ‘G major scale with arpeggios’ was done by increasing the

Table 1. Dosimeter DC112. Measurements in the first rehearsal.

Phase	L_{Aeqt}	L_{Ceqt}
1-Scale-Blanca	89.0	89.5
2-Scale-Carmen	90.0	90.1
3-Scale-Blanca	91.2	91.2
4-Bach-Carmen	91.0	91.2
5-Bach-Blanca	91.5	92.0
6-Bach-Carmen&Blanca	92.5	93.1
7-Rode-Blanca	92.3	92.8
8-Rode-Carmen	95.2	95.4
9-Rode-Carmen&Blanca	94.5	94.9

Table 2. Dosimeter DC112. Octave band measurements in the first rehearsal.

Phase	63	125	250	500	1000	2000	4000	8000
1-Scale-Blanca	53.7	54.2	74.6	83.7	87.2	79.4	74.2	63.0
2-Scale-Carmen	43.1	47.3	76.6	83.8	86.5	84.3	77.6	63.1
3-Scale-Blanca	44.2	49.2	75.9	84.2	87.6	86.3	76.6	66.1
4-Bach-Carmen	43.9	48.7	77.0	84.8	87.8	85.4	77.1	64.9
5-Bach-Blanca	43.8	45.3	77.3	87.5	88.1	85.2	74.6	60.1
6-Bach-Carmen&Blanca	42.8	42.9	78.1	88.4	90.8	82.1	75.8	63.2
7-Rode-Blanca	44.0	44.5	74.7	86.9	90.8	83.3	75.1	63.6
8-Rode-Carmen	46.5	47.4	79.1	91.0	90.8	88.8	86.0	67.7
9-Rode-Carmen&Blanca	44.9	47.4	79.3	90.1	91.4	87.2	82.8	65.0

speed in each rehearsal and in a loop. That is, when the performer would finish the 'Scale', she then repeated it straightaway from the beginning until the 3-minute cycle was completed. We can see the results obtained in the following tables (Tables 1 and 2, Figs. 5, 6, 7 and 8):

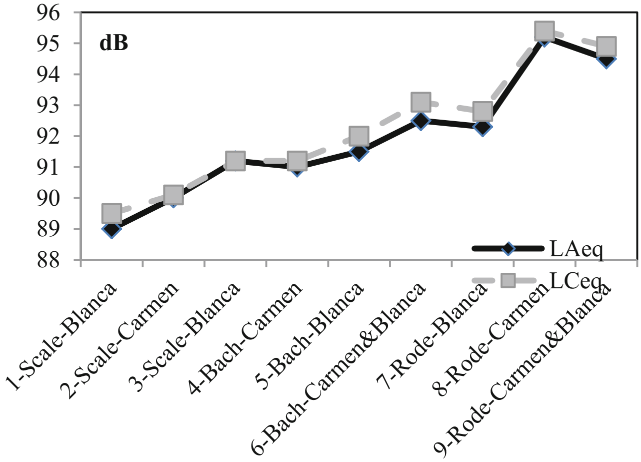


Fig. 5. Total Sound level of all the phases in the first rehearsal

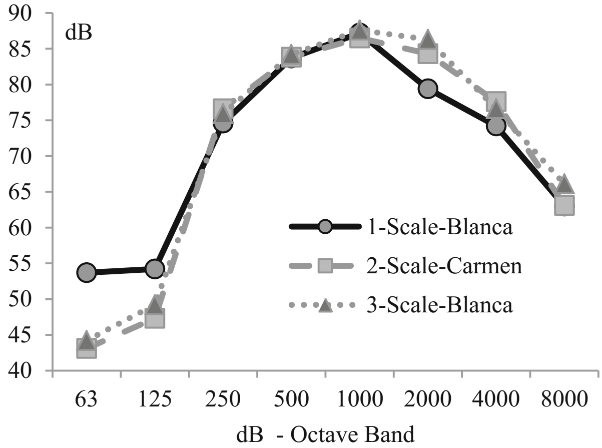


Fig. 6. Octave band Sound level in first, second and third phase of the first rehearsal.

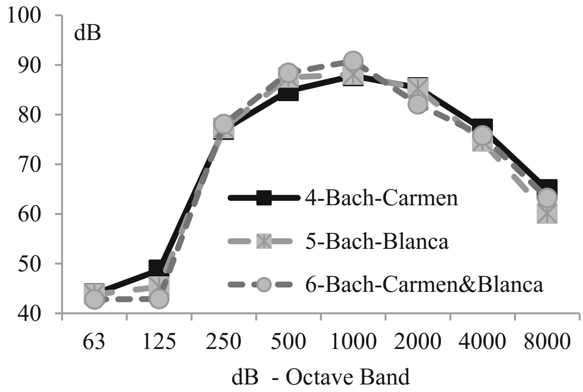


Fig. 7. Octave band Sound level in forth, fifth and sixth phase of the first rehearsal.

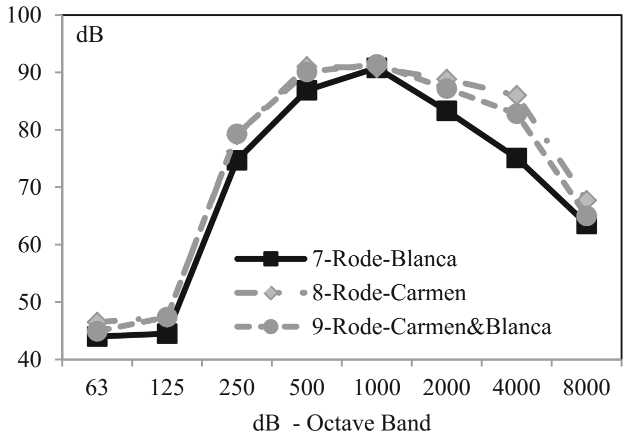


Fig. 8. Octave band Sound level in seventh, eighth and ninth phase of the first rehearsal

6 Dosimeter. Measurements in the Second Rehearsal

The second rehearsal took place in one block, in the auditorium of Leon, in the final rehearsal prior to the Christmas concert. The orchestra interpreted Strauss' 'Waltzes and Polkas', having the following results (Table 3, Figs. 9 and 10):

Table 3. Dosimeter DC112. Octave band measurements in the second rehearsal. The concert

Phase	L_{Aeq}	63	125	250	500	1000	2000	4000	8000
Concert	87,4	74.9	79.7	82.5	83.4	84.1	80.8	70.6	60.5

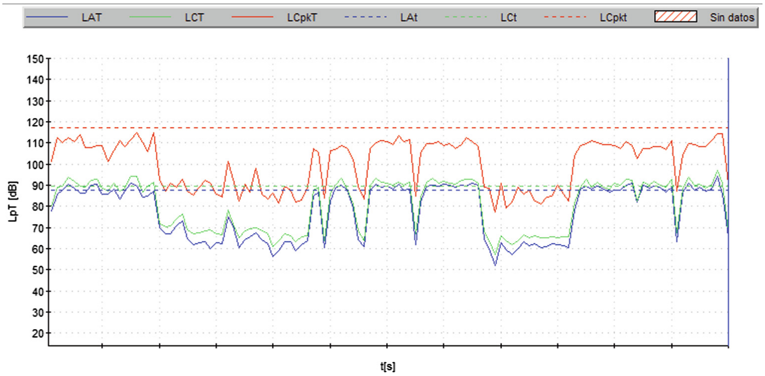


Fig. 9. Sound level in second rehearsal. Concert. It is measure per second, and shows the concer time and de pause time during the concert.

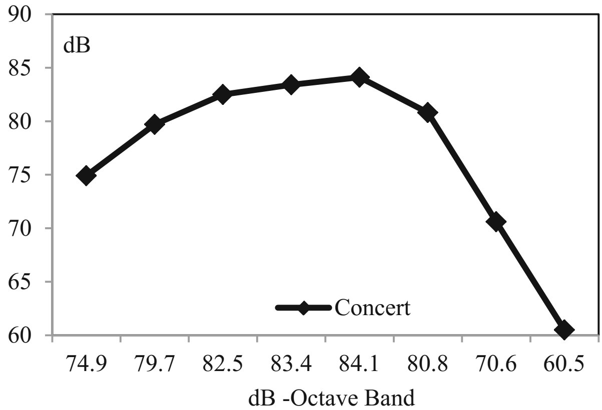


Fig. 10. Octave band Sound level in second rehearsal. Concert

7 Dosimeter. Measurements in the Third Rehearsal

The third rehearsal took place in three blocks, in the Profesional Conservatory of Music of Leon's Auditorium. We employed the size of the violin used as a reference for each block: we carried out the first block with a 1/4 violin, the second with a 3/4 one and the third on with a professional violin of 4/4. The musical pieces performed are the following: Kücher Concertino op. 11 and Portnoff Concertino op. 13. The results obtained are the following (Tables 4 and 5, Figs. 11 and 12):

Table 4. Dosimeter DS112. Measurements in the third rehearsal.

Phase	L_{Aeq}	L_{Ceq}
1-Violin 1/4	87.2	87.8
2-Violin 3/4	96.8	97.2
3-Violin 4/4	97.4	97.3

Table 5. Dosimeter DC112. Octave band measurements in the third rehearsal.

Phase	63	125	250	500	1000	2000	4000	8000
1-Violin 1/4	47.4	59.6	72.6	75.9	67.3	51.5	46.8	41.8
2-Violin 3/4	45.8	42.9	83.2	90	90.7	84.9	81.7	62.1
3-Violin 4/4	48.6	57.3	80.7	91.6	90.3	92.5	87.6	74.3

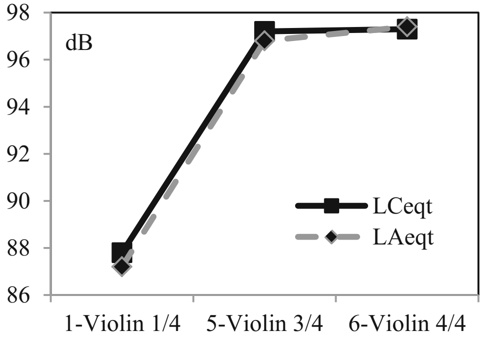


Fig. 11. Total Sound level of all the phases in the first rehearsal

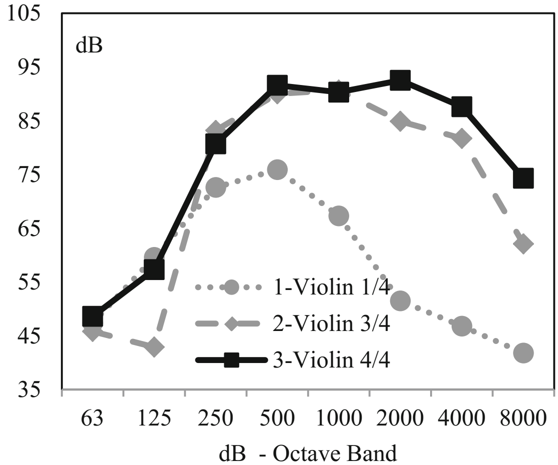


Fig. 12. Octave band Sound level in first, second and third phase of the third rehearsal.

8 Sound Level Meter. Measurements in the First Rehearsal

The first rehearsal took place in three blocks, in the assembly hall of the Industrial Engineering School of Leon. The performers interpreted the following pieces:

'G major scale with arpeggios', Bach's 'Violin Partita n2 Allemande' and Rode's 'Caprice n5'. The most representative value, which we use as a reference, is rehearsal 6, in which both violinists play Bach's 'Violin Partita n2 Allemande'. Here we provide the data obtained (Table 6, Fig. 13):

Table 6. Sound level meter. Measurements in the first rehearsal.

Position	L_{At}	L_{Ct}	L_{Cpeak}
Right	74,3	75,2	93,6
Central	74,2	75,1	93,4
Left	74,6	75,6	93,8

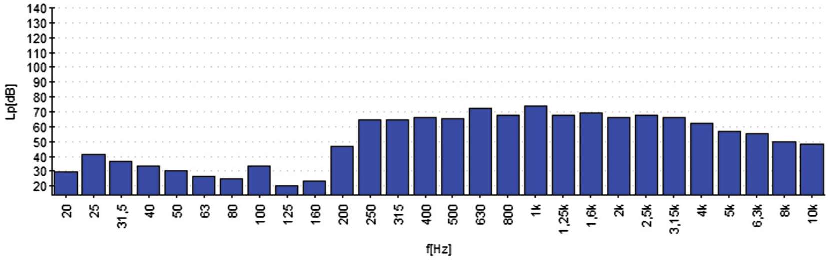


Fig. 13. Octave band Sound level in the first rehearsal. The interval between 20 dB to 200 dB shows the background noise.

9 Sound Level Meter. Measurements in the Second Rehearsal

We decided not to consider this measurement, due to the difficulty of receiving the sound generated by both violins inside the orchestra.

10 Sound Level Meter. Measurements in the Third Rehearsal

The third rehearsal took place in three blocks, in the Professional Conservatory of Music of Leon's Auditorium. We employed the size of the violin used as a reference for each block: we carried out the first block with a 1/4 violin, the second with a 3/4 one and the third on with a professional violin of 4/4. The musical pieces performed are the following: Kücher Concertino op. 11 and Portnoff Concertino op. 13. The results obtained are the following (Table 7):

Table 7. Sound level meter. Measurements in the third rehearsal.

Phase	L_{Aeq}
1-Violin 1/4	68,6
2-Violin 3/4	73,6
3-Violin 4/4	75,5

11 Conclusions

When we started this study on the sound energy generated by violins, the available references we had permitted us to infer estimated sound energy values. However, subsequent to this study, we have been able to verify that the real energy produced is a lot higher than the expected one. Another important factor we have had to decide upon is the musical pieces performed. This is important because, when performed, these pieces generate and determine the sound energy. The pieces chosen for our study have the characteristic of being not too fast, as we were looking for an energy reference based on minimums, and not the maximum energy this instrument can produce.

As a further developing of this point, it is interesting to emphasize how the chosen pieces have an apparently low sound level. In order to see this, it is sufficient to listen to Bach's 'Violin Partita n2 Allemande', which is a piece characterized by the stillness and peacefulness it transmits, in contrast to pieces such as Paganini's 'Caprice n24'. It is also important to point out that given decision of pieces and of basing our reference on minimums has to do with the fact that the violin is not an especially noisy instrument, in comparison with percussion instruments, for instance.

Impact on the Performer's Ear. If we analyze the results provided in the rehearsals, we can infer that if we play given pieces only for an hour each day, we reach LEX8 h levels superior to 80 dB (A). As a result, the final conclusions from an energy point of view are the following ones:

In the measurements, the minimum L_{Aeqt} value obtained was 87.2 dBA in a 1/4 violin, which children between 4 and 10 years old usually use when studying violin.

In one of the cases, we obtain a L_{Aeqt} value of 89dBA. Given value (as can be seen in the graph) takes into account stops in the performance and thus we can infer that the concert's net value is indeed greater. However, this will be a matter of another study.

In the remaining cases, the L_{Aeqt} value goes between a 90 dBA minimum and a 97.4 maximum.

Impact on the Near Audience´s Ear. The obtained results in all of the cases are energy levels between 68 dBA and 76 dBA, which highlight the fact that, starting from 1.5-meter distances, the energetic impact of one or two violins is minimum.

Reference

Meyer, J.: Acoustics and the Performance of Music. Springer Science+Bussines Media, Braunschweig (2009)

Neuroergonomics Theory and Design

EEG-Engagement Index and Auditory Alarm Misperception: An Inattentional Deafness Study in Actual Flight Condition

Frédéric Dehais[1(✉)], Raphaëlle N. Roy[1], Gautier Durantin[1], Thibault Gateau[1], and Daniel Callan[1,2]

[1] Université Fédérale de Toulouse, ISAE-SUPAERO, DCAS, 10, Avenue Édouard Belin, 31400 Toulouse, France
{Frederic.Dehais,Raphaelle.Roy,Gautier.Durantin, Thibault.Gateau}@isae.fr, dcallan@nict.go.jp

[2] Center for Information and Neural Networks CiNet, National Institute of Information and Communications Technology NICT, Osaka University, Osaka, Japan

Abstract. The inability to detect auditory alarms is a critical issue in many domains such as aviation. An interesting prospect for flight safety is to understand the neural mechanisms underpinning auditory alarm misperception under actual flight condition. We conducted an experiment in which four pilots were to respond by button press when they heard an auditory alarm. The 64 channel Cognionics dry-wireless EEG system was used to measure brain activity in a 4 seat light aircraft. An instructor was present on all flights and in charge of initiating the various scenarios to induce two levels of task engagement (simple navigation task vs. complex maneuvering task). Our experiment revealed that inattentional deafness to single auditory alarms could take place as the pilots missed a mean number of 12.5 alarms occurring mostly during the complex maneuvering condition, when the EEG engagement index was high.

Keywords: Inattentional deafness · Auditory alarm misperception · EEG engagement index · Real flight conditions

1 Introduction

Neuroergonomics is an exciting field of research that has gained momentum over the last decade. It promotes multidisciplinarity and the implementation of brain imaging devices to understand cognitive functioning in complex real-life situations [1]. Neuroergonomics opens promising perspectives for applied disciplines such as Human Factors and Ergonomics. Generally, these latter emphasize behavioral and subjective approaches to address human performance issues. Whereas the scientific contribution of these disciplines is of great importance, they may appear limited to provide explanations for more complex phenomena that require the investigation of the cerebral activity. This is the case for auditory alarm misperception that has been shown to be involved in several aircraft accidents [2, 3]. Indeed, the dominant theory to account for this phenomenon is that pilots consciously choose to ignore these warnings due to cognitive

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_21

biases [4] or poor design issues [5, 6]. Without denying the importance of these findings, recent Neuroscientific studies have postulated alternative perceptual and attentional explanations, known as the inattentional deafness hypothesis. There is a corpus of evidences that unexpected sounds may fail to reach awareness when highly engaged in visual tasks [7–9]. In these contexts, the visual modality may take over hearing via gating mechanisms at the visuo-auditory integrative [7, 10, 11] or higher levels [12].

Since flying is an activity that mainly solicits visual processing, inattentional deafness is more likely to take place in the cockpit, thus leading to auditory alarm neglect. Some experiments conducted in flight simulators have shown the existence of this phenomenon during the landing phase [3, 13, 14]. More recently, an electroencephalography (EEG) study involving a critical scenario in a flight simulator (i.e. approach with burning engine and smoke in the cabin) yielded a high rate of auditory alarm misperception. The EEG analyses revealed that misses were associated with lower N100 and P300 amplitude than hits [15], confirming the existence of an early and unconscious gating mechanisms. Interestingly enough, an EEG experiment conducted under real flight conditions disclosed that a reduction in phase resetting in alpha and theta band frequencies was a neural signature of inattentional deafness [16]. These studies demonstrate the importance of adopting a Neuroergonomics approach to underpin the neural mechanisms at the core of human performance and erroneous behavior.

There is still the need to understand the causal factors that promotes the occurrence of inattentional deafness. Excessive cognitive workload and limited resources theories are generally thought to be the main cause of such auditory attention impairment [4, 9]. However, cognitive workload should not be viewed as the resultant of an external demand applied on an individual passively adapting to it, but rather as an active process that depends on the human operator's level of engagement. Thus, the allocation of cognitive resources has to be considered as the product of the level of task demand by the level of task engagement. We state that level of engagement mainly depends on task utility/reward, associated risk and time on task (i.e. sunk cost effect - see [17, 18]). This explains why auditory misperception is more likely to occur during the landing (i.e. final destination) even during visual flight rules conditions [19, 20].

In the present study, we intend to investigate auditory misperception with EEG in more ecological settings than previous research [15, 19–21]. The main objective was to show that inattentional deafness could take place in the cockpit, especially during high level of engagement episodes. We manipulated the flying task to induce two levels of engagement. Additionally, we computed an EEG index defined by [22] to verify that our conditions were effectively leading to different levels of task engagement [22–24]. Eventually, an additional motivation was to show the feasibility of extracting this index with a dry electrodes system under ecological settings.

2 Material and Method

2.1 Participants

Four healthy male pilots (97.5 mean flight hours), participated in the study after they gave their informed written consent. All reported normal auditory acuity and normal or corrected-to-normal vision. The experimental protocol was approved by the European Aviation Safety Agency (EASA permit to fly approval number: 60049235).

2.2 Experimental Scenario

The experiment was conducted at Lasbordes airfield (Toulouse, France) in which the pilots were to respond by button press when they heard an auditory alarm (chirp sound). Two hundred and thirty stimuli were presented every 10 to 15 s. The experiment lasted approximately 1 h (i.e. from take-off to final taxiing). During the experiment, an instructor manipulated two levels of task engagement. The low task engagement condition involved simple navigation above 1700 feet and the high task engagement condition involved several complex maneuvering exercises such as simulated engine failure, off field emergency landing procedures, and low altitude circuit patterns. There were an equal number of auditory alarms between the two conditions.

2.3 Aircraft

The ISAE-SUPAERO DR400 light aircraft was used for the purpose of the experiment (Fig. 1). It was powered by a 180HP Lycoming engine and was equipped with classical gauges, radio and radio navigation equipment, and actuators such as rudder, stick, thrust and switches to control the flight. The participant was placed on the left seat and was equipped with the EEG dry electrode system. A switch button was attached to the stick to collect pilots' response. The participants wore a Clarity Aloft headset that was used to trigger auditory stimuli from a PC via an audio cable. The participant could still communicate with the other crew members, air traffic controllers when he received auditory alarm. The safety pilot was an ISAE-SUPAERO flight instructor. He was right seated and had the authority to stopping the experiment and taking over the control of the aircraft for any safety reason. The backseater was the experimenter: his role was to set the sensor, to trigger the experimental scenario and to supervise data collection.

Fig. 1. Left: a participant equipped with the 64-chanel Cognionics dry electrodes system. The aircraft was a Robin DR400. Right: the response button attached to the stick.

2.4 Neurophysiological Measurements and Analyses

The 64 channel Cognionics dry-wireless EEG system was used to measure brain activity. EEGLAB was used for analysis of the EEG data for each pilot. The continuous EEG data was filtered between 1–30 Hz, underwent automatic channel rejection, and was cleaned using automatic subspace reconstruction. For each condition, we computed the following EEG engagement index [22, 25]: average power in beta [13 30 Hz]/(average power in alpha [8 12 Hz] + average power in theta [4 8 Hz]).

3 Results

3.1 Behavioral Results

Our experiment revealed that inattentional deafness to auditory alarm could take place as the pilots missed a mean number of 12.5 alarms (SD = 5.6) with 71,2% of them (SD = 11%) occurring during the high engagement flying condition (Fig. 2).

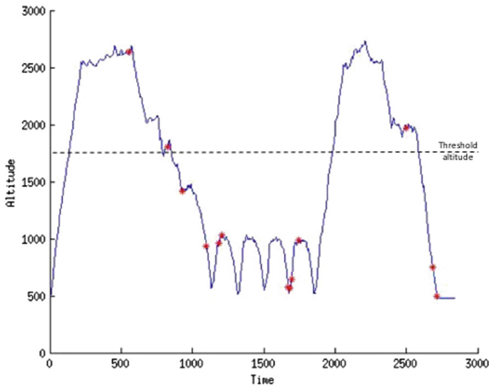


Fig. 2. Altitude in function of time. Red stars indicate misses. The threshold altitude was set at 1700 feet (dashed black line). Note that this latter participant missed 9 alarms out of 12 below this threshold (i.e. in the high task engagement condition)

3.2 EEG Results

As illustrated by Fig. 3 for two participants, the computed EEG engagement index appeared to be related to this phenomenon, as its mean was higher during the high engagement condition (Fronto-central area: 0.48, SD = 0.07; Parietal area: 0,46, SD = 0.07) than during the low engagement condition (Fronto-central area: 0.44, SD = 0.03; Parietal area: 0.51, SD = 0.12). Please report to Table 1 for more details.

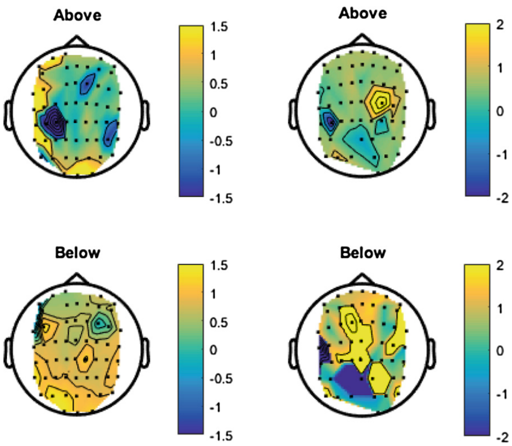


Fig. 3. Fontro-central region of interest: engagement ratio power spectral results for 2 participants above (i.e. low engagement condition) and below (i.e. high engagement condition) the altitude threshold.

Table 1. EEG engagement index for all the participants in the fronto-central and parietal regions of interest above and below the altitude threshold.

Fronto-central		Parietal	
Above	Below	Above	Below
0,44	0,58	0,50	0,64
0,42	0,47	0,51	0,58
0,48	0,50	0,42	0,43
0,42	0,39	0,41	0,38

4 Discussion

The objective of this study was two-fold: first, we wanted to show that auditory alarm misperception could occur under engaging flying conditions. Second, we aimed at measuring the neural correlates of this phenomenon. This was challenging as we collected data in highly ecological conditions with a dry electrodes device. To meet this goal, we designed a scenario involving two levels of task engagement (simple navigation task vs. complex maneuvering task) with four participants in an actual light aircraft.

Consistent with previous findings [16], our experiment showed that inattentional deafness to auditory alarms could take place in an actual cockpit. This is an important step as most of the studies that demonstrated the occurrence of this phenomenon were conducted in simulated conditions [19–21]. The four participants missed a mean number of 12.5 alarms which is important considering that any absence of response to such stimuli could jeopardize flight safety as revealed by aviation accidents [2, 26]. As expected, our behavioral results disclosed that the occurrence of auditory misses was higher in the high task engagement conditions when pilots faced complex and unexpected situations such as engine-off emergency landings. These complex engaging situations are known to increase pilots' visual load as they have to carefully scan several flight parameters, perform quick actions while controlling the trajectory, and finding a grass airfield or safe fields in the country side to land. On the other hand, the simple navigation task consisted of following predefined routes at higher altitude with no time pressure. These results confirmed basic studies revealing that auditory sounds could go unnoticed when visual load is high [7–9].

Our EEG analyses support the hypothesis that inattentional deafness occurs more often during high piloting task engagement. We computed an index using the average power in the beta (13–30 Hz), alpha (8–12 Hz) and theta (4–8 Hz) bandwidths that has been shown to be related to task engagement [22–25]. This index increased in the difficult flying condition that led to higher miss rate. This suggests that the pilots were particularly mentally engaged when performing the most critical landing maneuvers. Indeed, the utility/reward of these goals were high for the pilots that were particularly committed to achieve them. The miss rate and the few number or participants did not allow us to perform event related potential or inter-trial coherency analyses as respectively achieved by [16, 21]. However, our findings do bring complementary explanations and provide additional metrics to understand the phenomenon of inattentional deafness to auditory alarms. This study shows together with others [16, 27] that dry EEG electrode systems can be implemented in actual cockpits. It paves the way to the on-line monitoring of pilot's attentional state and cockpit adaptation for safer operation. This is of key importance as transportation aircraft manufacturers are currently developing the concept of single pilot operation. As the pilot would be alone in the cockpit, task demand and level of task engagement would be higher and he/she could not rely on a second pilot to assist him/her to detect alarms. Additional perspectives would be to test different designs in real flight conditions such as spatialized warnings that have been shown to be more efficient to capture attention [28], and further to implement an online estimation of the cerebral features associated with their processing with robust methods such as detailed in [29].

Acknowledgments. This work was supported by the AXA Research Fund ("Neuroergonomics for flight safety"). We would like to thank Stephane Juaneda (chief pilot), Etienne Perrin (flight instructor) and Fabrice Bazelot (chief mechanics) for their assistance in conducting the experiments.

References

1. Parasuraman, R., Rizzo, M.: Neuroergonomics: The brain at work. Oxford University Press, Oxford (2007)
2. Bliss, J.P.: Investigation of alarm-related accidents and incidents in aviation. Int. J. Aviat. Psychol. **13**, 249–268 (2003)
3. Dehais, F., Tessier, C., Christophe, L., Reuzeau, F.: The perseveration syndrome in the pilotís activity: guidelines and cognitive countermeasures. Hum. Error Saf. Syst. Dev. 68–80 (2010)

4. Bliss, J.P., Dunn, M.C.: Behavioural implications of alarm mistrust as a function of task workload. Ergonomics **43**, 1283–1300 (2000)
5. Peryer, G., Noyes, J., Pleydell-Pearce, K., Lieven, N.: Auditory alert characteristics: a survey of pilot views. Int. J. Aviat. Psychol. **15**, 233–250 (2005)
6. Hasanain, B., Boyd, A.D., Edworthy, J., Bolton, M.L.: A formal approach to discovering simultaneous additive masking between auditory medical alarms. Appl. Ergon. **58**, 500–514 (2017)
7. Molloy, K., Griffiths, T.D., Chait, M., Lavie, N.: Inattentional deafness: visual load leads to time-specific suppression of auditory evoked responses. J. Neurosci. **35**, 16046–16054 (2015)
8. Macdonald, J.S., Lavie, N.: Visual perceptual load induces inattentional deafness. Atten. Percept. Psychophys. **73**, 1780–1789 (2011)
9. Raveh, D., Lavie, N.: Load-induced inattentional deafness. Atten. Percept. Psychophys. **77**, 483–492 (2015)
10. Lebib, R., Papo, D., de Bode, S., Baudonnière, P.-M.: Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. Neurosci. Lett. **341**, 185–188 (2003)
11. Scannella, S., Causse, M., Chauveau, N., Pastor, J., Dehais, F.: Effects of the audiovisual conflict on auditory early processes. Int. J. Psychophysiol. **89**, 115–122 (2013)
12. Callan, D.E., Dehais, F., Gonthier, N., Durantin, G.: Inattentional deafness is signified by activity in brain regions involved with attentional bottleneck. Presented at the SFN 2016, San Diego (2016)
13. Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., Tremblay, S.: Failure to detect critical auditory alerts in the cockpit evidence for inattentional deafness. Hum. Factors J. Hum. Factors Ergon. Soc. **56**, 631–644 (2014)
14. Dehais, F., Causse, M., Vachon, F., Regis, N., Menant, E., Tremblay, S.: Inattentional deafness to auditory alarms in aviation: a safety issue. Hum. Factors J. Hum. Factors Ergon. Soc. (Submitted)
15. Scannella, S., Roy, R.N., Laouar, A., Dehais, F.: Auditory neglect in the cockpit: using ERPS to disentangle early from late processes in the inattentional deafness phenomenon. Presented at the Neuroergonomics 2016, Paris (2016)
16. Callan, D.E., Gateau, T., Gonthier, N., Dehais, F.: Neural signatures of inattentional deafness in airplane pilots recorded in real flight with dry-wireless EEG. Presented at the Neuroergonomics 2016, Paris (2016)
17. Causse, M., Péran, P., Dehais, F., Caravasso, C., Zeffiro, T., Sabatini, U., Pastor, J.: Affective decision making under uncertainty during a plausible aviation task: an fMRI study. NeuroImage **71**, 19–29 (2013)
18. Causse, M., Dehais, F., Péran, P., Sabatini, U., Pastor, J.: The effects of emotion on pilot decision-making: a neuroergonomic approach to aviation safety. Transp. Res. Part C Emerg. Technol. **33**, 272–281 (2012)
19. Dehais, F., Causse, M., Régis, N., Menant, E., Labedan, P., Vachon, F., Tremblay, S.: Missing critical auditory alarms in aeronautics: evidence for inattentional deafness? Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2012)
20. Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., Tremblay, S.: Failure to detect critical auditory alerts in the cockpit evidence for inattentional deafness. Hum. Factors J. Hum. Factors Ergon. Soc. **56**(4), 631–644 (2013)

21. Dehais, F., Roy, R.N., Gateau, T., Scannella, S.: Auditory alarm misperception in the cockpit: an EEG study of inattentional deafness. In: Schmorrow, D., Fidopiastis, C. (eds.) Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience, pp. 177–187. Springer, Heidelberg (2016)
22. Prinzel, L.J., Freeman, F.G., Scerbo, M.W., Mikulka, P.J., Pope, A.T.: A closed-loop system for examining psychophysiological measures for adaptive task allocation. Int. J. Aviat. Psychol. **10**, 393–410 (2000)
23. Chaouachi, M., Chalfoun, P., Jraidi, I., Frasson, C.: Affect and mental engagement: towards adaptability for intelligent systems. In: Proceedings of the 23rd International FLAIRS Conference. AAAI Press (2010)
24. Freeman, F.G., Mikulka, P.J., Prinzel, L.J., Scerbo, M.W.: Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. Biol. Psychol. **50**, 61–76 (1999)
25. Poussot, C., Roy, R.N., Bovo, A., Gateau, T., Dehais, F., Carvalho, C.: A Loewner-based approach for the approximation of engagement-related neurophysiological features. Presented at the Submitted to IFAC, Toulouse (2017)
26. Dehais, F., Tessier, C., Chaudron, L.: GHOST: experimenting conflicts countermeasures in the pilot's activity. In: Proceedings of International Joint Conference Artificial Intelligence IJCAI, vol. 18, pp. 163–168 (2003)
27. Callan, D., Durantin, G., Terzibas, C.: Classification of single-trial auditory events using dry-wireless EEG during real and motion simulated flight. Front. Syst. Neurosci. **9**, 11 (2015)
28. Andéol, G., Suied, C., Scannella, S., Dehais, F.: The spatial release of cognitive load in cocktail party is determined by the relative levels of the talkers. J. Assoc. Res. Otolaryngol. 1–8 (2017)
29. Roy, R.N., Charbonnier, S., Campagne, A., Bonnet, S.: Efficient mental workload estimation using task-independent EEG features. J. Neural Eng. **13**, 026019 (2016)

Aerospace Neuropsychology: Exploring the Construct of Psychological and Cognitive Interaction in the 100 Most Fatal Civil Aviation Accidents Through Multidimensional Scaling

Andreas Nidos[1,2(✉)], Spyros Stavrakis Kontostavlos[1,2], Petros Roussos[1], and Kostas Mylonas[1]

[1] National and Kapodistrian University of Athens, Athens, Greece
anidos@psych.uoa.gr
[2] Pine Assessment & Development, Athens, Greece

Abstract. The human factor in aviation is a complex and multidimensional construct, incorporating different levels of analysis. At the core of this complexity, cognitive and psychological variables occur in dynamic interaction. A model of aerospace neuropsychology is proposed in order to investigate the interaction between psychological variables and neurocognitive errors in the 100 most fatal civil aviation accidents. Recurring psychological and cognitive themes described in the accidents were subjected to multidimensional scaling. Five conceptual clusters pertaining to *individual pilot characteristics*, *sociotechnical cockpit performance*, *organizational/operational effect*, *training-level of involvement and decision making-overestimation,* were identified. Psychological and cognitive interaction is apparent throughout the accidents' human factor space, while individual characteristics are located within a dense area visually separated from sociotechnical and human-machine interaction features. Results are discussed in light of the need to review current aeromedical examination procedures under a scientific accurate and ad hoc methodology.

Keywords: Aerospace neuropsychology · Emotion and cognition interaction · Pilot assessment · Human factor in aviation · Real flight data · Aviation accident analysis · 100 most fatal civil aviation accidents · Multidimensional scaling

1 Introduction

Accident analysis has been for some time now the prevailing methodology of investigating and preventing human error in aviation accidents [1]. This error has had a manifold representation in the literature, for instance, under a cognition-based error model within a cognitive architecture [2], as a neurocognitive symptom with physiological underpinnings (e.g. spatial disorientation) [3], or as an occurring event in situations that increase human's innate vulnerability to error [4]. On the other hand, analysis of incidents related to the pilot's psychological state (e.g. stressful events), personality traits (e.g. inflexibility), psychiatric conditions (e.g. psychopathology) or even organizational factors (e.g. organizational pressures) have been mostly dealt with separately in the aviation literature. This "cognitive and psychological distinction" in the study of the human factor is also reflected on explanatory models in aviation and

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_22

ergonomics literature, and has also had an implication in the design and implementation of aeromedical assessment procedures. A new research design that takes into account both cognition and emotion is therefore warranted, in order to improve our understanding of the black box that is the human factor in aviation accidents.

Aerospace neuropsychology is the integration of neuropsychological methodology, theory and practice in aerospace settings, in order to study and assess individuals in every aspect of the human-machine interaction with an aim to fly. In a recent study[1], the 100 most fatal civil aviation accidents served as a research paradigm in order to explore the human factor under an aerospace neuropsychological framework [5]. Researchers identified and described 24 psychological and 17 cognitive recurring themes implicated in the accidents. A possible interaction–operationally defined as the presence of at least one cognitive and one psychological variable in an accident- was found in 53% of the cases (Table 1).

Table 1. Recurring cognitive and psychological themes in the 100 most fatal civil aviation accidents.

Cognitive (n = 17)	rf	Psychological (n = 25)	rf
Situational awareness	40	Communication	41
Monitoring	36	Violation of procedures	38
Attentional procedures	33	Risk involved	33
Decision making	27	Crew dynamics	28
Training	23	Operational pressure/goal oriented	22
Planning	17	Overestimation of capabilities	20
Language comprehension	13	Inflexibility	20
Multitasking	12	Passiveness	20
Prospective memory	11	Under reaction	18
Flawed assumptions	11	Performance under stress	16
Navigational error	11	Fatigue	13
Checklists	8	Other psychophysiological conditions	11
Spatial disorientation	8	Impulsivity	10
Computation	6	Sleep deprivation	8
Procedural steps	6	Obedience to authority	8
Level of alertness	5	Work conditions	8
Dyslexia	1	Aggressiveness	6
		Invulnerability	6
		Over-reliance on automation	6
		Excessive reaction	4
		Psychopathology	4
		Macho attitudes	3
		Murder suicide	2
		Real-life stressors	2

[1] Best poster award at the 5th European Congress of Aerospace Medicine (ECAM 2016) in Oslo, Norway.

Interaction between emotion and cognition is apparent both in clinical practice - with a characteristic example the case of cognitive dysfunction in depression- [6] and research [7]. Moreover, one of the most impressive scientific breakthroughs of our century has been the advances in the field of neuropsychology and the mapping of the brain. According to Young et al. [8] the amygdala which is considered the center of emotional processing connects with virtually every cortical area of the brain. Cortical areas are responsible for higher order cognitive functions with impressive complexity involving hundreds of paths that correlate emotional and cognitive functions. In this manner, the amygdala is hypothesized to be a strong candidate for integrating cognitive and emotional information.

Human factor has been found to account for at least 69% of the 100 most fatal aviation accidents [5] and consists of neurocognitive errors and psychological variables. The nature of the human factor itself is directly related to the scientific methodology of aerospace neuropsychology, which perceives human factors as cognitive and psychological variables' coexistence and dynamic interaction.

Two civil aviation accidents can serve as paradigms of either a possible effect of a psychological state or variable in cognitive performance or a single source human factor related only to the psychological state of the subject.

In flight YH708 of West Caribbean Airways on August 16, 2005 the crew of an MD-82 failed to take appropriate action to correct the stall of the aircraft, resulting in loss of control and 160 fatalities (all). From the official accident report and related sources, a number of psychological/emotional and cognitive contributing factors attributed to the crew can be observed [9]. More specifically, the psychological state of the captain at the time of the accident could be described by psycho/physical fatigue, anxiety, low self-esteem, stress and frustration, whereas the cockpit crew resources at the time of the accident are characterized by lack of effective communication and crew dynamics. Moreover, the generation gap between the captain and the co-pilot (40 and 21 years old respectively), in conjunction with their behavioral profile resulted in submissiveness/obedience and passiveness by the co-pilot without the verbalization or manifestation of any kind of emotion in a very stressful situation. A number of operational factors seem to have contributed to the psychological state of the crew with the airline having a number of financial problems affecting the operations by creating a climate of uncertainty and stress. Prior to take off, the aircraft was long delayed owing to payment of fuel resulting also in the deterioration of the alertness and fatigue of the crew. The captain was unpaid for 6 months and was obliged to work at a family owned bar/restaurant to gain some income, while infringements in crew rest hours, flying time and leave periods, failure to provide crew with regulation training and inconsistencies in aircraft records and flight documents could also be noted. The cognitive variables attributed to the accident are: training, attention, planning, checklists, alertness, situational awareness, decision making. There is a high probability that the respective psychological variables had a detrimental effect in the neurocognitive errors resulting in the loss of control and crash of the aircraft.

Germanwings flight 4525 raised a global interdisciplinary discussion regarding pilot assessment methods and human factors in aviation. On March 24, 2015 an Airbus 320 crashed in a mountainous area in southern France killing all 144 passengers and 6 crew members. The accident was due to a deliberate and planned action by the co-pilot to perform murder-suicide crashing the aircraft into the French Alps. The psychological state of the co-pilot, despite the fact that he was a holder of a class 1 medical certificate, was characterized by a mental disorder were active clinical psychopathology with psychotic symptoms was diagnosed and antidepressant and sleeping aid medication was prescribed. The co-pilot was referred to a psychotherapist and psychiatrist by a private physician one month before the murder-suicide and was diagnosed with a possible psychosis. The psychiatrist treating the co-pilot prescribed anti-depressant medication one month before the crash and other anti-depressants along with sleeping aid medication eight days before. The series of events included the captain of the aircraft leaving the cockpit at FL380 and the selected altitude changing from 38000 ft to 100 ft by the co-pilot while alone in the cockpit, with the aircraft starting a continuous and controlled descent on autopilot. The co-pilot never responded to the Marseille control centre or the French military defence system who tried to contact the aircraft on various attempts. The buzzer to request access to the cockpit by the pilot sounded 04:40 after the captain had left with no response from the co-pilot. Furthermore, none of the intercom calls elicited any answer, leaving the captain of the aircraft outside the cockpit until the collision with the terrain at 09 h 41 min 06. It is noted that the cockpit doors of the aircraft are designed for security reasons to resist penetration by small arms fire and grenade shrapnel and to resist forcible intrusions by unauthorized persons [10]. The Germanwings crash is an aviation disaster characterised by a single dimension of causal agent attributed to psychological variables and resulting in a conscious and deliberate act by the co-pilot.

The human factor in aviation is a complex and multidimensional construct, incorporating different levels of analysis. This is evident throughout aviation and ergonomics literature in theoretical frameworks, explanatory models and technical documentation of human factor accounts of civil aviation accidents. In a cognitive analysis model for example, a visually meaningful distinction has been made between practitioner and domain characteristics [11], one that places task performance on a continuum from human attributes (e.g. goals, motives, knowledge skills and strategies) to social/organizational structure and physical environment (e.g. sociotechnical organization, physical characteristics and constraints). Similarly, the Human Factor Analysis and Classification System (HFACS) as a comprehensive human error framework follows an analytical course from unsafe acts of operators to organizational influences [12]. The existence of many quantitative and qualitative distinctions between a successful performance or a fatal error while on air, calls for the use of more refined descriptive methods of the human factor in aviation. Aerospace neuropsychology attempts to further refine previous human factor constructs with the integration of cognition and psychology as latent and active variables affecting the human-organization continuum every step of the way.

The present study is an effort to explore an underlying construct for cognitive and psychological coexistence and possible interaction in the 100 most fatal civil aviation accidents.

2 Method

Recurring cognitive and psychological themes in the 100 most fatal civil aviation accidents were coded and organized in the respective variable categories "cognitive functions" and "psychological/emotinal variables". Two researchers performed qualitative and quantitative content analysis independently on the following: Official final/inconclusive/preliminary accident reports, cockpit voice recorder transcripts, judicial investigate reports, online databases, peer-reviewed articles, textbooks on aviation accidents. A third researcher calculated inter-rater reliability (Cohen's kappa = 0.79). Human factor was implicated in 69% of total cases. A possible effect of psychological factors to neurocognitive errors was defined as at least one psychological and one cognitive variable attributed to the pilot spatiotemporally during the accident and was accounted for 77% of causes related to human factor (n = 53). Variables from either a cognitive or a psychological source was found in 10% of human factor accidents (n = 7), while 13% of the human factor cases were considered as inconclusive (n = 9).

Results from content analysis were subjected to multidimensional scaling (MDS), with individual matrices for co-occurrence of cognitive and psychological variables simultaneously. Non-metric MDS (through ALSCAL) was applied in order to explore a two-dimensional graphical representation reflecting the map of human factors implicated in the accidents in question. The squared Euclidean distance between vectors of estimates was computed for each variable from the two human factor sources (cognitive and psychological). In this way distances represent composite measures of the spatial relationships of neurocognitive errors and psychological/emotional states in the sample. Further clarification of the underlying construct of the possible interaction matrix was attempted using a multidimensional scaling trigonometric solution (MDS-T), according to Mylonas [13–15]. Through this method the initial ALSCAL outcomes are constrained through trigonometric transformations on the circle (two-dimensional solutions) and/or the sphere (three-dimensional solutions) to reach better explanatory power as quadrant specific information is taken into consideration along with the relative distances among the stimuli analyzed [14, 15].

3 Results and Discussion

Psychological and mental health variables attributed to the crew were found adequate to cause an aviation disaster. These variables were directly related to the accident cause and were observed without the presence of a neurocognitive error. An act appearing at least twice in the 100 most fatal aviation accidents (Germanwings flight 4U4525, EgyptAir flight MS990) and is directly related to mental health variables and psychopathology is *murder-suicide*. This statistic does not take into consideration underdetermined causes in the sample where murder suicide is probable. Murder-suicide by pilot is a conscious and determinate double act in which the pilot in command deliberately crashes the aircraft as a way to kill himself while resulting in the death of one or more others on board. Despite murder suicide occurs relatively rare in aviation it is in large proportion to the number of deaths related to other accidents, while post hoc it is often appraised as an event that could have been prevented if the effect of psychological condition on human risk factor was adequately assessed. It is worth noting that a number of pilots that committed murder suicide had active psychopathology and were on psychiatric medication for related conditions at the time of the accident. *Suicide* and *murder-suicide* by pilot are often mistakenly considered as identical while they represent two independent and recognizable entities that in psychiatry are treated as separate events with distinct risk factors [16].

An individual matrix for possible interaction between cognitive and psychological variables (Fig. 1) reached a marginally acceptable stress of 0.142 with R-sq = 0.909. Multidimensionality was inferred from data organization with five visually recognizable conceptual clusters:

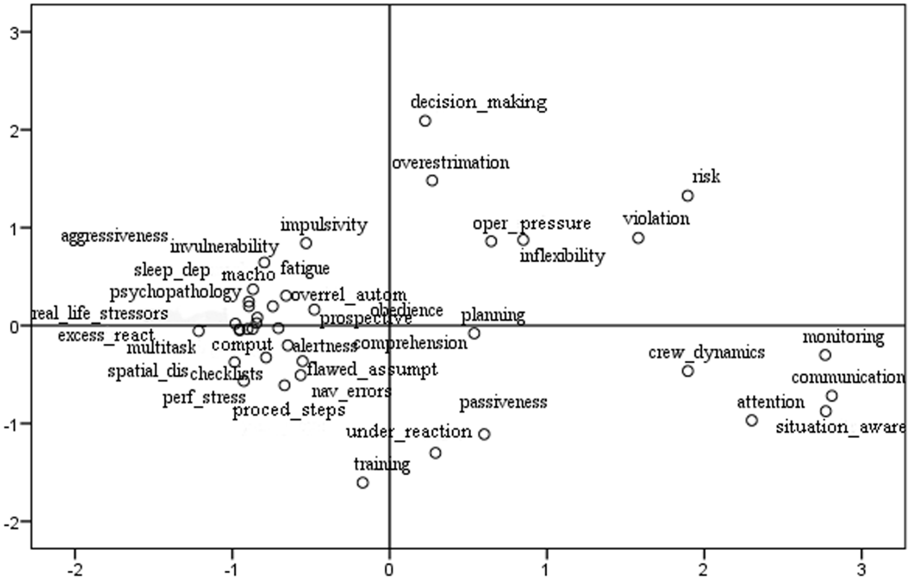


Fig. 1. Multidimensional scaling (*ALSCAL*) of cognitive and psychological variables coexistence and possible interaction in the 100 most fatal civil aviation accidents. Distances represent composite measures of the spatial relationships of neurocognitive errors and psychological/emotional states. Five conceptual clusters pertaining to (1) *individual pilot characteristics*, (2) *sociotechnical cockpit performance*, (3) *organizational/operational features* and supervising outliers *(4) training-level of involvement and (5) decision making-overestimation* are identified.

- *Individual Pilot Characteristics*: 61.5% of all data were accumulated in a space array roughly oriented vertically between −1 and 1 and horizontally between −1 and 0 (Fig. 1). Perceived individual characteristics and skills inside this shape formed a distinctive group, with further discrimination between cognitive[2] and psychological[3] variables bounded respectively below and above the X axis. A visual representation of coexisting cognitive and psychological variables inside this dense area is indicative of the possible effect of the psychological variables in the cognitive performance of the pilot. There is sufficient evidence to support that psychological variables related to personality traits, psychopathology, physiological states and real-life stressors can have a significant effect on cognitive functionality [7, 17].

- *Sociotechnical cockpit performance*: An apparent configuration of proximities on the opposite side, roughly oriented vertically between −1 and 0 and horizontally between 2 and 3, represents variables that could be conceptualized as recurring ad hoc inside the cockpit. Variables *situational awareness*, *monitoring, communication, attention* and *crew dynamics* are perceived as cognitive and psychological features that represent the interactive environmental challenges between systems and human resources of the airman during flight. These variables include but are not limited to Crew Resource Management involving interaction human-to-human (e.g. crew dynamics) and human-to-aircraft/human-machine interaction (e.g. monitoring). According to Stanton and colleagues [18, 19], situational awareness -a variable belonging to this cluster- is “an emergent property of the sociotechnical system which represents the unit of analysis rather than the individual agents working within it”.
- *Organizational/operational features*: *Operational pressure* (e.g. goal oriented behavior in order to meet the company’s flight schedule time tables under potentially dangerous situations) tend to form distinct proximities with individual characteristics that are separated from the dense area where most personality traits lie. These variables are: *risk-involved behavior*, *inflexibility* and *violation of procedures*. It appears that the co-occurrence of the above in conjunction with operational pressure holds a significant weight in the analysis of the 100 most fatal civil aviation accidents through aerospace neuropsychology as a prone to accident conceptual cluster.
- *Supervising Outliers: decision making and training*: *Decision making* and *training* appeared as vertically organized outliers (+2.1, −1.6 respectively). These cognitive variables seem to have a supervising role in the aerospace setting and have a detrimental effect should they breakdown. Two distinct conceptual entities directly related to training and decision making in aviation accidents could be further identified:

  (i) Level of involvement: Adequacy of *training* also seems to define the level of involvement in a potentially dangerous situation, forming a distinct cluster with the psychological variables *passiveness* and *under reaction*. Inadequate

[2] Prospective memory, computation, multitasking, use of checklists, alertness, comprehension, navigational errors, procedural steps, planning, spatial disorientation, flawed assumptions.

[3] Impulsivity, invulnerability, aggressiveness, macho attitude, excessive reaction, obedience to authority, over-reliance on automation, performance under stress, sleep deprivation, fatigue, work conditions, real-life stressors, psychopathology.

pilot training is often followed by passiveness and under reaction resulting in lower level of involvement in a critical situation.

(ii) Overestimated decision making: *Decision making* on the other hand has a proximity with *overestimation of capabilities,* a finding which is relevant to literature regarding over-assessment of abilities in pilots [20]. Probabilistic reasoning related to overestimation of capabilities by the pilot and its effect on decision making set the ground for the emergence of cognitive biases in aviation accidents.

This MDS solution was further tested through MDS-T, a trigonometric transformation on a circle to reach better explanatory power as quadrant specific information is taken into consideration along with the relative distances among the cognitive and psychological variables implicated in the accidents [14, 15]. Dimensionality and conceptual clusters described earlier were confirmed (Fig. 2).

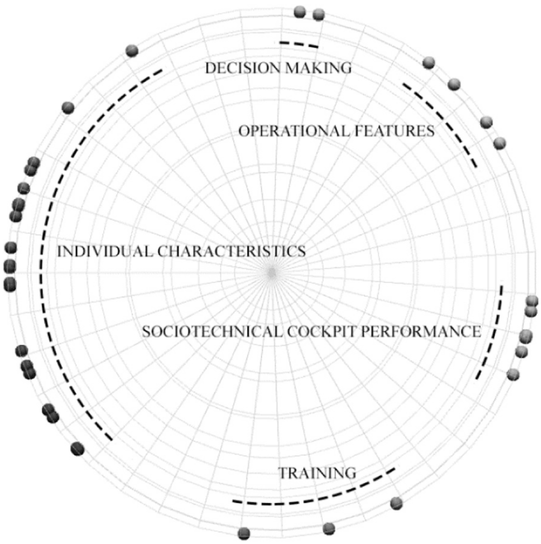


Fig. 2. Clustering of cognitive and psychological variables coexistence and possible interaction in the 100 most fatal civil aviation accidents through trigonometric transformations on the circle (MDS-T). Distances represent composite measures of the spatial relationships of neurocognitive errors and psychological/emotional states. Five conceptual clusters pertaining to (1) *individual pilot characteristics*, (2) *sociotechnical cockpit performance*, (3) *organizational/operational features* and supervising outliers (4) *training-level of involvement* and (5) *decision making-overestimation* are identified.

Single source human factors were identified in a few cases related to psychological variables mainly regarding psychopathology and murder suicide. The qualitative difference of these cases is their relevance to a determinate act by the pilot directly related to the accident cause and not to an error. Contrary, almost in each case where a neurocognitive error or characteristic was identified in an accident, at least one psychological variable was respectively identified pertaining to the same accident.

In the majority of cases, coexistence of psychological and cognitive variables yielded a complicated multidimensional visual structure, suggesting a possible psychological and cognitive interaction and a differentiation between organizational, human-machine/sociotechnical and individual characteristics clusters. Moreover, differentiated and individual variables such as “training” and “decision making” appeared to have a distinct effect in the structure, whereas adequate training was also related to the level of involvement in a potentially dangerous situation and decision making to overestimation of abilities.

The present study is to our knowledge the first effort to visually represent perceived cognitive and psychological coexistence and their possible interaction in aviation accidents through multidimensional scaling. Proximities and dissimilarities of cognitive and psychological variables that were implicated in the 100 most fatal civil aviation accidents were shown for the first time on a human factor map, while an attempt to explore the construct of emotion and cognition interaction was made. As multidimensionality was readily apparent in scaling this interaction, several meaningful constructs directly related to real-flight data were recognized. These human factor constructs are more than just face value since they are related to events and circumstances that could be reenacted over and over when in the air. Practitioner-individual characteristics, for instance, appear in our study inside a dense area which includes more than 2/3 of the total observations. This area could be considered as a “hot space” where psychological and cognitive phenomena appear in dynamic interaction. Moreover, this area is visually separated from domain characteristics such as organizational factors with a possible effect on cognitive or psychological performance and sociotechnical systems. The above distinction has been described in theoretical accounts of analysis of cognitive work [11] and is apparent with many technical details in human error frameworks such as the HFACS [12].

The human factor constructs and dimensions identified in the present study could be of ecological validity and hold a predictive value. In this light they should be linked to study methods and assessment procedures of the human factor in civil aviation. This is particularly important since aeromedical examination and assessment procedures are currently under great scrutiny and revision, especially after the recent Germanwings case. Results from this study are discussed towards the direction of a current need to review the existing examination procedures and create new reliable tools in an ongoing effort to calculate meticulously the necessary variables for safe flight. The development of new assessment tools, tests and prototype evaluation methods should be related to actual flight performance data under a scientific accurate and ad hoc methodology.

Accident analysis could be a valuable tool for gaining insight into the black box that is the human factor in aviation, similar to its role in the improvement of the safety procedures and mechanical reliability of the machine. When arguing about aerospace safety exhaustive control of exogenous factors is usually the rule of thumb, leaving an inadequate assessment framework in terms of endogenous human variables. As a conclusion, it is argued that the role of a neuropsychological approach towards the identification of neurocognitive errors, psychological variables and their constant and dynamic interaction in aviation mishaps could be the method of choice as it is directly related to the nature of the human risk factor itself.

References

1. Shappell, S., Wiegmann, D.: Applying reason: the human factors analysis and classification system (HFACS). Hum. Factors Aerospace Saf. **1**, 59–86 (2001)
2. Busse, D.K., Johnson, C.W.: Using a cognitive theoretical framework to support accident analysis. In: Proceedings of the Working Group on Human Error, Safety and System Development. International Federation for Information Processing, Seattle, WA (1998)
3. Gibb, R., Ercoline, B., Schafff, L.: Spatial disorientation: decades of pilot fatalities. Aviat. Space Environ. Med. **82**, 717–724 (2011)
4. Dismukes, R.K., Berman, B.A., Loukopoulos, L.D.: The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents. Ashgate, Aldershot (2008)
5. Stavrakis-Kontostavlos, S., Nidos, A., Roussos, P.: Aerospace neuropsychology: a theoretical framework and research paradigm from the study of the human factor in the 100 most fatal civil aviation accidents. In: Poster Session Presented at the 5th European Congress of Aerospace Medicine, Oslo (2016)
6. Hammar, Å., Årdal, G.: Cognitive functioning in major depression – a summary. Front. Hum. Neurosci. **3**, 26 (2009)
7. Mueller, S.C.: The influence of emotion on cognitive control: relevance for development and adolescent psychopathology. Front. Psychol. **2**, 327 (2011)
8. Young, M.P., Scannell, J.W., Burns, G.A., Blakemore, C.: Analysis of connectivity: neural systems in the cerebral cortex. Rev. Neurosci. **5**, 227–249 (1994)
9. Junta Investigadora de Accidentes de Aviación Civil (JIAAC) – Venezuela.: Report Number JIAAC-9-058-2005, 13 August 2010 (2010)
10. Bureau d'Enquêtes et d'Analyses (BEA) – France, Report Number BEA2015-0125, 13 March 2016 (2016)
11. Bisantz, A., Roth, E.: Analysis of cognitive work. In: Boehm-Davis, D.A. (ed.) Reviews of Human Factors and Ergonomics, vol. 3, pp. 1–43. Human Factors and Ergonomics Society, Santa Monica (2008)
12. Shappel, S.A., Wiegmann, D.A.: The human factors analysis and classification system–HFACS. U.S. Federal Aviation Administration, Office of Aviation Medicine, Washington, D.C. (2000)
13. Mylonas, K.: Statistical analysis techniques based on cross- cultural research methods: cross-cultural paradigms and intra- country comparisons. Psychol. J. Hellenic Psychol. Soc. Special Issue: Cross- Cult. Psychol. **16**(2), 185–204 (2009)
14. Mylonas, K.: Factor structures across countries and across clusters of countries: a 36-country study on the family values scale. Adv. Soc. Sci. Res. J. **3**(8), 63–76 (2016). doi:10.14738/assrj.38.2157
15. Papazoglou, S., Mylonas, K.: An examination of alternative multidimensional scaling techniques. Educ. Psychol. Measur. 1–20 (2016). doi:10.1177/001316441666182
16. Kenedi, C., Friedman, S.H., Watson, D., Preitner, C.: Suicide and murder-suicide involving aircraft. aerospace medicine and human performance, Aerospace Med. Assoc. **87** (4), 388–396 (2016)
17. Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., van IJzendoorn, M.H.: Threat-related attentional bias in anxious and non-anxious individuals: a meta-analytic study. Psychol. Bull. **133**, 1–24 (2007). doi:10.1037/0033-2909.133.1.1
18. Stanton, N.A., Chambers, P.R.G., Piggott, J.: Situational awareness and safety. Saf. Sci. **39**, 189–204 (2001)

19. Stanton, N.A., Salmon, P.M., Walker, G.H., Salas, E., Hancock, P.A.: State-of-science: situational awareness in individuals, teams and systems. Ergonomics (2017). doi:10.1080/00140139.2017.1278796
20. Wilson, D.R., Fallshore, M.: Optimistic and ability biases in pilots' decisions and perceptions of risk regarding VFR flight into IMC. In: Proceedings of the 11th International Symposium on Aviation Psychology, 5–8, March 2001, Columbus, OH (2001)

Physiological Model to Classify Physical and Cognitive Workload During Gaming Activities

Néstor Arroyo-Gómez, José Laparra-Hernández(✉), Andrés Soler-Valero, Enrique Medina, and Helios de Rosario

Instituto de Biomecánica de Valencia (IBV), Universitat Politècnica de València, Edifici 9C Camino de Vera s/n, 46022 Valencia, Spain
{Nestor.Arroyo, Jose.Laparra, Andres.Soler, enrique.medina, helios.derosario}@ibv.upv.es

Abstract. Some of new approaches in Human Factors and Ergonomics are based on the assessment of cognitive and physical workload using physiological measurements. Nevertheless, the relationship between both requires to get in depth about its causes and effects. The main goal of this work was to develop a model to distinguish the impact of physical and cognitive workload, leaning on physiological response analysis. To do this, senior citizens performed a set of tasks of video games, where the predominance of each type of workload is known. Facial electromyography, galvanic skin response and electrocardiogram signals from subjects were recorded while they performed the tasks. The parameters extracted were used to design a classification model to predict the type of workload involved in a task. The designed model is based in a reduced number of variables and it achieves a 75.6% of success to differentiate physical and cognitive demands.

Keywords: Gaming tasks · Cognitive/physical demands · Physiological model · Senior citizens

1 Introduction

Workload is an indicator of the mental and/or physical effort required to carry out one or more tasks at a specific performance level [1]. Workload assessment techniques, for both physiological and cognitive components, can be classified into two categories: subjective and objective. Subjective assessment techniques are based on an individual's personal feelings and perceptions, and they are typically considered less reliable than workload assessment with objective techniques [2]. Those last ones are usually based on quantitative data such as the level of performance or physiological data, and they are often more reliable.

The relationship between physiological response and workload has been studied in the past. As an example, Kohlisch and Schaefer [3] conducted an experiment within they investigated the differences between mental workload and motor activity while subjects performed computer tasks. Motor activity was forced by the speed of compensatory keystrokes and mental workload was forced by mental arithmetical tasks.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_23

One of the findings from their study suggested that the heart rate was a useful indicator of the attentional aspects of mental load [3]. Further, the cardiovascular system responds to both physical and cognitive stresses, by causing an increase in both heart rate (HR) and blood pressure [4]. In the experimental study they found that the rate–pressure product (RPP), given by the product of HR and systolic blood pressure, is significantly modified by both cognitive and physical tasks. Moreover, the findings from this study indicated that RPP can be used as an objective measure to separate the components of physical workload and cognitive stress in combined tasks.

The main goal of this research was to get in depth of the relationship between physiology and workload. Workload has been often studied without combine different physiological measurements, thus, the starting point was a broad assessment of the physiological response. In addition, different gaming tasks were performed, in order to link the predominant workload type and the physiological state changes.

2 Materials and Methods

2.1 Participants

Twenty subjects were involved into the experiment. They were divided in two groups according to their age. On one hand, a group of 10 older subjects (5 women and 5 men) with age ranging between 65 and 75 years (mean age 62.3 and standard deviation 2.2). On the other hand, a group of 10 subjects (4 women and 6 men) with age ranging between 45 and 55 years (mean age 49,4 and standard deviation 2.8). As inclusion criteria, they should not suffer any severe pathology (physical or cognitive) and they should not have previously played video games.

2.2 Experimental Procedure

Each user took part in 3 game sessions with at least 3 days of separation between sessions. At the beginning of each experimental session, the subject was instrumented with the equipment described below. Then, the technician requested to the subject to sit down and rest during 30 s, in order to record the basal physiological state. During this period the subject should not be influenced by any type of stimulus. Later, the user performed the 4 gaming tasks. The tasks sequence was repeated three times throughout the session. In order to stabilize the physiological state, subjects maintained a relaxed state during 10 s before the beginning of the next gaming task.

2.3 Equipment

Physiological signals were recorded using BioPLUX. It is a device that collects and digitalizes the signals from different kind of sensors. The device transmits the data to a computer via Bluetooth, where it is displayed in real time. BioPLUX was used to acquire the electromyography (EMG) signals on the zygomaticus major and corrugator supercilii muscles of the non-dominant side of subject’s body. Zygomatic major EMG is related to the action of smiling and to positive emotional valences. Moreover, corrugator supercilii EMG is related to the action of frowning and to negative emotional valences. In addition, BioPLUX was used to record subjects' electrocardiography (ECG) and the galvanic skin response (GSR). GSR electrodes were placed on the palm of the non-dominant hand (Fig. 1).

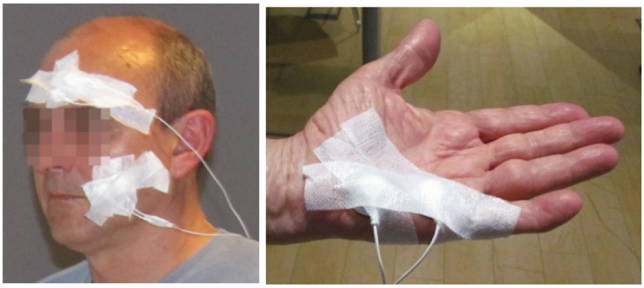


Fig. 1. Facial electromyography (*left*) and galvanic skin response (*right*) electrodes placement.

2.4 Tasks Description

The consoles and video games used in the study were chosen taking into account the predominant type of workload (physical and cognitive), the difficulty of use for the subjects, and the movements to be performed during the game. Table 1 shows the main characteristics of the tasks. In addition, each task is described in detail below and Fig. 2 shows some screenshots of the games.

Table 1. Tasks performed during the experimentation.

#	Game	Console and accesories	Workload
1	Big brain academy: Balloon Burst	Nintendo Wii	Cognitive
2	Wii Fit Plus: Ski Slalom	Nintendo Wii + Balance board	Physical
3	Dr. Kawashima's Brain and Body Exercises	XBOX 360 + Kinect	Cognitive
4	Kinect Adventures: River Rush	XBOX 360 + Kinect	Physical

Task #1. Several numbered balloons appear on the screen and the user has to pop them from lowest to highest number or viceversa. To do this, the subject uses the Wii remote control to point and the "A" button to select.

Task #2. The avatar simulates a skier. The objective is to ski between the gates that are signposted with arrows. The player controls the direction of the skis distributing his/her body weight from right to left on the "Wii balance board".

Task #3. Several figures are presented at the same time, and the goal is to match the two figures that are identical in shape and color. User controls the avatar arms by moving his/her arms.

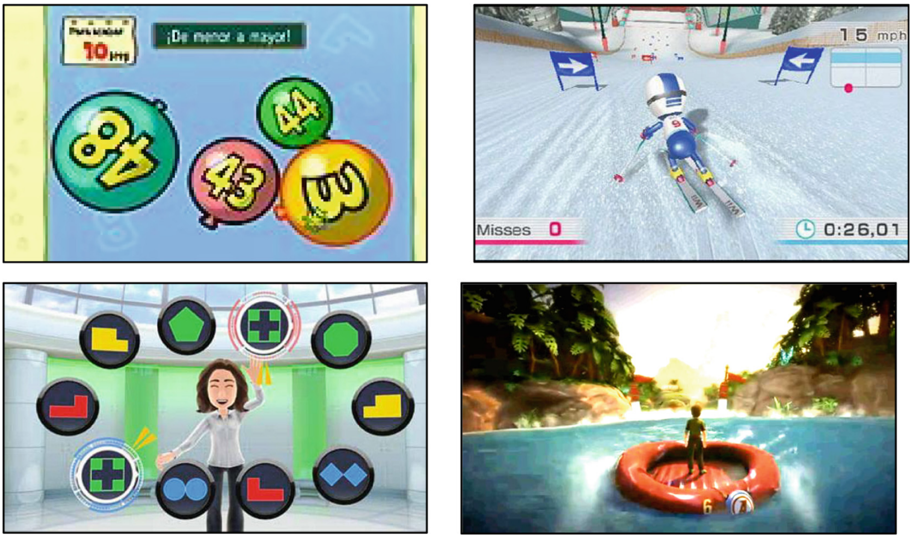


Fig. 2. Screenshots of the games used for the tasks. Task #1: Big brain academy: Balloon Burst (*up to the left*), Task #2: Wii Fit Plus: Ski Slalom (*up to the right*), Task #3: Dr. Kawashima's Brain and Body Exercises (*down to the left*) and Task #4: Kinect Adventures, River Rush (*down to the right*).

Task #4. The objective is to guide a raft along the rapids of a river collecting the largest number of coins. The user changes the direction of the raft by moving his body to the right or to the left. The game allows jumping, but, in this experimentation, users were prevented to jump during the game to avoid potential injuries.

2.5 Experimental Design

The order of performance of the tasks (Table 2) for each user was randomized for avoiding order influence. As the tasks combinations were the same for both groups of subjects, there were two leftover combinations to be discarded. To make the decision about the combinations to be discarded, consoles were prioritized against the type of workload. Then, combinations in which the same console (Wii or XBOX 360) were tested at beginning and the end were discarded: "#3, #2, #1, #4" and "#1, #4, #3, #2".

2.6 Signals Processing

EMG and ECG signals were filtered with a 50 Hz and 100 Hz notch filter to avoid first and second order interference from power lines. Moreover, the EMG signals were filtered with a 0.2 Hz high-pass filter to reject baseline effect and were fully rectified and filtered with a 4 Hz low-pass filter to acquire the envelope of the signal.

ECG signals were filtered with a 0.1 Hz high-pass filter to diminish baseline effect and artefacts due to body movement (e.g. breathing). In addition, a modified version of

Table 2. Order of the tasks performed for each subject during the experimentation.

Subject number		Tasks order of performance			
Older group	Younger group	1st	2nd	3rd	4th
1	11	#4	#1	#2	#3
2	12	#1	#2	#4	#3
3	13	#3	#1	#4	#2
4	14	#1	#3	#2	#4
5	15	#4	#3	#1	#2
6	16	#2	#1	#3	#4
7	17	#2	#4	#1	#3
8	18	#2	#3	#4	#1
9	19	#4	#2	#3	#1
10	20	#3	#4	#2	#1

Pan-Tompkins algorithm [5] was applied to detect "R" peaks and calculate the Heart Rate. ECG signals were visually reviewed in order to avoid mistakes on R peaks detection. Then, Heart Rate Variability (HRV) was calculated using the axis of Poincare Plots.

GSR signals were decomposed as tonic response (baseline) and a phasic response, (fast fluctuations) that are the result of external stimulus and related to the arousal level [6]. The phasic response was extracted using a 0.1 Hz high-pass filter.

2.7 Data Normalization

Normalization of physiological data is critical. In fact, many studies cannot get results due the high variability of the data. EMG and GSR data are especially dependent on the near time states and on the subject.

The method used to normalize EMG and GSR signals is shown in (1). Here, the normalized variable N depends on X, the 75th percentile of the signal acquired during each task, and B, which is the 75th percentile of the signal during the initial 30 s of relaxation state. This method allowed to reducing the intra- and inter-subject variability.

$$N = \frac{X - B}{B}. \quad (1)$$

2.8 Model Design

In order to design the model, a logistical binary regression was made, using the backward elimination method of Wald, since it allows to design a model with the maximum percentage of success. This method introduces in the model a set of variables and eliminates in each step the variable that less contributes to explain the differences.

The same procedure was done including different sets of variables: all the 20 variables extracted, only those related to a single type of physiological phenomena, and those that were normalized following the same criterion.

At last, two workload classification models were obtained. One with the objective to achieve the higher success rate, and another one focused on reducing the number of input variables.

3 Results

3.1 Workload Classification Model with the Higher Success Rate

The classification model with the higher success rate, was obtained introducing as input all the variables (20 variables). This model achieved a percentage of success of 78.3% (Table 3). However, it was based on 12 variables and some of them were highly related: mean and standard deviation of R-R intervals, SD2 axis of Poincaré plot, both phasic and tonic response of GSR and, finally, EMG parameters resulting from steps of their treatment and normalization process. Table 4 summarizes model characteristics and the results of Hosmer and Lemeshow test, which validates the model because it is non-significant ($p < 0.05$).

Table 3. Classification table of the model with the higher success rate.

Observed		Predicted		
Physical	Cognitive	Physical	Cognitive	Correct (%)
360	0	269	91	74.7
0	360	65	295	81.9
		Overall success (%)		78.3

Table 4. Summary and Hosmer and Lemeshow test for the model with the higher success rate.

Summary				Hosmer & Lemeshow	
Step	−2LL	Cox & Snell R^2	Nagelkerke R^2	χ^2	p-Value
1	705.531	0.334	0.445	11.684	0.166
2	705.546	0.334	0.445	12.894	0.116
3	705.569	0.334	0.445	12.144	0.145
4	705.598	0.334	0.445	13.256	0.103
5	705.775	0.334	0.445	12.258	0.140
6	706.022	0.333	0.445	13.490	0.096
7	706.831	0.333	0.444	13.739	0.089
8	708.247	0.331	0.442	13.116	0.108

Table 5. Input variables of the workload type classification model.

Symbol	Variable
V^{a}_{corr}	75th percentile of EMG corrugator supercilii potential
V^{a}_{zig}	75th percentile of zygomaticus major EMG potential
V^{a}_{GRST}	75th percentile of tonic GSR potential
STD (RR)	Standard deviation of R-R intervals of ECG (ms)
SD2	Poincaré plot axis (ms)

[a]Parameters normalized respect the value at the beginning of the task.

3.2 Understandable Workload Classification Model

In order to achieve a more understandable model, variables related with the same physiological phenomena were not introduced in the model. The resulting model was based on 5 input variables (see Table 5), and it predicts the predominant workload type with a percentage of success of 75.6% (see Table 6). Table 7 summarizes model characteristics and the results of Hosmer and Lemeshow test, which validates the model because it is non-significant ($p < 0.05$).

Table 6. Classification table of the "optimum" model.

Observed		Predicted		
Physical	Cognitive	Physical	Cognitive	Correct (%)
360	0	254	106	70.6
0	360	70	290	80.6
		Overall success (%)		75.6

Table 7. Summary and Hosmer and Lemeshow test for the "optimum" model.

	Summary			Hosmer & Lemeshow	
Step	−2LL	Cox & Snell R^2	Nagelkerke R^2	χ^2	p-Value
1	768.024	0.274	0.365	10.083	0.259
2	768.155	0.273	0.365	10.546	0.229
3	768.376	0.273	0.364	11.473	0.176

The model allows to predict the workload according following (2), where "k" is described in (3). Forecast value, F, has to be interpreted as follows: If $F \in (0, 0.5)$, the predicted workload is physical, if $F \in (0.5, 1)$, the predicted workload is cognitive. The behavior of the model in terms of workload type prediction for each variable is shown in Fig. 3.

$$F = \frac{e^k}{1 + e^k} \tag{2}$$

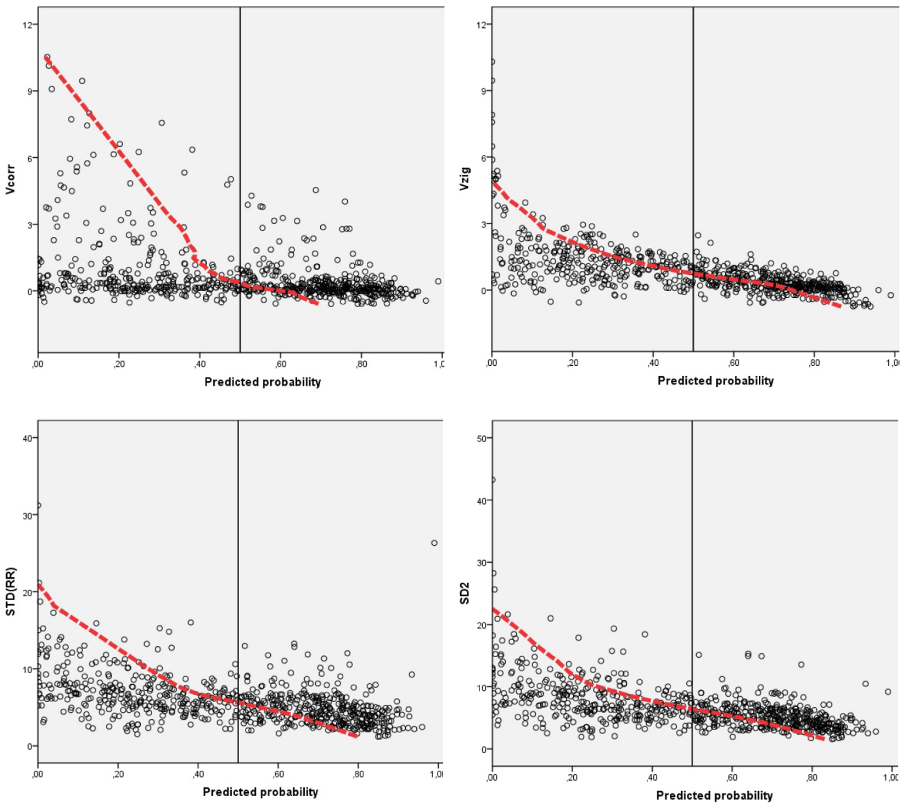


Fig. 3. Behaviour of model data in terms of predicted probability for each variable. Predicted probability values from 0 to 0.5 mean that physical demand predominates in the task, since values from 0.5 to 1 mean that cognitive demand predominates.

$$k = -0.419 \cdot V_{corr} - 1.123 \cdot V_{zig} + 0.216 \cdot V_{GRST} + 0.235 \cdot STD - 0.460 \cdot SD2 + 2.536. \quad (3)$$

The cognitive games seems to stress more than physical ones, as it considerably reduces the variability of heart rate that is related with higher demands. However, the effect on GSR seems to be similar. On the other hand, the physical games seems to be more engaging, producing greater valence responses, in both sides, positive (EMGz) and negative (EMGc) emotions, although being more similar in the case of negative valences.

4 Conclusions

Physiological measurements allow to detect the physical workload and the cognitive demands. As it was find in past studies, ECG variables are a good predictor of workload type. In addition, other physiological variables have been checked that can contribute to get more information about the type of workload. The designed model has a high classification rate (>75%) and it is comprehensible from physiological point of view, using a reduced a set of parameters.

Further research is necessary to get in depth about the nature of the cognitive workload, especially when the demands are higher or are really mixed. Physiological response and its relations with workload will provide valuable information for designing products and services, and also for training cognitive and physical capabilities of people with disabilities or senior citizens.

References

1. Stramler, J.H.: The Dictionary for Human Factors/Ergonomics. CRC Press, Inc., Boca Raton (1993)
2. Mital, A., Govindaraju, M.: Is it possible to have a single measure for all work? Int. J. Ind. Eng.-Theory Appl. Pract. **6**(3), 190–195 (1999)
3. Kohlisch, O., Schaefer, F.: Physiological changes during computer tasks: responses to mental load or to motor demands? Ergonomics **39**(2), 213–224 (1996)
4. Fredericks, T.K., Choi, S.D., Hart, J., Butt, S.E., Mital, A.: An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. Int. J. Ind. Ergon. **35** (12), 1097–1107 (2005)
5. Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. IEEE Trans. Biomed. Eng. **3**, 230–236 (1985)
6. Heino, A., Van der Molen, H.H., Wilde, G.J.S.: Risk-homeostatic Processes in Car Following Behaviour: Electrodermal Responses and Verbal risk Estimates as Indicators of the Perceived Level of Risk During a Car-driving Task (Report VK 90-22). Traffic Research Centre, University of Groningen, Haren (1990)
7. Rubio-Valdehita, S., Díaz-Ramiro, E.M., López-Higes, R., Martín-García, J.: Effects of task load and cognitive abilities on performance and subjective mental workload in a tracking task. Ann. Psychol. **28**(3), 986–995 (2012)
8. Dreiseitl, S., Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review. J. Biomed. Inform. **35**(5), 352–359 (2002)

Predicting Stimulus-Driven Attentional Selection Within Mobile Interfaces

Jeremiah D. Still[1(✉)], John Hicks[1], Ashley Cain[1], and Dorrit Billman[2]

[1] Department of Psychology, Old Dominion University, Norfolk, VA, USA
{jstill, jhick008, acain001}@odu.edu
[2] Research Foundation, San Jose State University, San Jose, CA, USA
dorrit.billman@nasa.gov

Abstract. Masciocchi and Still [1] suggested that biologically inspired computational saliency models could predict attentional deployment within webpages. Their stimuli were presented on a large desktop monitor. We explored whether a saliency model's predictive performance can be applied to small mobile interface displays. We asked participants to free-view screenshots of NASA's mobile application Playbook. The Itti et al. [2] saliency model was employed to produce the predictive stimulus-driven maps. The first six fixations were used to select values to form the saliency maps' bins, which formed the observed distribution. This was compared to the shuffled distribution, which offers a very conservative chance comparison as it includes predictable spatial biases by using a within-subjects bootstrapping technique. The observed distribution values were higher than the shuffled distribution. This suggests that a saliency model was able to predict the deployment of attention within small mobile application interfaces.

Keywords: Human-computer interaction · Mobile interface · Cognitive engineering · Saliency model · Visual search

1 Introduction

Interface designers attempt to develop digital products that are easy to search. It is apparent which designers are successful because interface search seems intuitive and easy. In these cases, we are guided by top-down (prior knowledge and experiences) and by bottom-up (stimulus-driven) influences. Top-down influences are most often considered during human-centered development (e.g., mental models, conventions). Designers make searches effortless by thoughtfully selecting familiar objects and placing them in familiar locations. However, designers often overlook bottom-up influences. These stimulus-driven properties are inherent in an interface. Bottom-up influences are automatic, occur early in the visual processing system, and operate independently of a user's expectations. For example, searching for a yellow dot amongst many green dots is easy. The number of green dots does not modulate task performance because the processing of the features within the search display is completed in parallel [3]. Our visual system is biased towards selection of regions containing visually unique information. Unfortunately, cognitive processes like these are not open for introspection, making it

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_24

difficult for designers to identify stimulus properties and estimate the influence of saliency across an entire display. However, designers can employ a biologically inspired computational saliency model to reveal the influence of low-level visual features.

1.1 Saliency Research

The cognitive psychology literature has shown stimulus-driven influences such as visual saliency reduce search times and often facilitate successful task completion [4]. According to Theeuwes [3, 5], visually salient objects capture attention even when they are irrelevant to the current task demands. For instance, search efficiency decreases when there is a salient distractor in the display, even when participants are instructed to ignore the distractors. These findings suggest that salient objects are difficult to ignore, highlighting the importance of recognizing the influence of saliency within interfaces.

It can be difficult to determine varying degrees of salience within complex and visually rich images (like nature scenes). The vision science community has addressed this issue by developing computational models to measure salience (e.g., [2]). These models capture saliency by detecting local feature contrast differences across major bottom-up dimensions (specifically: intensity, color, and orientation). Considering regional uniqueness appears to be an effective means to predict attention. Parkhurst et al. [6] demonstrated that a saliency model was able to predict fixations within images of home interiors, buildings, natural landscapes, and fractal images.

Masciocchi and Still [1] showed that a saliency model could predict stimulus-driven selection within web pages and suggested that users tend to look at regions of high salience before regions of low salience [7]. Co-locating essential elements with higher salience can nudge users to attend to certain areas over others. Split second differences can become important for interactions that are brief. It might mean the difference between an element being recognized or completely missed.

1.2 Mobile Display Research

Usage of mobile devices has grown at an astonishing rate. Small display interactions have become ubiquitous and a major design focus for interface developers. Often interactions with smaller displays amplify usability issues [8]. According to Christie et al. [9], the physical screen size of an interface can have a larger impact on task performance than actual task complexity. Other researchers have found a variety of mobile design considerations when attempting to improve the usability of mobile systems [10–12].

In the current study, we explore whether a saliency model can be applied to small mobile displays. The previous literature explored images displayed at 43.18 cm or larger [6, 7]. It is possible that display size impacts the guidance visual saliency provides. Small displays mostly fall within our fovea requiring only a few eye movements to explore the interface. Nevertheless, stimulus-driven influences ought to continue guiding attention. This study is the first to explore whether saliency can account for attentional selection within small mobile interfaces.

2 Method

2.1 Participants

Forty undergraduate students from Old Dominion University participated in the study (28 females; 38 right-handed; 35 English is native language). The University Institutional Review Board approved all study procedures.

2.2 Materials and Equipment

The stimuli were composed of 50 screenshots taken from NASA's Playbook mobile application [13]. Playbook helps astronauts schedule operations through timelines. Playbook images were displayed on a monitor at an iPad mini size (a diagonal of 20 cm). The images were visually homologous containing colored blocks and text (see Fig. 1). A Tobii X3-120 system recorded eye movements. This eye tracker gathers binocular data and performs tracking using both dark and bright pupil data at a sampling rate of 120 Hz. The system's tracking accuracy was (M = 0.70°, SD = 0.27°). Further, Tobii Studio (3.4.6) was used to present the stimuli and define fixations.

The saliency value maps were created by employing Harel et al. [14] MATLAB implementation of the Itti et al. [2] algorithm. This original saliency model (targeting the color, intensity, and orientation channels) identifies the degree of visual saliency for every pixel within an image. The model was run on each image, and the maps were normalized by dividing all values by the maximum value of that map and multiplying by 100. Low values indicate regions of the image that are low in salience (unlikely to be fixated) while high values indicate regions high in salience (likely to be fixated). The fixation x, y locations were used to locate and extract values from the saliency map.

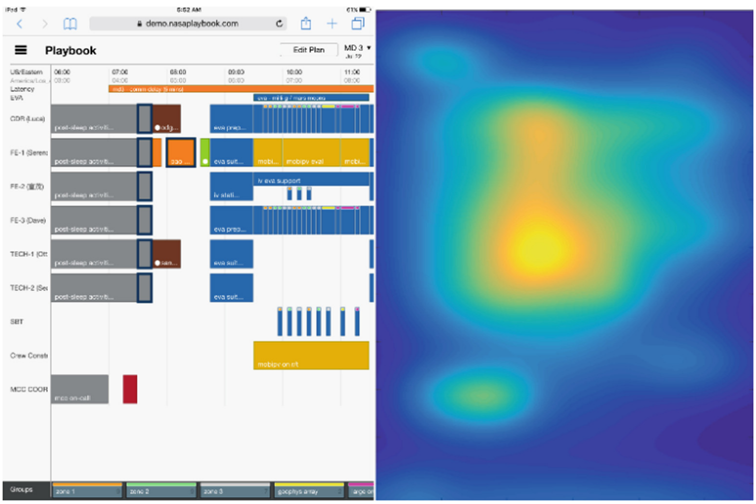


Fig. 1. The original playbook screenshot showed with its corresponding saliency map. Warmer regions within the map indicate higher values and colder regions lower values.

2.3 Procedure

Participants were simply instructed to view a series of Playbook app screenshots. The experiment began with two nine-point calibration sequences to set up the system and determine tracking accuracy. Each trial started with participants fixating on a central fixation cross. After a short pause, a randomly selected screenshot was shown for 3000 ms. Then, the fixation cross display reappeared signaling the start of the next trial. This study took about 10 min.

3 Results

The fixation x, y locations were used to select values from the saliency maps' bin, which formed the observed distribution. A bootstrapping technique was used to assess a very conservative chance performance (that includes individual spatial biases). Using this technique, each participant's x, y fixations and corresponding saliency values are extracted from every image map except for the map from which the fixations belong. These data formed a within-in subject 'shuffled' distribution.

A 2 (Distribution: observed, shuffled) X 6 (Fixation: 1, 2, 3, 4, 5, 6) repeated measures ANOVA, with Greenhouse-Geisser corrections, was employed to determine whether the difference between distributions varied by fixations. The main effect of distribution was significant, $F(1, 39) = 133.49$, $p < .001$, $\eta_p^2 = .77$. The observed distribution ($M = 45.39$, $SD = 7.67$) contained higher values than the shuffled distribution ($M = 41.24$, $SD = 7.13$). The main effect of fixation was also significant, $F(3, 117) = 13.48$, $p < .001$, $\eta_p^2 = .26$. The interaction between distribution and fixation was not significant, $F(4,165) = 0.34$, $p = .86$, $\eta_p^2 = .01$. (see Fig. 2). Post hoc tests using Bonferroni adjustments revealed that the first fixation (ps < .001) was associated with significantly higher saliency values than the other fixations. Together, these findings suggest that a saliency model can be employed to predict the deployment of attention within small mobile application interfaces (compare Figs. 1 and 3).

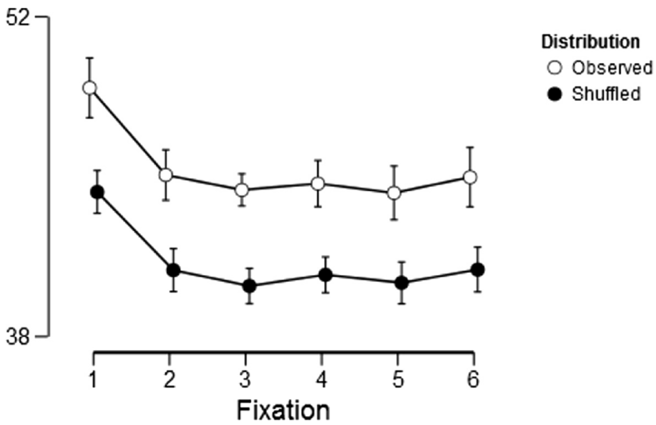


Fig. 2. Mean saliency values at the first six fixation locations for the observed and shuffled databases. The figure error bars represent 95% confidence intervals.

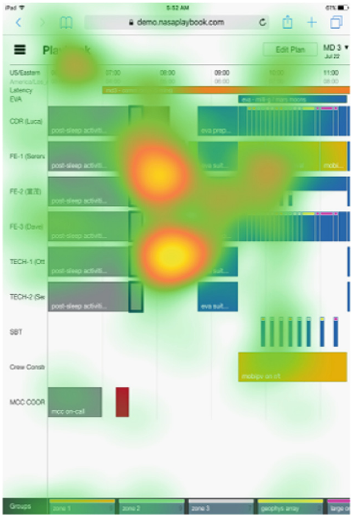


Fig. 3. The heatmap reflects the participants' fixations during the first 2000 ms of viewing this interface. Warmer regions within the map indicate more fixations and colder regions fewer fixations.

4 Conclusion

Designers strive to create intuitive interface interactions. One way to achieve this goal is to consider the influence of visual saliency on searches. Visual saliency can be determined by employing a computation saliency model; it determines the probability of saliency, and therefore the likelihood of fixation, for each pixel in an image. Masciocchi and Still [1] suggested that saliency can predict attentional deployment within webpages. Further, they showed that users tend to fixate on high salience regions [7]. Notably, the model appears to be better at predicting the first fixation than later fixations [similar to 6]. This finding is not surprising given stimulus-driven influences occur early. The present study explored whether a saliency model could predict attentional deployment within mobile interfaces (previous research used stimuli displayed on a large monitor). Therefore, we displayed NASA's Playbook application at under half the size of stimuli presented in previous studies; this display size simulates the actual size the application would be displayed at during typical usage. We found that a saliency model was able to predict fixations within much smaller displays. With this knowledge, designers ought to utilize the model to reveal how stimulus-driven influences are guiding their users' searches within mobile interfaces.

4.1 Future Work

Employing a saliency model ought to help designers make decisions about the automatic deployment of attention that are more informed. Clearly, formal models can facilitate design decisions, but more work is needed. Rosenholtz et al. [15] suggest specifically that more development is needed to bridge the research and practice gap. For instance, the model needs to go beyond marking a pixel's saliency to recognizing design elements and suggesting impactful feature modifications. Even with these advances, research will be needed to determine when this model should be employed during the development process.

We propose that this salience model ought to facilitate lower cost and faster prototype development. Currently, designers have to collect eye-tracking data to determine how an interface is guiding users during a search. This requires time, eye tracking expertise, special equipment, and participants throughout the iterative prototype development cycles. The employment of this computation model ought to decrease the number of development cycles needed to meet performance requirements and this could be done without the need for special equipment or participants. However, designers ought not abandon eye-tracking altogether. We recommend that a saliency model be used during initial high fidelity prototyping. Then, during the 'final' summative testing phase, eye-tracking can be employed to verify users are looking at critical elements in the design.

Even with advances in understanding the contributions of saliency to the deployment of attention in interfaces, research that explores the interaction between bottom-up and top-down processing is still needed. We are assuming, based on decades of research in cognitive psychology, that stimulus-driven influences (bottom-up) continue to guide attention during a directed interface search (top-down). Although it is reasonable to assume that top-down influence would have a larger impact on search performance than bottom-up influences, bottom-up influences like saliency ought to have an impact regardless of the search activity. If future research shows this to be true, the applications for the saliency model reach beyond predicting free-viewing behavior or mindless 'internet surfing.' We could use the model to facilitate a variety of searches ranging from life-critical systems to e-commerce sites. If important interface elements are not salient, changes to the design could be made to increase their salience. This redesign process ought to help designers direct users' attention to certain element over others. This attentional guidance might result in greater situation awareness within a life-critical system or higher profit margins within e-commerce sites.

Another advantage to using this type of attentional guidance is that it might offer a less invasive means of communication. Currently, designers employ alerts that exogenously demands attention by making an element blink or shake. This approach takes advantage of low-level automatic neurological orienting mechanisms, which is acceptable if the interface element must immediate capture the user's attention. However, most of the time, a system either is not designed to identify what ought to be important or if it is designed to "guess" what the user should prioritize, it may be wrong an unacceptable (and annoying) amount of the time. Saliency, as defined here, does not demand immediate attentional orientation; it simply guides or nudges users to some areas or others. We hope future research explores the practical application of using saliency to nudge users to attend to certain interface elements over others.

Acknowledgments. The authors thank Steven Hillenius for providing access to the Playbook software through a demo site. In addition, we thank the Virginia Space Grant Consortium for financially supporting this research.

References

1. Masciocchi, C.M., Still, J.D.: Alternative to eye tracking for predicting stimulus-driven attentional selection within interfaces. J. Hum.-Comput. Interact. **34**, 285–301 (2013)
2. Itti, L., Koch, C., Niebur, E.: A model of saliency-based fast visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254–1259 (1998)
3. Theeuwes, J.: Top-down search strategies cannot override attentional capture. Psychon. Bull. Rev. **11**, 65–70 (2004)
4. Wolfe, J.M.: Guided Search 4.0: Current Progress with a Model of Visual Search. In: Gray, W. (eds.) Integrated Models of Cognitive Systems, pp. 99–119, New York, Oxford (2007)
5. Theeuwes, J.: Perceptual selectivity for color and form. Percept. Psychophys. **51**, 599–606 (1992)
6. Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vis. Res. **42**, 107–123 (2002)
7. Still, J.D., Masciocchi, C.M.: A saliency model predicts fixations in web interfaces. In: 5th International Workshop on Model Driven Development of Advanced User Interfaces Computer-Human Interaction Conference, pp. 25–29 (2010)
8. Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G.: Improving web interaction on small displays. J. Comput. Netw. **31**, 1129–1137 (1999)
9. Christie, J., Klein, R.M., Watters, C.: A comparison of simple hierarchy and grid metaphor for option layouts on small-size screens. Int. J. Hum.-Comput. Stud. **60**, 564–584 (2004)
10. Garcia-Lopez, E., Garcia-Cabot, A., de-Marcos, L.: An experiment with content distribution methods in touchscreen mobile devices. J. Appl. Ergon. **50**, 79–85 (2015)
11. Lim, J.J., Feria, C.: Visual search on a mobile device while walking. In: Mobile HCI Conference, pp. 295–304, San Francisco, CA (2012)
12. Watters, C., Duffy, J., Duffy, K.: Using large tables on small display devices. Int. J. Hum.-Comput. Stud. **58**, 21–37 (2003)
13. Marquez, J.J., Pyrzak, G., Hashemi, S., Ahmed, S., McMillin, K., Medwid, J., Chen, D., Esten, H.: Supporting real-time operations and execution through timeline and scheduling aids. In: 43rd International Conference on Environmental Systems, pp. 1–11. ICES (2013)
14. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: NIPS, vol. 1, pp. 5–12 (2006)
15. Rosenholtz, R., Dorai, A., Freeman, R.: Do predictions of visual perception aid design? ACM Trans. Appl. Percept. **8**, 1–20 (2011)

Evaluating ANN Efficiency in Recognizing EEG and Eye-Tracking Evoked Potentials in Visual-Game-Events

Andreas Wulff-Jensen(✉) and Luis Emilio Bruni

Department of Architecture, Design and Media Technology,
Aalborg University Copenhagen,
AC Meyervænge 15, 2450 Copenhagen, Denmark
{awj,leb}@create.aau.dk

Abstract. EEG and Eye-tracking signals have customarily been analyzed and inspected visually in order to be correlated to the controlled stimuli. This process has proven to yield valid results as long as the stimuli of the experiment are under complete control (e.g.: the order of presentation). In this study, we have recorded the subject's electroencephalogram and eye-tracking data while they were exposed to a 2D platform game. In the game we had control over the design of each level by choosing the diversity of actions (i.e. events) afforded to the player. However we had no control over the order in which these actions were undertaken. The psychophysiological signals were synchronized to these game events and used to train and test an artificial neural network in order to evaluate how efficiently such a tool can help us in establishing the correlation, and therefore differentiating among the different categories of events. The highest average accuracies were between 60.25%–72.07%, hinting that it is feasible to recognize reactions to complex uncontrolled stimuli, like game events, using artificial neural networks.

Keywords: Artificial neural network · Machine learning · Electroencephalogram · Eye-tracking · Games · Pupillometry · Game events · Psychophysiology

1 Introduction

Games can be viewed as complex multimodal stimuli with visually appealing content and multilayered audio stimuli, which pose cognitively demanding tasks on the players. All of which the gamer perceives holistically eliciting embodied reactions towards the next course of actions. The perception process and the immediate unconscious physiological reaction of the body are rather multifaceted and can tell us a lot about the state of the person and the stimulus he is receiving. However, taking two complex streams of signals and relating them directly to each other comes with a huge amount of uncertainty, as psychophysiological signals have the property of being mapped many to one [1]. Videogames can provide us with a controllable environment in which we can design categories of events that can be repeated or undertaken multiple times. Then,

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_25

there will be the possibility that certain observed patterns of psychophysiological signals can be correlated to specific categories of events.

The psychophysiological signals considered in this study are Electroencephalogram (EEG), Eye-tracking and pupillometry. These measurements can generate large amounts of data, thus investigating them manually for patterns and relations might be troublesome and time consuming. Therefore, inquiries in deep machine learning algorithms, such as artificial neural networks (ANN) [2–4], are relevant in this domain.

2 Related Works: Classifying EEG and Psychophysiological Data Through the Use of ANN

The idea of using ANNs to classify and recognize psychophysiological data streams is not novel, as it has been looked at for a few decades, particularly for detecting EEG artefacts [5]. In the literature, ANNs have been used to detect the EEG P300 event-related-potential (ERP) components elicited by counting and expectancy [6–8] and to recognize EEG frequency bands, heart rate, eye-tracking and respiration in relation to working memory [9–11]. In the ERP experiments the subjects counted the attended sound frequencies [6] or the appearance of preselected images [7]. In the expectancy experiment, the subjects where prompted to expect “the right” match of congruent audio and visuals [8]. These studies provided the ANN with simple binary classification problems, either target ERP or non-target ERP. The accuracy in these studies reflects their simplicity, as it varies between 69% [8] to 97.3% [7]. Furthermore, the datasets and ANNs used in these studies have been enhanced by a few algorithms to improve the performance such as grand averaging across EEG channels [8], back propagation [7] and autoregressive feature selection [7]. Apart from these works a review article has debated the accuracies of ANN trained on EEG data [12]. The article concluded that there is a promising future for using ANN to recognize and classify EEG.

In the quantitative EEG realm, [11] addresses the recognition of working-memory-load, both, within a single task and across different tasks. There were three memory tasks with two difficulty levels each: one low and one high. Classifying these difficulty levels within each task gave accuracies between 86–89%. Moreover, when the ANN was trained on one task, but tested with data from another task, the average accuracy was 44.3% [11]. The same accuracy tendencies have been found within workload classification of real life challenges such as air traffic monitoring, in which eye-tracking, heartrate and respiration rate have been used as well [9, 10].

In the videogames domain, one investigation included predicting levels of enjoyment in a game using heartrate, which yielded 76% accuracy through cross-validation [13]. Another study with heartrate and skin conductivity features from dissimilar games attempted to recognize user experience across both games [14]. When training the ANN with the heartrate features, the accuracies fluctuated between 58.22% and 66.30%. Likewise, the ANNs trained on skin conductivity yielded accuracies fluctuating between 57.93% and 64.67% [14]. The studies seem to consider mostly the tonic response (e.g.: continuous EEG) and not so much the subtle phasic responses (e.g. ERPs) in relation to different events and actions, which are occurring multiple times during a game session. These might include some cognitive, psychological or affective processes uniquely bound to the recorded signal thus possibly detectable and recognizable by the ANN. A demonstration of segmenting a game into different events and analyzing them through descriptive statistics can be found in [15]. However, the similarity of the recordings within categories of events was not recognized through a machine learning algorithm.

3 The Game Stimuli

For the present study, we chose the 2D platformer game Super Mario Bros for the testbed, given that it has been the subject of numerous affective-games and AI studies, and due to the possibility to control and generate content within it [16–21]. This is a rather simple game with recurring events and a limited palette of controls. The game is about controlling the game-character Mario through different levels from left to right. These levels are created with platforms containing boxes, collectables and enemies. Some of the collectables allow Mario to morph and evolve, providing him with new powers such as enhanced strength, the possibility of shooting fireballs and acquiring extra life.

For this study the game has been dissected into different event components elicited through the player's interactions (see Table 1). These events were classified in different

Table 1. Event categories and type of events found in Super Mario Bros.

Event categories	Event types
Mario gaining or loosing qualities when colliding with an object:	Colliding with mushroom: enhances Mario's strength
	Colliding with a flower: enables the use of fireballs
	Colliding with an enemy: losses one of the gained qualities
Box collision events:	Hitting an empty box
	Hitting a box containing a coin
	Hitting a box containing a mushroom
	Hitting a box containing a flower
Killing enemy events:	Killing an enemy by a stomp
	Killing an enemy by a fireball
	Killing an enemy by a shell
Ending level events:	Dying by an enemy
	Winning a level
	Losing a level
	Failing to jump a gap
Automated ending level events:	Automated jump animation when dying
	Automated transition to black
	Automated transition back to the game
Avoiding dying in a gap event	Succeeding a jump
Object collision without effect	Picking up a coin

categories according to semantic similarity. For instance, the “box collision events” category consists of all the events which are triggered by Mario hitting a box when jumping into it.

In order to ensure that each subject was exposed to enough repetitions of each event-type for measuring purposes (i.e. ERP and gaze averaging and classification), ten random but yet homogeneous game levels were generated. These levels were composed of a jumping area, where the playing character could fall into gaps, and a fighting area populated with enemies and with the presence of boxes (the source code used to generate the ten levels are re-elaborated from [16]).

4 Experimental Setup

The game was played on a 27-inch iMac. The eye-tracking device, an eye-tribe eye-tracker [22], was connected to this computer. Temporal onset information about the game events and the eye-tracking signals were send through User Datagram Protocol Networking (UDP) to a second computer which synchronized the signals together with the EEG recordings in a Matlab Simulink environment. The device used to record the EEG signals was g.Tec g.Gammabox and g.Tec g.USBAmp [23]. The subject was situated 110 cm from the iMac and the distance to the eye-tracker was 65 cm. This setup promotes heightened eye-tracking accuracy, because the screen has a smaller visual angle than what has been tested previously, and which have been found to give poor accuracy along the borders of the screen [24, 25] (the calculation of the visual angle can be found in [26]). The sample rates from the different signals were 256 Hz for the EEG and 60 Hz for the Eye-tracker. Since the computer recording the EEG is the master recorder there is only an eye-tracking sample for every 4.26 EEG samples. To avoid null signals, the EEG samples with no corresponding Eye-tracking signals, have been assigned zeroes in replacement of the null eye-tracking signal.

The eye-tribe provided X-Y coordinates for the gaze position and pupil diameter for the left and right eye, generating in total four signal streams.

The EEG was recorded from 16 channels spatially divided to cover most of the scalp within the 10–20 system. These channels were: F3, FZ, F4, T7, C3, CZ, C4, T8, P7, P3, PZ, P4, P8, O1, OZ and O2. In total, these streams generated 20 different data signals which could be influenced by the different game events that the subject could encounter.

5 Subjects and Experimental Procedure

31 subjects were recruited through non-probabilistic convenient sampling with the help of social media and the Aalborg University intranet at the Copenhagen campus. They were requested to fill out a consent form, which informed them that brain-signal and gaze would be recorded in a non-harmful way. A short pre-test questionnaire was administered to get basic demographic data (gender, age and occupation) and asking whether the participant had consumed any alcohol or caffeine, and whether he or she had been diagnosed any relevant disorder (they were advised about this in the

recruiting process). Be that the case, his or her dataset has been excluded when performing across-participant analysis, but the experiment was anyway conducted. Lastly, the subjects were asked how often they play video games and whether they had played Mario, as this could have an impact on the data.

The Subjects comprised 7 females and 24 males, average age = 26.16 (SD = 6.9). 27 were students or interns (18 graduates and 9 undergraduates). Their gaming behavior showed that 29 of them have played Super Mario Bros before, and 21 play video games on a weekly basis. This makes the sample a homogenous group of videogame playing students in their twenties.

After the pre-test questionnaire, the subjects were asked to sit in front of the computer which ran the Mario game. The EEG cap was mounted and a conductive gel applied. After that, the signals were checked to assure proper conductivity and impedance and that the signals being recorded were EEG. The subjects were asked to sit as comfortable and still as possible while anyway being able to look and interact with the computer running the Mario game. A native eye-tracking calibration program was executed to calibrate the device to the eyes of the particular user.

After all the data acquisition devices were calibrated, the Simulink patch, which received and recorded the psychophysiological signals and the triggers for each event, was started. The gameplay recording program was turned on. In order to record the EEG baseline, each subject was exposed for one minute to a grey screen with a fixation cross in the middle. Right after the baseline recording the Mario game was started. While the subjects were playing the game, the experimenter monitored the EEG and eye-tracking signals, maintaining a log of any eventual external sources of artifacts or noise in order to exclude these parts of the recordings from the analysis. After the subjects had played the game for 10 levels, they were debriefed and thanked.

6 Data and ANN Preparation

6.1 Data Preparation

The recorded data was segmented into 3-seconds' epochs for each of the relevant game events: 1 s before the onset and 2 s after the onset. This range was chosen to avoid losing any important information from either the EEG or the Eye-tracking [27–31, 41]. The EEG data was filtered with a notch filter of 50 Hz [27] and a band pass filter between 0.1 and 100 Hz. Afterwards, a visual inspection was conducted in order to remove the epochs containing eye and muscle artifacts. All the remaining epochs were counted and linked to their respective game events. The events having more than 20 epochs in one play session, and which appeared with such frequency in at least 10 subjects, were taken into consideration for further analysis. Through this process four different events from Table 1 were considered for analysis: Hitting an empty box (EB), Hitting a box containing a coin (CB), Killing an enemy with a stomp (KS) and Killing an enemy with a fireball (KF). Furthermore, the exclusion process reduced the number of subjects from 31 to 13. In these subjects the amount of repetitions for each of the four events varied. To reduce the bias of variance, the number of these repetitions per subject was reduced to the lowest common denominator – which was 20 repetitions – so each of the 13 subject's dataset consisted of 20 EB, 20 CB, 20 KS and 20 KF.

Once the epochs for the relevant events were identified, the quality of the signal was further improved by applying the Independent Component Analysis (ICA) [32, 33] in order to detect possible artifacts not caught in the previous procedure. The same segmentation procedure was applied to the Eye-tracking signals, in order to link both data streams to their respective stimulus/event.

6.2 ANN Preparation

To prepare the ANN[1] for epoch classification, the model was defined in terms of layers and neurons, how many data points from each epoch should the ANN considered and the number of iterations [2, 3]. Through iterative testing of procedural generated ANN models [18, 34], the models shown in Table 2 were optimized to find the highest accuracy for EEG or Eye-tracking data. Furthermore, to reach the best possible result, and to take random weight initiation into account, all the models were tested 10 times. All models were tested for 1000 iterations. To reach the amount of optimal iterations, the resulting 3 best models were then tested every 100 iterations up to 10000 iterations.

Table 2. ANN models optimized for EEG and Eye-tracking epoch data

Modality	Input layer (amount of data points)	Hidden layers	Output layers	Learning function	Activation function	Iterations	Number of tested models
EEG	256 (1000 ms after stimuli onset)	16, 10, 9	4	Resilient backpropagation [37]	Tanh for hidden layers. Sigmoid for output layer	7700	4341
Eye-tracking	768 (1000 ms before and 2000 ms after stimuli onset)	16, 12	4	Resilient backpropagation [37]	Tanh for hidden layers. Sigmoid for output layer	2400	1000

6.3 Data Modelling for Recognition

Each subject yielded 20 datasets: 16 EEG channels and 4 eye-tracking data streams. Each of these datasets contained the segmented epochs of the four game events that were repeated 20 times in total. In order to be submitted to the ANN each of these 20 datasets per subject were divided into a training-dataset (70% of the total game event epochs – 14 for each type of event) and a test-dataset (30% of the total game event epochs – 6 of each type of event). Furthermore, through the training session the K Fold Cross validation (K = 10) was applied, which has previously been found as a good

[1] The ANN used in this study was created through the Encog Library [40].

technique to avoid overfitting [35, 36]. Once the algorithm reached the amount of prefixed iterations, it was tested with the test-dataset. Each dataset was run through the ANN 10 times in order to compensate for different weight initiations. This yielded 10 different levels of accuracy and precision per single channel per single event.

With this data structure we could then analyze whether there would be significant differences in either accuracy or precision of the ANN model when matching EEG or Eye-tracking patterns to their respective event types. In other words, we were interested in analyzing whether the data from the different EEG channels and eye-tracking parameters would result in similar levels of accuracy and precision by the trained ANN. To access the different accuracies and precisions between each of the psychophysiological datasets with respect to each event type, Kruskal–Wallis one-way ANOVA test (K-WANOVA) ($p < 0.05$) [37] was used.

7 Results

The results of the analysis are documented in the tables on the following pages. They present an overview of different accuracy and precision benchmarks. These include the highest/lowest results for both, individual subject/channel (e.g. the accuracy in channel C4 in a single subject) and the averaged across subjects per channel (e.g. (13 subjects X EEG channel C4's accuracy)/13). All the averages were compared in all permutations of channel-pairs (i.e. C3-P3, Pz-P3, etc.) through K-WANOVA to find significant differences between them. The tables only present the pairs in which there was a significant difference ($p < 0.05$). Similar comparisons are presented for the eye-tracking parameters.

7.1 Accuracy Levels in Matching EEG Epochs to Game Events

A K-WANOVA comparison was run for each of the four game events. When comparing all 16 EEG channels (averaged for the 13 subjects) for each event type, there

Table 3. Comparisons of accuracies in different EEG channels per game events

Game events	Averaged channels pairs with significant differences $P < 0.05$	Channel with highest averaged accuracy	Channel with highest accuracy	Channel with lowest average accuracy	Channel with lowest accuracy
EB:	C3 > P3, Pz > P3	C4 = 66.53%	F3 = 74.16%	P3 = 63.81%	O1 = 54.58%
	P9 > P3				
CB:	T7 > P3	T = 63.01%	C3 = 72.50%	P3 = 60.51%	C4 = 52.91%
KS:	Oz > C3, Oz > Pz	Oz = 64.19%	Fz = 71.66%	Pz = 60.51%	P7 = 52.50%
	Oz > P8, Oz > O2				
	T8 > Pz, T8 > O2				
KF	F3 > C4, T7 > C4	Oz = 72.07%	Pz = 77.08%	C4 = 70.06%	O2 = 62.08%
	Oz > C4				

were no significant differences between the accuracies yielded by the ANN when matching the epochs to the respective game events (*P* > *0.05*). However, pairwise differences were found as shown in Table 3 (game event abbreviations follow from Sect. 6.1)

7.2 Precision Levels in Matching EEG Epochs to Game Events

The K-WANOVA tests showed no significant difference of precision levels when comparing all 16 averaged channels in relation to each of the four game events (*P* > *0.05*). However, as in the case of accuracy levels, pairwise significant differences occurred. See Table 4 for results.

Table 4. Comparisons of precision in different EEG channels per game events (channel-pairs marked with (*) showed a significance level p < 0.01)

Game events	Averaged channels pairs with significant differences *P* < *0.05*	Channel with highest averaged precision	Channel with highest precision	Channel with lowest average precision	Channel with lowest precision
EB:	F3 > Fz, F4 > Fz	F4 = 25.98%	C3 = 51.12%	C4 = 17.69%	C4 = 1.53%
	F4 > T7, F4 > C4				
CB:	O2 > P3, O2 > P8	O2 = 27.35%	O2 = 45.54%	P3 = 20.37%	Pz = 4.72%
KS:	Cz > F3, Oz > F3	Oz = 27.66%	Fz = 45.73%	Pz = 20.32%	P7 = 5.00%
KF	F3 > O1, Cz > P4*	Pz = 22.58%	Fz = 60.00%	O1 = 10.52%	Fz, P3, P4 = 0.00%
	Cz > O1*, T8 > O1				
	Pz > P3, Pz > P4				
	Pz > O1*, Pz > C4				
	P8 > O1, O2 > C4				
	O2 > P4*, O2 > O1*				

7.3 Accuracy Levels in Matching Eye-Tracking Epochs to Game Events

Four K-WANOVA tests compared accuracy differences between all averaged eye-tracking parameters with respect to each of the four game events. Only one of these tests showed a significant difference – between the pupil dilation of the left eye and the three other parameters in the KF game event (*P* < *0.01*). However, pairwise inspection revealed significant differences, which can be found in Table 5.

7.4 Precision Levels in Matching Eye-Tracking Epochs to Game Events

The K-WANOVA tests showed no significant difference of precision levels when comparing all 4 averaged eye-tracking parameters for each game event (*P* > *0.05)*. Only one of the events (CB) showed a significant difference in the precision scores between left and right pupil dilation. More benchmarks can be found in Table 6.

Table 5. Comparisons of accuracies in different Eye-tracking parameters per game events. Left Pupil Dilation (LP), Right Pupil Dilation (RP), X Gaze Position (XG) and Y Gaze Position (YG). (parameters-pairs marked with (*) showed a significance level $p < 0.01$)

Game events	Averaged channels pairs with significant differences $P < 0.05$	Channel with highest averaged accuracy	Channel with highest accuracy	Channel with lowest average accuracy	Channel with lowest accuracy
EB:	LP > YG	LP = 60.25%	RP = 67.5%	YG = 55.96%	YG = 46.66%
CB:	LP > RP	LP = 60.80%	XG = 70.00%	RP = 57.85%	XG = 46.67%
KS:	LP > YG	LP = 61.63%	XG = 69.16%	YG = 55.57%	YG = 37.91%
KF	LP > RP* LP > YG*	LP = 70.99%	LP = 75.00%	XG = 64.16%	XG = 55.41%
	LP > XG*				

Table 6. Comparisons of precisions in different Eye-tracking parameters per game events

Game events	Averaged channels pairs with significant differences $P < 0.05$	Channel with highest averaged precision	Channel with highest precision	Channel with lowest average precision	Channel lowest precision
EB:	None	RP = 27.58%	LP = 42.54%	XG = 23.28%	XG = 12.59%
CB:	LP > RP	LP = 27.94%	XG = 42.25%	RP = 24.03%	YG = 10.82%
KS:	None	RP = 26.76%	RP = 39.44%	YG = 23.30%	LP = 13.16%
KF	None	YG = 26.13%	YG = 48.33%	RP = 18.89%	XG = 9.08%

8 Discussion

The accuracies for the KF event (killing with a fireball) were always higher than the accuracies for the other events. However, its precision was always lower. This observation could suggest that the epochs for that event look different than the epochs of other events, but the lower precision could suggest that the KF's epochs are quite dissimilar to each other. This remark can be validated when looking into how this game event occurs. During the game, multiple enemies are spawned in the same area provoking the user to shoot more fireballs after each other. In case more than two fireballs hit enemies in a very short time frame, the epochs of each of these hits will be influenced by each other. Furthermore, the other events have higher average precisions than KF, thus suggesting that epochs related to these events are more homogeneous. However, they are also somewhat similar across the three events as the accuracies are lower, implying that epochs from one event may be matched to another. Another factor which can have huge influence on the epochs is the specific game sequence of the subject. The game events are certainly the same, but the sequence in which they occur is random. This can have a strong impact on the psychophysiological signals as they reflect the reaction to specific sequences in idiosyncratic ways. The epochs which are said to be alike might be similar in small components, and these components may appear in a different time span within the epoch. This can be due to the player constantly evaluating his options and in which order he should perform his actions [38]. However, it is encouraging to see accuracies reaching 70%, thus indicating that there might be some similar signal patterns across game events. Taking into consideration the sequences of the events could result in higher accuracies, given that these sequences have more than one element of the game in common. Such analysis could open for the exploration of the cognitive processes of the player, for example, how the particular succession of events allows the player to plan and structure beforehand his goals and his reactions to the coming targets.

In previous studies utilizing ANN to recognize EEG signals, the accuracies are higher than the highest average observed in the present study: 69%–97.3% vs 63.01%–72.07% [6–8, 11]. There can be many reasons for this, one perhaps being the complexity of the stimuli. The previous studies worked with controlled experimental procedures, either to elicit the P300 component or observe differences in the continuous EEG realm. Compared to their stimuli the game stimulus in this study is complex and uncontrollable as discussed before.

9 Conclusion

In this study, we have recorded the subjects' electroencephalogram and eye-tracking data (gaze position and pupillometry) while they were exposed to different events in a 2D platform game (Super Mario Bros). EEG and Eye-tracking signals have customarily been analyzed and inspected visually in order to be correlated to the controlled stimuli. This process has proven to yield valid results as long as the stimuli of the experiment were under complete control (e.g.: the order of presentation). In our case we had no control over the order in which these actions were undertaken. Therefore, this study addressed the feasibility of using an ANN for efficiently recognizing and correlating psychophysiological signals to complex stimuli represented by random events in a videogame context. A further study has to be carried out in order to compare different ANNs in this domain (for example the use of convolutional neural networks (CNN) [39]. The levels of accuracies (60.25%–72.07%) and precisions (22.58%–27.94%) obtained in this experiment suggest a promising avenue for exploring the automatic recognition of multimodal psychophysiological patterns and their correlation to complex random stimuli. These could be useful when creating meaningful biofeedback loops, or for online cognitive characterization of the subject.

References

1. Koelstra, S., Mühl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: DEAP: a database for emotion analysis; using physiological signals. IEEE Trans. Affect. Comput. **3**, 18–31 (2012). doi:10.1109/T-AFFC.2011.15
2. Jain, A.K., Mao, J.: Artificial neural network: a tutorial. Communications **29**, 31–44 (1996). doi:10.1109/2.485891

3. Basheer, I.A., Hajmeer, M.: Artificial neural networks: fundamentals, computing, design, and application. J. Microbiol. Methods **43**, 3–31 (2000). doi:10.1016/S0167-7012(00)00201-3
4. Andrews, R., Diederich, J., Tickle, A.B.: Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowl.-Based Syst. **8**, 373–389 (1995). doi:10.1016/0950-7051(96)81920-4
5. Kubacki, A., Jakubowski, A., Sawicki, Ł.: Detection of artefacts from the motion of the eyelids created during eeg research using artificial neural network. Robot. Meas. Tech. Adv. Intell. Syst. Comput. **440**, 447–457 (2016). doi:10.1007/978-3-319-29357-8_24
6. Wu, F.Y., Slater, J.D., Honig, L.S., Ramsay, R.E.: A neural network design for event-related potential diagnosis. Comput. Biol. Med. **23**(3), 251–264 (1993)
7. Turnip, A., Hong, K.S.: Classifying mental activities from EEG-P300 signals using adaptive neural networks. Int. J. Innov. Comput. Inf. Control **8**, 6429–6443 (2012)
8. Gupta, L., Molfese, D.L., Ravi, T.: An artificial neural-network approach to ERP classification. Brain Cogn. **27**, 311–330 (1995)
9. Wilson, G.F., Christopher, R.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Hum. Factors **45**, 635–643 (2004). doi:10.1518/hfes.45.4.635.27088
10. Wilson, G., Russell, C.: Operator functional state classification using multiple psychophysiological features in an air traffic control task. Hum. Factors **45**, 381–389 (2003)
11. Baldwin, C.L., Penaranda, B.N.: Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. NeuroImage **59**, 48–56 (2012). doi:10.1016/j.neuroimage.2011.07.047. Elsevier Inc.
12. Robert, C., Gaudy, J.-F., Limoge, A.: Electroencephalogram processing using neural networks. Clin. Neurophysiol.: Off. J. Int. Fed. Clin. Neurophysiol. **113**, 694–701 (2002). doi:10.1016/S1388-2457(02)00033-0
13. Yannakakis, G.N., Hallam, J., Lund, H.H.: Entertainment capture through heart rate activity in physical interactive playgrounds. User Model. User-Adap. Interact. **18**, 207–243 (2008). doi:10.1007/s11257-007-9036-7
14. Perez Martínez, H., Garbarino, M., Yannakakis, G.N.: Generic physiological features as predictors of player experience. Affect. Comput. Intel. Interact. 267–276 (2011). doi:10.1007/978-3-642-24600-5_30
15. Weber, R., Behr, K.M., Tamborini, R., Ritterfeld, U., Mathiak, K.: What do we really know about first-person-shooter games? An event-related, high-resolution content analysis. J. Comput.-Med. Commun. **14**, 1016–1037 (2009). doi:10.1111/j.1083-6101.2009.01479.x
16. Togelius, J., Karakovskiy, S., Shaker, N.: MarioAI (2012). http://www.marioai.org/LevelGeneration/source-code. Accessed 14 Mar 2016
17. Shaker, N., Asteriadis, S., Yannakakis, G.N., Karpouzis, K.: A game-based corpus for analysing the interplay between game context and player experience. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, vol. 6975, pp. 547–556 (2011). doi:10.1007/978-3-642-24571-8_68
18. Shaker, N., Asteriadis, S., Yannakakis, G.N., Karpouzis, K.: Fusing visual and behavioral cues for modeling user experience in games. IEEE Trans. Cybern. **43**, 1519–1531 (2013). doi:10.1109/TCYB.2013.2271738
19. Shaker, N., Shaker, M., Abou-zleikha, M.: Towards generic models of player experience. In: Proceedings, the Eleventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-2015), 191–197 (2015). http://www.aaai.org/ocs/index.php/AIIDE/AIIDE15/paper/download/11548/11378, ISBN 157735740X 9781577357407

20. Yannakakis, G.N., Togelius, J.: Experience-driven procedural content generation. IEEE Trans. Affect. Comput. **2**(3), 147–161 (2011). doi:10.1109/T-AFFC.2011.6
21. Togelius, J., Shaker, N., Karakovskiy, S., Yannakakis, G.N.: The Mario AI championship 2009-2012. AI Mag. **34**, 89–92 (2013). doi:10.1609/aimag.v34i3.2492
22. The_EyeTribe: The Eye Tribe Tracker (2015)
23. G.tec, medical engineering: Advanced Biosignal Acquisition, Processing and Analysis Products 2013–2014, pp. 1–96 (2013)
24. Kowalik, M.: Do-it-yourself eye tracker: impact of the viewing angle on the eye tracking accuracy. In: Proceedings of CESCG 2011: The 15th Central European Seminar on Computer Graphics, pp. 1–7 (2011). http://old.cescg.org/CESCG-2011/papers/Szczecin-Kowalik-Michal.pdf, ISBN 978-3-9502533-3-7
25. Ooms, K.: Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental set- ups **8** 1–20 (2015). doi:10.16910/jemr.8.1.5
26. Kosslyn, S.M.: Measuring the visual angle of the mind's eye. Cogn. Psychol. **10**, 356–389 (1978). doi:10.1016/0010-0285(78)90004-X
27. Key, A.P.F., Dove, G.O., Maguire, M.J.: Linking brainwaves to the brain: an ERP primer. Dev. Neuropsychol. **27**, 183–215 (2005). doi:10.1207/s15326942dn2702_1
28. Başar, E., Başar-Eroglu, C., Karakaş, S., Schürmann, M.: Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG? Neurosci. Lett. **259**, 165–168 (1999). doi:10.1016/S0304-3940(98)00934-3
29. Sprague, N., Ballard, D., Robinson, A.: Modeling embodied visual behaviors. ACM Trans. Appl. Percept. **4**, 11 (2007). doi:10.1145/1265957.1265960
30. Reilly, R.: Triangulating the reading brain: eye movements, computational models, and EEG. In: Current Trends in Eye Tracking Research, pp. 131–139. Springer International Publishing (2014)
31. Laeng, B., Sirois, S., Gredeback, G.: Pupillometry: a window to the preconscious? Perspect. Psychol. Sci. **7**, 18–27 (2012). doi:10.1177/1745691611427305
32. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Netw. **13**, 411–430 (2000). doi:10.1016/S0893-6080(00)00026-5
33. Makeig, S., Bell, A.J., Jung, T.-P., Sejnowski, T.J.: Independent component analysis of electroencephalographic data. Adv. Neural. Inf. Process. Syst. **8**, 145–151 (1996). doi:10.1109/ICOSP.2002.1180091
34. Shaker, N., Yannakakis, G.N., Togelius, J.: Towards automatic personalized content generation for platform games. In: AIIDE, October 2010. https://www.aaai.org/ocs/index.php/AIIDE/AIIDE10/%20paper/viewFile/2135/2546
35. Nogueira, P.A., Aguiar, R., Rodrigues, R., Oliveira, E.: Designing players' emotional reaction models: a generic method towards adaptive affective gaming. In: 2014 9th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6. IEEE. (2014). doi:10.1109/CISTI.2014.6877079
36. D'Mello, S., Chipman, P., Graesser, A.: Posture as a predictor of learner's affective engagement. In: Proceedings of the 29th Annual Meeting of the Cognitive Science Society, vol. 1 pp. 905–910 (2007)
37. MacDonald, J.H.: Kruskal-Wallis test. In: Biological Handbook of Statistics (2009)
38. Lindley, C.A., Sennersten., C.C.: Game play schemas: from player analysis to adaptive game mechanics. Int. J. Comput. Games Technol. **2008**, 1–7 (2008). doi:10.1155/2008/216784
39. Cecotti, H.: A time-frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses. Pattern Recognit. Lett. **32** 1145–1153 (2011). doi:10.1016/j.patrec.2011.02.022

40. Heaton, J.: Encog: library of interchangeable machine learning models for Java and C#. J. Mach. Learn. Res. **16**, 1243–1247 (2015)
41. Kliegl, R., Dambacher, M., Dimigen, O., Sommer, W.: Oculomotor control, brain potentials, and timelines of word recognition during natural reading. In: Current Trends in Eye Tracking Research, pp. 141–155. Springer International Publishing (2014)

Fundamental Cognitive Workload Assessment: A Machine Learning Comparative Approach

Colin Elkin(✉), Sai Nittala, and Vijay Devabhaktuni

EECS Department, The University of Toledo, Toledo, OH, USA
{Colin.Elkin,
SaiKameshwarRao.Nittala}@rockets.utoledo.edu,
Vijay.Devabhaktuni@utoledo.edu

Abstract. Mental workload remains an essential but challenging aspect of human factors, while machine learning serves as an emerging and expanding research realm to a wide variety of applications. This paper aims to comprehensively bridge the two areas by comparing present state-of-the-art machine learning approaches that are currently utilized for assessing cognitive workload, primarily artificial neural networks and support vector machines. To address and evaluate both approaches, we obtain a physiological data set used to study fear conditioning and cognitive load and format the data to focus primarily on the latter. Ultimately, the results indicate that both techniques can effectively model the data with up to 99% accuracy. Furthermore, under optimal parameter selection, the neural network model produces the highest possible accuracy under a comfortable level of deep learning while the support vector machine model employs greater speed and efficiency while still enjoying a respectably high level of accuracy.

Keywords: Cognitive workload · Machine learning · Deep neural networks · Support vector machines

1 Introduction

Mental workload continues to be a topic of great interest in human factors, particularly the means of assessing cognitive workload. Furthermore, the current state of machine learning serves as an expansive research realm to a variety of fields. This paper aims to bridge the two areas by identifying the advantages and disadvantages of present state-of-the-art machine learning approaches that are currently in the assessment of cognitive workload.

1.1 Cognitive Workload

Cognitive workload refers to the overall amount of effort being utilized in working memory. In a certain light, the concept can be regarded as a type of balancing act between underload and overload of such memory. For instance, many prior human factors research efforts have sought to address the problem of overload [1], such as when an individual is given tasks that are too numerous or too difficult, making one's

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_26

overall performance of such work less than optimal. Conversely, however, we must also address the possibility of underload. For instance, a traveler engaging in a long and monotonous drive may in fact benefit from increasing mental workload, as a low amount may find the driver in a "highway hypnosis," dulling his senses and leaving him unable to react effectively to sudden changes, such as the immediate slowdown of a leading car.

Overall, the ability to accurately and effectively assess one's cognitive state has remained a vital challenge in the field of neuroergonomics. Furthermore, the desire to predict, classify, and otherwise manipulate cognitive workload assessment has been greatly aided by the many novel facets of machine learning.

1.2 Machine Learning

Machine learning is a broad and well-recognized research field that spans across a wide variety of applications. Cognitive workload is no exception and has been addressed by this field in a vast plethora of ways. Artificial neural networks (ANNs) appear to be most established in a variety of such problems, including classification of mental workload [2], qualitative analysis based on air traffic data [3], and cross-task workload [4]. A subset of ANNs known as deep neural networks (DNNs) have been applied to similar but differing applications, such as emotion recognition [5], brain-computer interfaces [6], and analysis of human activity [7]. Support vector machines (SVMs) are well-known for assessing the cognitive state of a driver [8–11] and variations of mental workload [12]. Like with any machine learning problem, the key to forming an accurate and effective model is to properly decide the relevant inputs and outputs. Inputs in this field are typically physiological in nature and have included respiratory [2], electrocardiographic (ECG) [2], and electroencephalogram (EEG) [13] data, the latter of which seems to be the most common [2, 5, 13]. Electrodermal activity (EDA) also appears to be a promising physiological input in cognitive activity [14]. This paper aims to form a cost-benefit analysis of ANNs and SVMs in addressing cognitive workload, using EDA data as a key physiological input.

1.3 Overview of Subsequent Sections

The remainder of this paper is organized as follows. Section 2 provides a more thorough technical background of ANNs and SVMs. Section 3 provides an overview of how the research was conducted, while Sect. 4 presents the results and the discussion thereof. Afterwards, Sect. 5 presents concluding remarks.

2 Theory and Technical Background

This section establishes the technical backgrounds of the two machine learning techniques addressed in this paper. As such, ANNs and SVMs are organized as two distinct subsections. In addition, a further subsection presents the details of deep neural networks.

2.1 Artificial Neural Networks

An artificial Neural Networks (ANNs) is a statistical learning algorithm inspired by biological neural networks. A Multilayer Perceptron (MLP) is a neural network model that comprises of multiple layers of nodes [15]. MLP employs a supervised learning technique for training a neural network that consists of two steps. The first step is to select an appropriate architecture for the neural network problem, and the second step is to update the connection weights of the network. A neural network consists of individual units known as neurons, which take the form of three possible categories: input neurons, output neurons, and hidden neurons, which are interconnected and allocated at different layers. In a feedforward network, all the connections are forwarded from input neurons to hidden neurons, which are then routed to output neurons.

In recent years, many learning algorithms have been introduced that include direct optimization methods as well as global search techniques like Genetic Algorithm and Swarm Intelligence. The Genetic Algorithm (GA) is a global search technique based on biological evolution and genetics [16]. This technique searches from one population of solutions to another, propelling towards the eventual discovery of an optimal solution. Swarm Intelligence (SI) is an innovative bio-inspired computational tool that is inspired by animal or insect behavior. Among many Swarm Intelligence based techniques, Ant Colony Optimization (ACO) imitates the natural behavior of ants [17].

Deep Neural Networks. A deep neural network is one that consists of more than one hidden layer. In this type of feedforward network, the connections are still forwarded from the input neurons to the hidden neurons, but in this case, there is a first layer of hidden neurons that receives connections from the inputs, which then routes new connections to a secondary layer. These connections continue until reaching the last hidden layer, which in turn feeds a final set of connections to the output neurons. This phenomenon of deep learning is advantageous in the fact that we can apply a vast number of hidden neurons and therefore a greater number of variables for optimally tweaking a neural network model. In addition, an ANN can run much more efficiently with several layers of a few hidden neurons per layer than under a single layer of the same total number of hidden neurons.

2.2 Support Vector Machines

Support vector machines are a two-group classification problem that maps data points onto a high dimensional space [18]. In this technique, a support vector network maps multiple input factors into such a space through nonlinear mapping techniques [18]. The term support vector refers to the vectors that determine the largest margin of separation between two groups. In general, SVMs offer fast training in a distributed manner with efficient use of processing resources. Due to the relatedly binary nature of SVM outputs, this technique is more beneficial when a problem seeks qualitative outputs rather than quantitative ones. In addition to the training and validation accuracy metrics that are commonly sought under an ANN model, an additional metric, known as the F1 score, can be applied to such binary classifiers as SVMs [19].

3 Simulation Setup

To thoroughly address the costs and benefits of both techniques, we conducted a variety of simulation scenarios using the MATLAB scientific programming environment. An overview of the default simulation parameters in the experimental setup is as follows:

- Results presented as the averages of 30 independent trials for the sake of redundancy
- Samples of data randomized during each trial with 80% used for training and 20% used for validation
- ANN tested from 1 to 20 hidden layers of neurons and from 3 to 20 hidden neurons per layer
- ANN trained with Ant Colony Optimization containing α value of 0.1, 30 discrete points, 10 ants per iteration, an error maximum threshold of .001, and a maximum of 30 iterations
- SVM trained with a radial basis function kernel as well as with variable penalty terms and γ values, each consisting of 0.1, 1, 5, and 10.

The simulation was conducted on a Windows desktop with a quad core processor and 8 GB of RAM, with relevant conditions executed consistently between both machine learning mechanisms.

3.1 Data Acquisition

The data utilized in this research is based on an experiment by [14] in 2016. Natarajan et al. collected the data for the purposes of evaluating fear conditioning and cognitive logo. In turn, we have formatted the data so that it focuses entirely on the latter. To address cognitive workload, the data measures electrodermal activity (EDA) during a series of tests in which a human subject is required to either add 2, subject 7, or rest for 30-second periods. These mental tasks were chosen for the specific purpose of assessing cognitive workload, based on an individual's ability to remember a number and repeatedly, as well as interchangeably, add to or subtract from it. For this research project, the data was repurposed so that an ANN or SVM could be modeled to answer one of many possible questions:

- Given a task number and a time step, what will be a subject's EDA?
- Given a subject's EDA and a time step, which task was he/she performing?
- Given every subject's EDA and a time step, what (common) task were they performing?

To demonstrate each algorithm's capacity to accurately model the maximum number of inputs feasible that this data set can form, the data in this research is formatted to focus on addressing the third question. Under this arrangement, the EDA for 24 separate subjects form 24 inputs, and their common time step serves as an additional input for a total of 25. The sole output is the activity number, which represents the task being performed. Although there are only three unique activities, each nonconsecutive instance of a recurring task is its own separate activity number.

3.2 Algorithmic Implementation

The algorithm was programmed by following the flowcharts provided in Fig. 1. The SVM algorithm was conducted from built-in MATLAB functions, while the ANN algorithm was coded manually. Thus, only the ANN training algorithm is covered in detail, as indicated by the flowchart on the right. The left diagram serves, on the other hand, as a general training and testing process for both techniques.

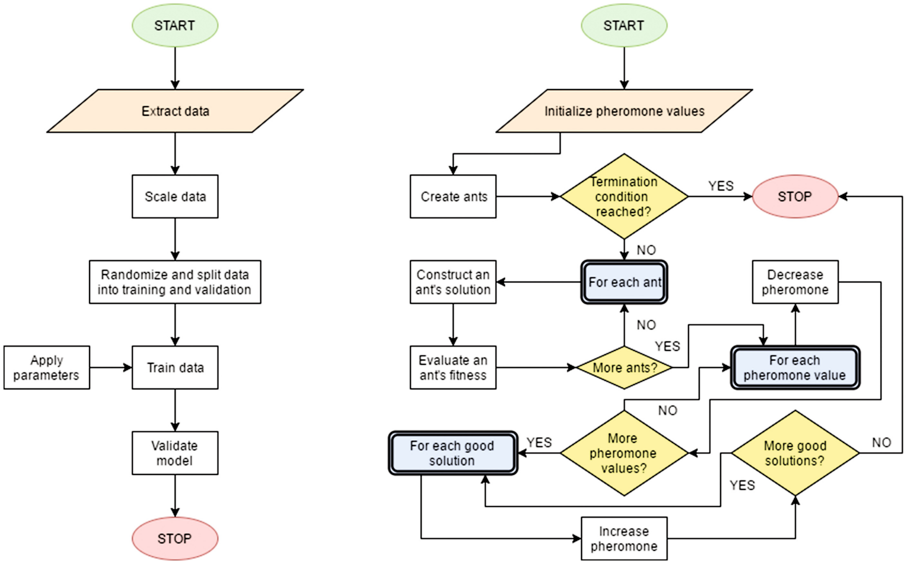


Fig. 1. Flowcharts of general machine learning process and of ANN training by ant colony optimization.

4 Results and Discussion

This section provides the results for the MATLAB simulation, in the forms of both accuracy and computational runtime. In addition, a discussion of the pros and cons of both ANNs and SVMs is given. Also discoursed are the key distinctions and observations between deep neural networks and more traditional, single layer ANNs.

4.1 Neural Network Results

Figures 2 and 3 provide surface plots of the training and validation accuracies of the ANN simulation, respectively. As the training accuracy indicates, the data is better modeled under a DNN than under a conventional ANN. Validation accuracy, on the other hand, provides less conclusive results between hidden neurons and accuracy.

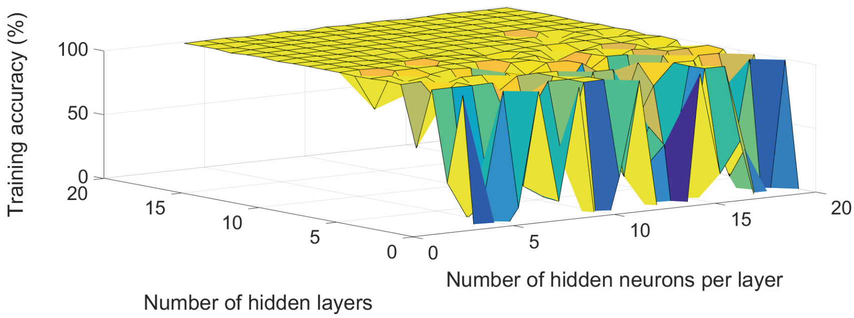


Fig. 2. Surface plot of training accuracy by number of hidden layers and by number of hidden neurons per layer.

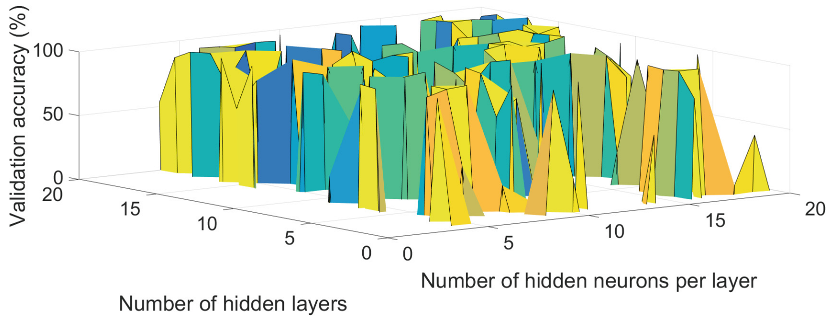


Fig. 3. Surface plot of validation accuracy by number of hidden layers and by number of hidden neurons per layer.

To remedy this, the experiment was then refined so that only the optimal number of neurons per hidden layer is considered for each number of layers, as given in Fig. 4. In addition, Fig. 5 provides the efficiency of the ANN model under the same two variables.

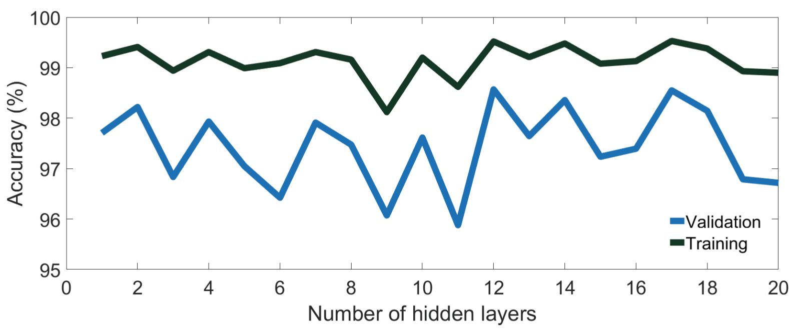


Fig. 4. Optimal accuracy by number of hidden layers.

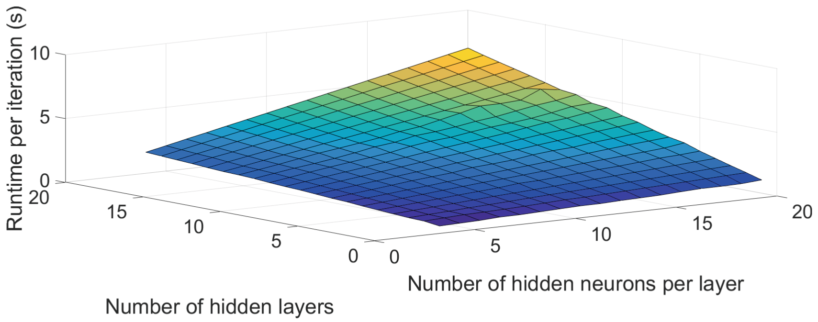


Fig. 5. Surface plot of computational runtime per iteration by number of hidden layers and by number of hidden neurons per layer.

When factoring the efficiency presented in the latter figure with the accuracy given in the former, we can declare that the ANN model works optimally under the range of 11–18 layers, thereby proving that deep neural networks are overall better suited for this realm of cognitive workload when compared to traditional single layer ANNs.

4.2 Support Vector Machines Results

Figure 6 provides a comprehensive analysis of the SVM model. The top row of figures provides the respective training and validation accuracies by γ value and by penalty term. The bottom row provides the F1 score and the total computational runtime in terms of the same two variables.

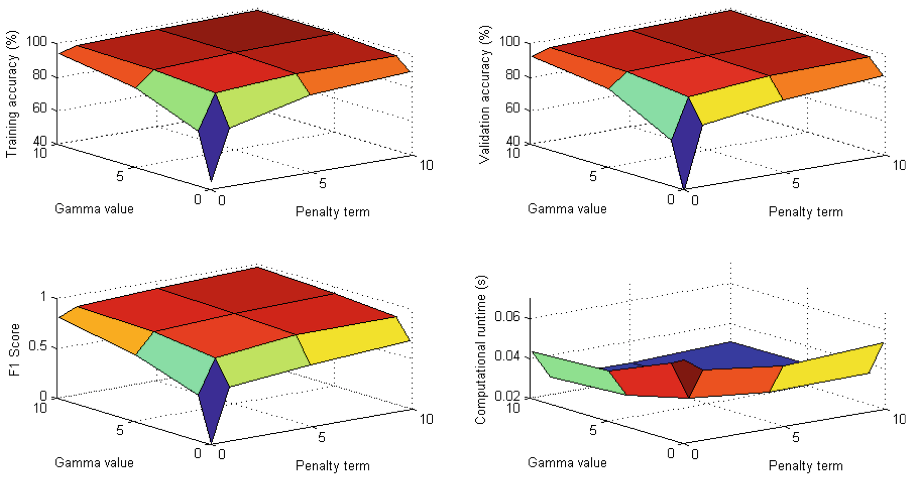


Fig. 6. Surface plots of (top left) training accuracy, (top right) validation accuracy, (bottom left) F1 score, and (bottom right) computational runtime in terms of gamma value and penalty value.

As the first three plots indicate, accuracy follows a consistent and somewhat predictable pattern, as an increase in each of the variables results in a higher accuracy and a higher F1 score. The bottom right plot follows an inverted but consistent pattern, in that an increase in either variable results in a shorter runtime. To maintain consistency with presentation of the previous technique's results, Fig. 7 depicts the accuracy results for the optimal penalty term in each gamma value.

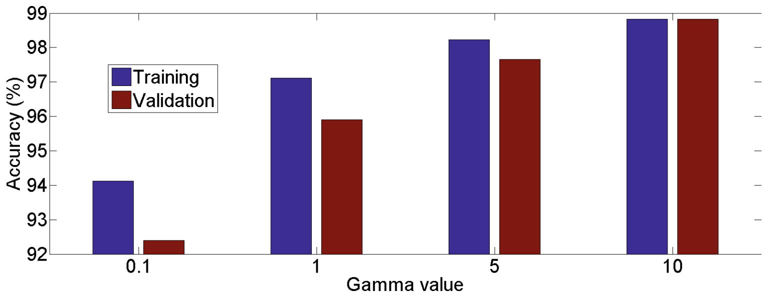


Fig. 7. Optimal accuracy by gamma value.

As the results indicate, both forms of accuracy remain respectably high under the optimal penalty value, with the training accuracy in the range of 94–99% and validation accuracy spanning 92–99%.

4.3 Discussion

The most apparent observation when comparing the two sets of results is the stark difference in runtime. The SVM in its entirety runs significantly faster than that of even a single ANN iteration. A likely reason for this lies within the added complexity of incorporating a global method of training (ACO in this case), as opposed to a quicker gradient based method.

In terms of accuracy, however, the ANN model maintains a slight edge. Under optimal amounts of hidden neurons per layer, training accuracy remains consistently above 98% while validation accuracy exceeds 97% in nearly two thirds of the hidden layer choices. The SVM model, on the other hand, only reaches these accuracy thresholds for half of the tested gamma values. Thus, we can conclude that in this particular application of cognitive workload assessment, SVMs are the more ideal machine learning mechanism for speed and efficiency. In the instance of absolute need for high accuracy, however, ANNs produce a noticeably higher level of accuracy, particularly under certain conditions of deep learning.

5 Conclusions and Future Work

This paper presented a detailed analysis of the use of artificial neural networks and support vector machines for assessing cognitive workload. These machine learning mechanisms were tested on a set of physiological data aimed to relate the electrodermal activity of multiple human subjects to classification of a cognitive task. The former technique was evaluated in terms of a traditional single layer ANN and in terms of a deep neural network. The results ultimately indicate that the SVM model has a noticeable edge in computational runtime, while the ANN model produces a consistently higher accuracy.

Because this paper focused only on two machine learning techniques, the most prominent direction for future research lies within the evaluation of additional machine learning techniques, such as deep belief networks, neuro-fuzzy systems, and regression trees. In addition, ANNs and SVMs can be expanded for evaluation in terms of more parameters, more training methods, and in the case of the latter, more kernel functions. Exploration of additional cognitive workload data sets also remain a relevant path for future work.

Acknowledgments. This research is supported by the Dayton Area Graduate Studies Institute (DAGSI) fellowship program for a project titled "Assessment of Team Dynamics Using Adaptive Modeling of Biometric Data." The authors wish to thank their DAGSI sponsor Dr. Gregory Funke and his colleagues Dr. Scott Galster and Mr. Justin Estepp for their continued guidance and support throughout the project. The authors also thank the EECS Department at the University of Toledo for partial support through assistantships and tuition waivers.

References

1. Berka, C., et al.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat. Space Environ. Med. **78**(5), B231–B234 (2007)
2. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Hum. Factors **45**(4), 635–644 (2016)
3. Chatterji, G.B., Sridhar, B.: Neural network based air traffic controller workload prediction. In: Proceedings of the 1999 American Control Conference, pp. 2620–2624 (1999)
4. Baldwin, C.L., Penaranda, B.N.: Adaptive training using an artificial neural network and eeg metrics for within- and cross-task workload classification. NeuroImage **59**(1), 48–56 (2012)
5. Zheng, W., Lu, B.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Auton. Ment. Dev. **7**(3), 162–175 (2015)
6. Hennrich, J., Herff, C., Heger, D., Schultz, T.: Investigating deep learning for fNIRS based BCI. In: 37th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), pp. 2844–2847, New York. IEEE Press (2015)
7. Sarkar, S., Reddy, K., Dorgan, A., Fidopiastis, C., Giering, M.: Wearable EEG-based activity recognition in PHM-related service environment via deep learning. Int. J. Progn. Health Manag. **7**, 1–10 (2016)
8. Jin, L. et al.: Driver cognitive distraction detection using driving performance measures. Discret. Dyn. Nat. Soc. **2012**, 1–12 (2012)
9. Son, J., Oh, H., Park, M.: Identification of driver cognitive workload using support vector machines with driving performance, physiology and eye movement in a driving simulator. Int. J. Precis. Eng. Manuf. **14**(8), 1321–1327 (2013)

10. Putze, F., Jarvis, J., Schultz, T.: Multimodal Recognition of cognitive workload for multitasking in the car. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 3748–3751, New York. IEEE Press (2010)
11. Liang, Y., Reyes, M., Lee, J.: Real-time detection of driver cognitive distraction using support vector machines. IEEE Trans. Intell. Transp. Syst. **8**(2), 340–350 (2007)
12. Yin, Z., Zhang, J.: Identification of temporal variations in mental workload using locally-linear-embedding-based eeg feature reduction and support-vector-machine-based clustering and classification techniques. Comput. Methods Programs Biomed. **115**(3), 119–134 (2014)
13. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from EEG with deep recurrent-convolutional neural networks. In: 4th International Conference on Learning Representations (2016)
14. Natarajan, A., Xu, K.S., Eriksson, B.: Detecting divisions of the autonomic nervous system using wearables. In: IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), pp. 5761–5764, New York. IEEE Press (2016)
15. Malsburg, C.: Frank Rosenblatt: principles of neurodynamics: perceptrons and the theory of brain mechanisms. In: Palm, G., Aertsen, A. (eds.) Brain Theory, Proceedings of the First Trimeste Meeting on Brain Theory, pp. 245–248. Springer, Berlin (1986)
16. Morris, G.M., et al.: Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. **19**(14), 1639–1662 (1998)
17. Socha, K., Blum, C.: An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training. Neural Comput. Appl. **16**(3), 235–247 (2007)
18. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
19. Rijsbergen, C.J.: Information Retrieval, 2nd edn. Butterworth, London (1979)

Relationship Between EEG and ECG Findings at Rest and During Brain Activity

Yuto Nakahata[1(✉)] and Hiroshi Hagiwara[2]

[1] Graduate School of Information Science and Engineering, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan
is0178xs@ed.ritsumei.ac.jp
[2] College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan

Abstract. The development of information technology has enabled increasingly easy access to biological information. Humans need to rest in accordance with individual biological status. This study aimed to determine relationships between findings from electroencephalography (EEG) and electrocardiography (ECG) as physiological indices at rest and during brain activity. Correlations between seated rest and seated rest on a massage chair were low, indicating that the power of alpha brain waves decreased as R-R interval (RRI) increased. However, this study did not show correlations between seated rest + the auditory 2-back task and seated rest on the massage chair. RRI and the power of alpha waves during tasks changed depending on whether the participant was relaxed or engaged in brain activity, with an inverse correlation between the two factors. The relationship between EEG and ECG findings at rest and during brain activity reflects the relationship between the central and autonomic nervous systems at rest.

Keywords: Autonomic nervous system · Central nervous system · Auditory 2-back task · Massage chair

1 Introduction

It is becoming possible to acquire, store, and utilize biological information on the individual level with the development of information technology applications such as the Internet of Things (IoT) and big data. In many developed countries, including Japan, stress factors such as long work hours, sleep deprivation, and interpersonal relationships are present in daily life. Accumulation of stress is associated with depression and lifestyle-related diseases, and represents an obstacle to living a healthy everyday life. To prevent such outcomes, the ability to accurately gauge the status of the individual by measuring biometric information and to provide appropriate countermeasures is important. In this research, we focused on grasping the condition of the individual.

Findings on electroencephalography (EEG) reflect the status of the central nervous system and are used to estimate depth of sleep and workload [1] of the brain. On the other hand, findings on electrocardiography (ECG) reflect the status of the autonomic

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_27

nervous system and are used to evaluate stress [2] and relaxation. Since the electrical signal on EEG is very small (around 1–100 μV), the results tend to be very noisy. Measurement on a daily basis is thus difficult. On the other hand, the electrical signal from ECG is comparatively easy to measure on a daily basis, since the electrical signal (100 μV to 10 mV) is larger than that from EEG. As a result, a large amount of research has examined methods for measuring ECG in daily life. In addition, Airitmo (Daikin Industries, Osaka, Japan) [3] and Hitoe (Toray Industries, Tokyo, Japan) have developed products that can measure heartbeat in daily life. Quality of life is expected to be improved by measuring biological information during daily life and offering feedback to users.

Based on this research background, the purpose of the present study was to elucidate the relationships between EEG, ECG, and subjective evaluation during brain activity and rest. By elucidating these relationships, we believe that information on not only the autonomic nervous system but also the central nervous system and subjective interpretations will be obtained from EEG information.

2 Experimental Method

2.1 Subjects

Participants comprised 12 healthy, non-medicated college students (9 men, 3 women; age range, 18–23 years). All subjects provided written informed consent prior to participation. Subjects refrained from excessive eating and drinking the night before the experiment. In addition, subjects refrained from engaging in prolonged or strenuous exercise on the morning of the experiment.

2.2 Experimental Protocol

The experimental protocol is shown in Fig. 1. The experimental protocol was performed using alpha attenuation test (AAT)-specific activity (eyes closed [30 s], eyes open [30 s] × 3) [4] for 3 min, followed by a 10-min task, then the 3-min AAT again. The time of one cycle was 16 min. The Roken Arousal Scale (RAS) was evaluated before and after measurements, with 5-min breaks between each cycle. The order in which tasks were performed was randomized in every experiment to exclude any effect of order.

2.3 Tasks

The four tasks in the experiment comprised seated rest, seated rest + auditory 2-back task, seated rest on a massage chair, and seated rest on a massage chair + auditory 2-back task. In the state of seated rest, the subject sat on a chair with eyes closed. In the state of seated rest on the massage chair, the stimulation position for massage was set to the shoulder. The subject sat on the massage chair with eyes closed.

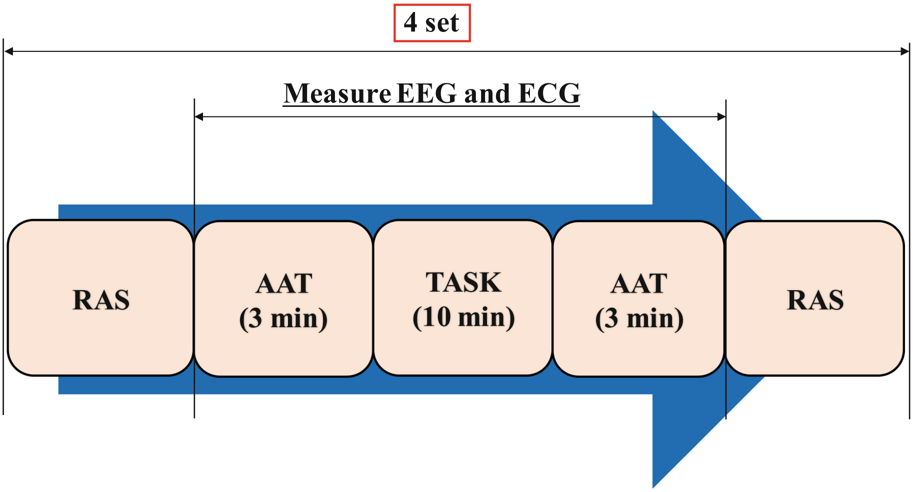


Fig. 1. Schematic of the experimental design. Tasks were performed continuously and sequentially under the following four conditions: (1) rest; (2) rest + auditory 2-back task; (3) massage; and (4) massage + auditory 2-back task.

The auditory 2-back task consisted of the participant listening for an audible random number between 0 and 3, delivered at 3 s intervals. Subjects were required to press a button when the presented number was identical to the number presented two numbers previously (Fig. 2). This auditory 2-back task is a representative task used to evaluate working memory.

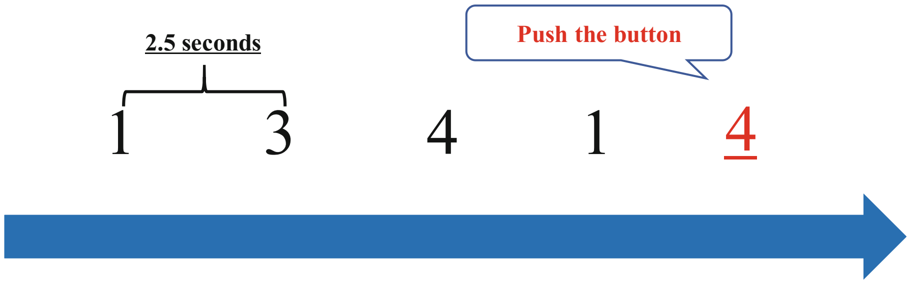


Fig. 2. Schematic of the auditory 2-back test. The subject was required to press a button when the presented number was identical to the number presented two numbers previously.

2.4 Experimental Equipment and Electrode Fixation Points

An EEG1100 (Nihon Kohden, Tokyo, Japan) was used for EEG and ECG. EEG was recorded using the International 10–20 system (Fig. 3a). Electrodes were attached to the head, and measurements were obtained from electrodes C3, C4, O1, O2, Fz, Pz, A1, and A2. Cutaneous sebum was removed using a polisher before electrode application to reduce electrical resistance. ECG was recorded using the three-point lead system (Fig. 3b). The difference between signals recorded from the electrodes over the right clavicle and left ribs was measured. Respiratory rate was measured using the respiratory sensor wrapped around the thoracoabdominal region.

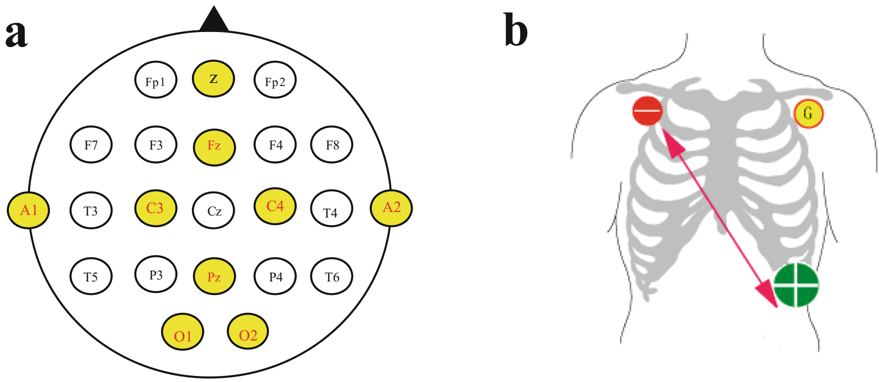


Fig. 3. (a) International 10–20 system for measurement of EEG. (b) Three-point lead system for measurement of ECG.

3 Analysis

3.1 Roken Arousal Scale (RAS)

The RAS is a psychological evaluation method that quantitatively rates psychological values of fatigue and alertness [5]. The RAS provides a quantitative index of the following six states: sleepiness, activation, relaxation, strain (tension), difficulty with attention and concentration, and lack of motivation. Each of the six states is evaluated using two similar questions, with the average of the two similar questions defined as the state value.

3.2 ECG

In this experiment, RR interval (RRI) was used as ECG data. The purpose of the research was to obtain more information from EEG data. For this reason, we used RRI, which would be the easiest data to obtain, without analyzing ECG data such as heart rate variability (HRV) or high frequency (HF).

3.3 EEG

Brain wave noise was removed using low-pass filtering (120 Hz), high-pass filtering (0.1 Hz), and band-stop filtering (57–60 Hz). Frequency analysis was then performed using a fast fourier transform (FFT). EEG signals obtained from O2-A1 were analyzed, since alpha waves appear in EEGs recorded from the parietal and occipital regions under conditions of wakefulness, rest, and with closed eyes. In this study, we defined the power spectrum at frequencies from 8 Hz to 13 Hz as the power of alpha wave.

4 Results

4.1 Auditory 2-Back Task

Figure 4a shows the average reaction time in the auditory 2-back task for the 12 subjects. Comparing differences between rest and massage, the average correct answer rate was greater at seated rest than with seated rest on the massage chair. Standard deviation of the correct answer rate in seated rest was smaller than that in seated rest on the massage chair.

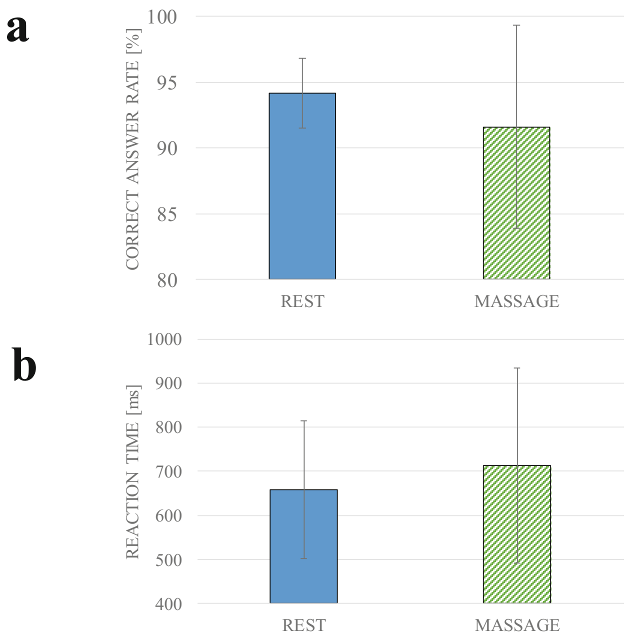


Fig. 4. (a) Average correct response rate for the auditory 2-back task in the 12 subjects. (b) Average reaction time in the auditory 2-back task for the 12 subjects.

Figure 4b shows the average correct answer rate in the auditory 2-back task for the 12 subjects. Comparing differences between rest and massage, the average correct answer rate in seated rest was lower than that of seated rest on the massage chair. Standard deviation of the correct answer rate in seated rest was higher than that of seated rest on the massage chair.

4.2 RAS

Figure 5 shows average values for sleepiness and relaxation calculated from the RAS score. Sleepiness value increased in all four tasks. A significant increase ($p < 0.05$) was seen before and after the task in seated rest. Relaxation value also increased in all four tasks. A significant increase ($p < 0.05$) was seen between before and after the task in seated rest and seated rest + the auditory 2-back task.

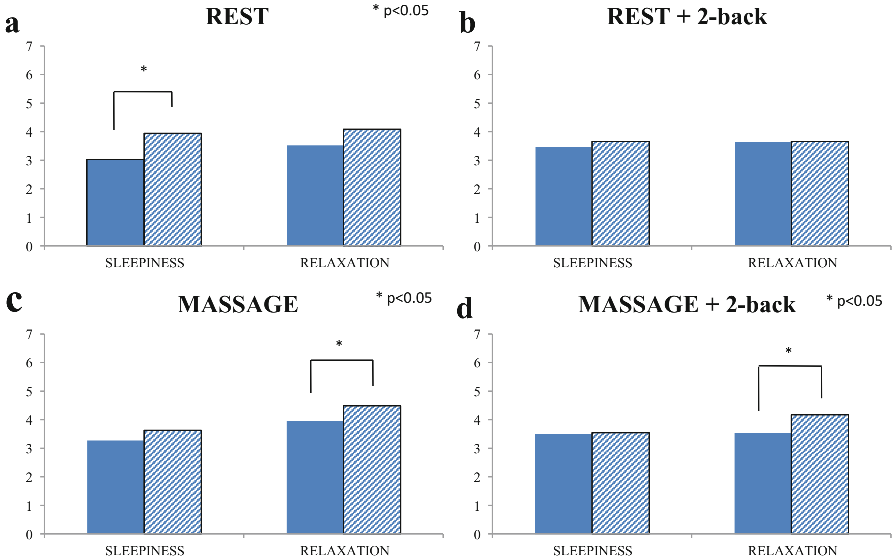


Fig. 5. Bars show differences between RAS scores before and after the task in: (a) seated rest; (b) seated rest + auditory 2-back task; (c) seated on massage chair; and (d) seated on massage chair + auditory 2-back task.

4.3 EEG

RRI change during 10 min is shown in Fig. 6. Each point represents the average value of alpha waves every 2 min. Circles represent “rest” and triangles represent “massage”. Solid lines represent “absence of the auditory 2-back task”, and dotted lines represent “presence of the auditory 2-back task”.

The power of alpha waves in all tasks showed a decreasing trend over time. No significant differences were seen between the auditory 2-back task and absence of the auditory 2-back task in both rest and massage at 0 to 2 min and 8 to 10 min. Conversely, average power of the alpha wave in absence of the auditory 2-back task was lower than in the auditory 2-back task.

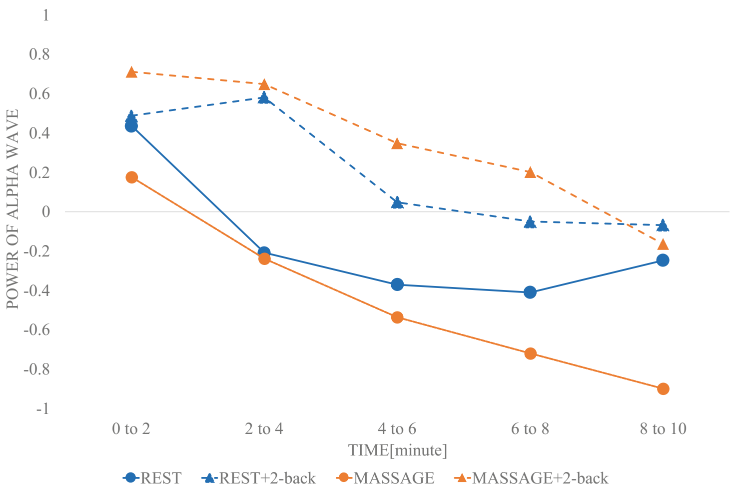


Fig. 6. Average power of the alpha wave in all four tasks for the 12 subjects.

4.4 ECG

Results of the RRI change during 10 min are shown in Fig. 7. Each point is the average value of the RRI every 2 min. Circles represent "rest" and triangles represent "massage". Solid lines represent "absence of the auditory 2-back task" and dotted lines represent "presence of the auditory 2-back task".

The RRI in seated rest showed an increasing trend and RRI in seated rest + auditory 2-back task was roughly constant over time. RRI in seated rest on the massage chair showed a slight increasing trend and RRI in seated rest on the massage chair +

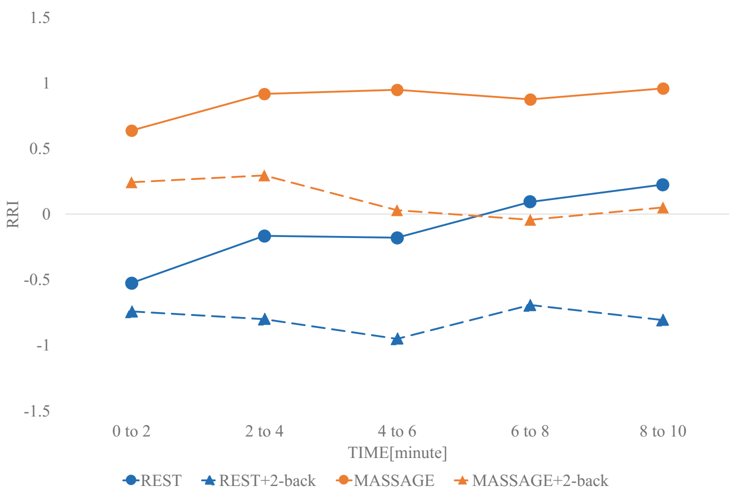


Fig. 7. Average RRI in all four tasks for the 12 subjects.

the auditory 2-back task displayed a decreasing trend over time. Next, we described differences depending on the presence or absence of tasks. No significant differences were evident between the auditory 2-back task and absence of the auditory 2-back task in both rest and massage at 0 to 4 min. Conversely, average RRI in the absence of the auditory 2-back task was higher than in the auditory 2-back task.

4.5 Relationship Between EEG and ECG

Figure 8a shows a graph of seated rest with the power of the alpha wave on the vertical axis and RRI on the horizontal axis. Figure 8b shows a graph of seated rest + an auditory 2-back task. In these graphs, the average of each value for every 30 s for the 600 s of the task was plotted and connected by a line. In addition, circles indicate the start point (from 0 s to 30 s in each task), and triangles indicate the end point (from 570 s to 600 s in each task). RRI increases as the power of the alpha wave decreases in seated rest. A similar tendency was seen in seated rest on the massage chair. Conversely, no tendency was seen for RRI to increase as the power of alpha waves decreased in both seated rest + auditory 2-back task and seated rest on the massage chair + auditory 2-back task.

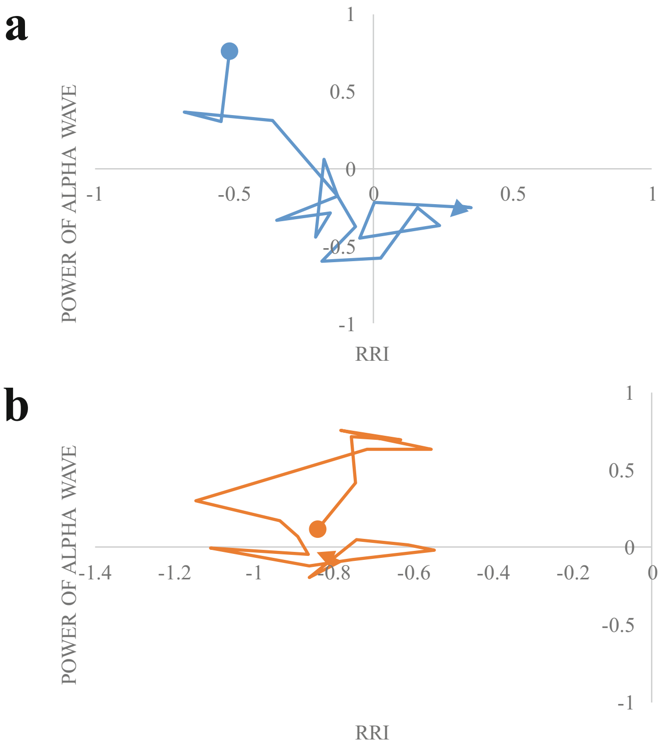


Fig. 8. Relationship between power of the alpha wave and RRI. (a) Seated rest; (b) seated rest + auditory 2-back task.

We found a correlation between power of the alpha wave and RRI in 10 min of task performance for each subject. Ten of the 12 subjects showed a negative correlation and 7 of those 10 individuals showed a strong correlation (5% significance level) in seated rest. Nine of the 11 subjects showed a negative correlation and 4 of those 9 people showed a strong correlation (5% significance level) in seated rest. No constant trend was evident, and different correlation coefficients were shown by subjects in both seated rest + auditory 2-back task and seated rest on the massage chair + auditory 2-back task.

5 Discussion

From 2 min to 8 min, significant differences in the power of the alpha wave were seen depending on the presence or absence of the auditory 2-back task. The subject became sleepy towards the end of the task, and the power of the alpha wave could shift to the power of a theta wave. Moreover, brain activity is considered to be promoted and attenuation of the power of alpha waves was suppressed by the subject performing the auditory 2-back task. Likewise, regarding RRI, significant differences were also seen depending on the presence or absence of the auditory 2-back task in both seated rest and seated rest on the massage chair, except task 0 min to 2 min. As time passed, a tendency was seen for the difference between the RRI for the absence of the auditory 2-back task and for the increased presence of the auditory 2-back task.

From the correlation coefficient between the power of the alpha wave and RRI, a relationship was seen whereby RRI increases as the power of alpha wave decreases in seated rest with eyes closed. From this, the state of brain activity was suggested as able to be estimated from the heart beat interval. Also, no relationship in the auditory 2-back task was identified. This is probably because the motivation and degree of concentration of each subject differed due to the large error of performance result from the auditory 2-back task. We plan to focus on the following two issues in the future: (1) to analyze not only the power of alpha waves but also the power of theta waves and so on, and to analyze the relationship between EEG and ECG; and (2) to classify subjects according to the results of the auditory 2-back task and see the relationship for each class.

6 Conclusion

In this study, we aimed to increase information obtained from ECG by examining the relationship between ECG, EEG, and subjective evaluation. Since the correlation between the ECG and EEG was observed in seated rest and seated rest on a massage chair, ECG was suggested as potentially able to be estimated from the electrocardiogram. Under this thesis, we examined the relationship between the ECG and EEG, but we hope to investigate the relationship between ECG and subjective evaluations in future research. We hope that this research will prove useful for biometric information-sensing technology.

Acknowledgements. This study was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT)-Supported Program for the Strategic Research Foundation at Private Universities, 2013–2017.

References

1. Holm, A., Lukander, K., Korpela, J., Sallinen, M., Müller, K.M.: Estimating brain load from the EEG. Sci. World J. **9**, 639–651 (2009)
2. Vrijkotte, T.G.M., van Doornen, L.J.P., de Geus, E.J.C.: Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. Hypertension **35**(4), 880–886 (2000)
3. Daikin Industries, Ltd.: "Airitmo". http://www.airitmo.jp/
4. Hagiwara, H., Araki, K., Michimori, A., Saito, M.: A Study on the quantitative evaluation method of human alertness and its application. Psychiatr. et Neurol. Jpn. **99**(1), 23–34 (1997)
5. Takahashi, M., Kitagima, H., Honjo, Y.: Analysis of the relationship between sleepiness and relaxation using a newly developed mental work strain checklist. J. Sci. Labour **72**(3), 89–100 (1996)

Empathy in Design: A Historical and Cross-Disciplinary Perspective

Yumei Dong, Hua Dong(✉), and Shu Yuan

College of Design and Innovation, Tongji University,
1239 Siping Road, Shanghai 200092, China
donghua@tongji.edu.cn

Abstract. From a historical and cross-disciplined perspective, this paper reviewed the evolvement of empathy in philosophical aesthetics, sociology and psychotherapy. Three aspects of ambiguities about empathy were summarized, namely, affection and cognition, subject-oriented and object-oriented perspective, and attitude and ability, based on which a system of coordinates was created to indicate three pairs corresponding problems of empathy in design research and practice. Through this system of coordinates, the authors wanted to build a reference for designers and researchers to deeper understand the subtle nature of empathy in design.

Keywords: Empathy · Ambiguities · Affection and cognition · Human-centered design

1 Introduction

With the growing complexity of design, understanding users and their experience has moved to the central place, and Human-Centered Design (HCD) has become a powerful and popular notion and concept in both design academic and industrial circles. Building on a long history of HCD, empathic design has attracted an increasing attention [1].

The adjective "empathic" was first introduced to the design field in late-1990s [2] when companies started to realize that only listening to customers' responses on questionnaires was not enough to develop successful products [3, 4]. In the following two decades, empathy was frequently emphasized and discussed on public occasions in both academic and industrial circles. However, it seemed that empathy just acted as an umbrella term for designers and design researchers who tried to show their "Human-Centered" mindset, while the deep understanding of the nature of empathy was missing. It made the discussion about empathy in design filed vague and obscure.

In order to clarify the meaning of empathy in design, we reviewed the papers about the development of the term "empathy" in the disciplines of philosophical aesthetics, psychology and sociology. Based on a comparison of empathy in history, we found that it experienced a controversial process on defining empathy, while there were also inconsistency through a cross-disciplinary analysis.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_28

2 A Historical View of Empathy

2.1 From Late 19c to Early 20c

Robert Vischer, a German art historian and philosopher, first discussed empathy in philosophical aesthetics in 1873. Vischer named a German term, *Einfühlung,* to express observers' feelings elicited by works of art [5]. It was used to describe the projection of human feelings onto the natural world and inanimate objects [6]. In 1903, Theodore Lipps used *Einfühlung* to refer to the tendency of perceivers projecting themselves into the perceived objects [7]. He conceived of empathy as a psychological resonant phenomenon, in which people projected their emotion to external objects. The experience of beauty was characterized as 'objectified self-enjoyment' [8], which can be considered as a kind of animism [9]. In 1909, Edward Bradner Titchener coined the term empathy as a rendering of *Einfühlung,* which meant a 'process of humanizing objects, of reading or feeling ourselves into them [10]'.

From these early explanations of empathy in aesthetics, it was found that empathic experience was focused more on the reflection from the person who was interacting with the object rather than the object itself, although Lipps tended to regard it as an automatically emotional reaction while Titchener thought it is a cognitive process based on imagination and reconstruction. These slight differences of the explanations about empathy could be regarded as an initial divergence on empathy component between emotion and cognition.

2.2 From the Early 20c to the Middle 20c

As early as 1897, Lipps brought empathy from aesthetics to psychology. In 1905, Sigmund Freud started to use 'empathy' to describe the psychodynamics of putting oneself in another person's position [6]. Thus, the empathizing object turned into emotional human from natural objects and non-human.

Another change is the emphasis on cognitive component of empathy. Köhler thought empathy was more about the understanding of others' feelings than sharing of them [11]. Piaget, a child developmental psychologist, also emphasized the cognitive over the emotional part [12]. He addressed empathy as 'a cognitive function and to the ideas of what is required of an individual in order to decenter and imagine the role of another' [13].

Besides, Mead (1934), a social psychologist, recognised the potential of role-taking as the key to social and ethical development of children [14]. Kohler, Piaget and Mead, all placed a huge emphasis on the individual's capacity to take on the role of other persons to understand how they viewed the world, which manifested the importance of cognitive part in empathy.

After that, more and more researchers paid attention to empathy as a capacity in social interaction. For example, Heinz Kohut, a psychoanalyst, was best known for his self-psychology, in which empathy was an essential component. He viewed empathy as 'a capacity to think and feel oneself into the inner life of another person', and also a kind of 'vicarious introspection', by which the therapists can use their own experiences to understand the clients in similar circumstances [15]. Rogers, a coeval humanist psychologist, also emphasized empathy as a core component in psychotherapy [16]. Rogers considered it was important to understand how clients viewed the world rather than the actual circumstances. By his "Client-Centered Therapy", Rogers believed that people were fundamentally good, thus he advocated an equal relationship between therapists and clients. This is the early concern about the relationship between the subject and the object in an empathic process.

2.3 From 1970s to Present

Being affected by Rogers and Kohut, empathy gained popularity in the field of psychotherapy. Beside the discussion on the nature of empathy, empathy scales aroused a big interest of research.

From 1970's to present, there was a radical divergence on the discussion about the components of empathy. Stotland, Hoffman and Baston defined empathy from an affective perspective [17–19]. Hoffman saw empathy as an affective response more appropriate to another's situation than one's own, while Baston defined empathy not only referring to others-oriented feelings but also to a vicarious emotion such as compassion, warmth, and concern.

However, Wispé and Ickes preferred to focus on a cognitive component of empathy [20, 21]. Ickes coined the term 'empathic accuracy' to refer to one's capability to accurately infer the specific content of another person's thoughts and feelings.

There were also researchers who viewed empathy as a multidimensional phenomenon. One of them was Davis, who thought empathy contained two different parts, the cognitive role taking and the affective reactivity to others. Davis developed the Interpersonal Reactivity Index (IRI) to test the individual level of empathy [12], which contained four separate but related self-report subscales, namely perspective taking, fantasy, empathic concern, and personal distress.

3 Ambiguous Aspects in the Evolvement of Empathy

From the origin to present, the concept of empathy has evolved and transformed on many aspects. The development of modern neuroscience offered many evidence for understanding empathy. However, the nature of empathy has yet to reach an agreement among researchers. Based on the literature review, we identified 3 aspects of unclear aspects of empathy.

3.1 Affection vs. Cognition

From 1920s to 1950s, researchers tended to discuss empathy from a cognitive view. After about 20 years of silence, researchers from experimental psychology began to pay attention to the affective component of empathy. From 1970s till now, it shows a divergence among affective advocators, cognitive advocators and multi-dimensional advocators.

Some of them considered empathy as an automatic affective response while others preferred to define it as an advanced and active cognitive processing such as role-taking. Because of different views to the construction of empathy, the measurements of empathy also concerned different aspects. For example, Hogen Empathy Scale (HES) (Hogen 1969) focused on the cognitive structure while Questionnaire Measure of Emotional Empathy (QMEE) [23] measured empathy from an affective perspective. Even when empathy was translated into Chinese, there were also two terms ‘Gongqing’ and ‘Tongli’, which related to affective and cognitive parts respectively.[1]

With affective and cognitive views, neuroscience studies suggest that empathy is a psychological construct regulated by both cognitive and affective components. There are two possible systems for empathy: a basic emotional contagion system and a more advanced cognitive perspective-taking system [24].

Researchers also argued that empathy contains three components. Glastein considered behavior as the third component, apart from the affective and cognitive ones [25], while Rogers regarded communication as the third part [16]. Motor was also considered as the third part in recent research [26]. However, the affection and the cognition are still the primary concerns.

3.2 Subject-Oriented vs. Object-Oriented

Empathy happens between two roles. In philosophical aesthetics, empathy occurs between aesthete and works of art. While in psychotherapy, it occurs between therapists and clients. Therefore, empathy refers to the relationship between two main bodies, the subject and the object. Then we have to discuss who is in the center of this relationship and whose experience has the priority. Theodor Lipps defined empathy as a process of ‘feeling ourselves into them (“them” means aesthetics object here)’. In this context, the empathetic experience derived from the aesthete, i.e., the empathizer. It is the observer’s subjective initiative and experience, rather the object’s, that is dominant in formulating an empathetic experience.

However, in Client-centered therapy, Roger advocated an unconditional positive concern to clients without judgment. The empathizing subject just played a role as a mirror of the object, which showed a completely respect to the empathized object and encouraged the object itself to develop a healthier view of the world. In this context, the object’s experience was in the central place in this empathic relationship.

Although researchers in sociology emphasized more on objects’ experiences, the practitioners inevitably projected themselves in the process of understanding others. Some scholars empathized the importance of self-reflection in empathetic understanding [27]. They argued that trying to connect one’s own experiences similar to the empathic object could be beneficial for empathic arousing. Experiments on pain also offer the evidence for the effect of self-reflection on empathy.

The relationship between the subject and the object in the empathic process also involves the direction of emotion, self-oriented and other-oriented. Davis considered

[1] The Chinese term ‘Gongqing’ means sharing emotion with others, while ‘Tongli’ refers to understanding the same.

empathy as a multidimensional phenomenon, which contained both other-oriented feelings, such as sympathy, compassion, warmth and concerns for unfortunate others and self-oriented feelings of personal anxiety and unease in tense interpersonal settings [28].

3.3 Attitude vs. Ability

Greenson (1960) defined empathy as the capacity 'to share, to experience the feelings of another person' [29]. McDonagh also received empathy as the intuitive ability to identify other people's thoughts and feelings [30]. The explanation about empathy as a kind of capacity was highly recognized among researchers. Under this viewpoint, researchers thought people could improve the ability of empathy with proper trainings. In many applied fields such as psychotherapy, education and nursing, more and more researchers concentrated on techniques for interventions of empathy training. In many controlled experiments, it is also proved that empathy training group showed much more improvement in the ability of empathetic understanding such as emotional facial expressions decoding [31]. This suggests empathy can be improved by certain techniques of training. In a more applied perspective, Kohler also thinks empathy was a 'means' and 'a scientific method' for psychoanalysis to collect data [32].

However, the dogmatic training on empathic techniques might also bring trouble, especially in psychotherapy. Rogers disagreed to consider empathy as a specific technique or a way of response, but a part of the whole attitude [33]. He insisted that overemphasizing on capacity or techniques might result in an expert superiority of therapists, which would make patients uncomfortable. This problem could be found in many applied fields.

4 The Ambiguous Understandings About Empathy in Design

After almost 100 years of research on empathy in philosophy, sociology and psychology, it finally attracted the attention of design researchers. To some extent, the controversial nature of empathy and the ambiguities in its evolvement impacted on researchers' understanding of design. In current discussion on empathy in design and empathetic design practice, problems and ambiguities often comes from the above three aspects as shown in the 3-dimensional system of coordinates (Fig. 1). When utilizing the concept of empathy in design or research activities, designers and researchers need to consider which stance you are on at all these three aspects.

4.1 Affection or Cognition

Empathy was first introduced to design when traditional inquiry methods were far from users' feelings and moods and few insights obtained from the market survey [3]. Under this background, researchers focused more on the affective part of empathy when discussing about empathy in the design field.

The affective empathy included positive and negative emotions. When utilizing empathy to explain the interaction between two persons, most researchers tended to

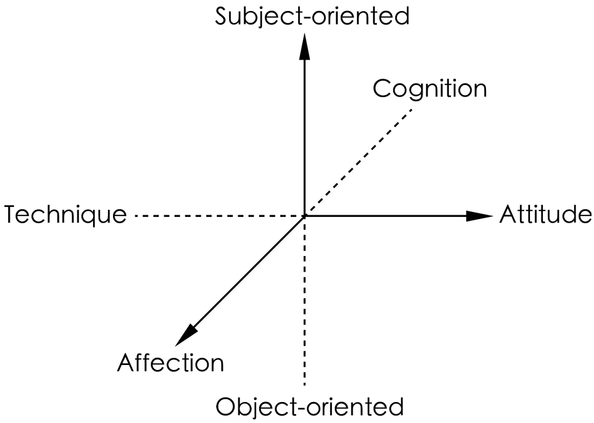


Fig. 1. A system of coordinates on empathy

focus on negative feelings rather than on positive emotions. Baston considered classical empathy as other-oriented feelings [34], which included compassion and sympathy. So, the majority of empathy scales paid more attention to the empathic concern and emotional sharing ability to people who suffered from pain and misfortune. That was the reason why affective empathy was often confused with the term sympathy.

Empathy was often related to design for vulnerable groups in the design field. Under this point of view, first, users were regarded as problematic. Consequently, their capacity was underestimated or ignored, which is further discussed in Sect. 4.2. Secondly, too much focus on negative emotions may lead to emotional exhaustion. Eva Koppen and Meinel considered design activities as a form of emotional work [35] such as social workers, nurses when empathy was discussed in design activities. This understanding may hint a kind of devoted attitude on designers.

The following example demonstrates the problem of an overemphasis on the emotional component. In 2016, the first author attended the DESIS Symposium in Jiangnan University. After a speaker shared her experience of devoting herself to the work of helping mentally disabled people by design interventions, one commented, ‘when you are sharing your work, you show high empathy with disabled people.’ Surprisingly, the speaker responded, ‘actually, I am hard-hearted now. After so many years concentrating on this group of people, I find we cannot solve their problems only by an empathic understanding and warm-heartedness. I have to rethink my work rationally.”

This case, on one hand, revealed that people tended to understand empathy as one kind of emotional connection. On the other hand, it indicated that an overemphasis on emotional inputs in design practice may result in negative effects.

4.2 Object-Oriented or Subject-Oriented

In empathic design, researchers are required to understand user experience in his or her world. This is not an easy task. So technically research in design field emphasized on the importance of self-reflection, which meant 'designers connect with users by recalling explicitly upon their own memories and experiences' [27] in order to enhance empathic understanding with users. Here, self-reflection is a specific technique to help designers gain empathic understanding with users.

The paper 'Spark Innovation through Empathic Design' (published in 1997) [3] may be the first research paper about empathy in design. Steen argued that the researchers' and designers' knowledge and experience were privileged in empathic design since empathic design attempted to fuel creativity by empathizing with end-users [36]. Then we could see, in most design practice with empathic techniques, the user just played a role as a sensitizer, and another kind of tool serving for design innovation. Obviously, it violated the original intention when empathy was introduced to the design field. Users should be served through better empathic understanding.

This kind of perception treats users with indignity, which often impacts the relationship between users and designers. For example, in a project [37] of design for the elderly, some design students in HongKong PolyU went to the nursing home to find problems of the elderly. They observed their living surroundings and interviewed active older people. When asked what difficulties they faced in their life, a grandma told students that 'everything is fine'. The answer made students feel puzzled, as they thought the grandma would complain about the difficulties in her daily life. Then they asked a series of questions: 'does you son often visit you?', 'do you feel lonely?', 'how can you carry heavy things by yourself?' These questions made the grandma angry. She shouted 'why do you always feel we have problems? We are fine! We have happy days!'

In this case, it seemed that the students only focused on identifying a topic for their project rather than paying real attention on the grandma. This example revealed how designers' sense of taking-it-for-granted hurt users. Empathic design sometimes faced with this problem if we tried to understand users only through our own perspectives and experiences.

4.3 Attitude or Technique

The majority of discussion of empathic design tended to focus on tool making. Design researchers have proposed a variety of methods, tools and techniques to facilitate an empathic design process, such as design probes [38], contextmapping [39] and experience prototyping [40]. All these methods and techniques contributed to enhance empathic understanding to users, and help designers embody empathy.

However, a focus on empathy as a capacity and technique draw designers' attention to how we design rather than why we design. Empathy technique then is easy to be considered as a step–by–step operational guideline by many design students and inexperienced design researcher. Ian Hargraves challenged the prevalence of overemphasizing designers' capacity in human-centered design in his Ph.D. dissertation [41]. He argued that HCD should be emphasized for its core principle, i.e. caring for people, not the designers' capability of utilizing human-centered design methods. Similarly, Battarbee believed that empathic understanding went beyond one's knowledge [42]. When empathizing, one does not judge, but 'relate to the user and understand the situation and why certain experiences are meaningful to these people'. Kouprie viewed that training empathy, including both improving the research skills and 'supporting an active and open attitude towards users' [43].

5 Conclusion

In comparison with aesthetic, sociology and psychology, design is a relatively new domain where empathy is discussed. Since the concept of empathy was developed with different interpretations in history, it leads to different understandings and uncertainties when empathy was introduced into design research and design practices.

In order to deeply understand different viewpoints, we built a system of coordinates with three dimensions, which indicates three pairs of divisions, namely, affection and cognition, subject-oriented and object-oriented perspective, and attitude and technique. These three dimensions somehow revealed problems and confusions when utilizing and talking about empathy in design domain. The position in axes indicated six tendencies. However, by this system of coordinates, we do not attempt to make a judgment on which side of the axes is better. Instead, we provided a reference for designers and researchers to map and reflect his or her mindsets when doing research and practice relating to empathy.

References

1. Mattelmäki, T., Vaajakallio, K., Koskinen, I.: What happened to empathic design? Des. Issues **30**(1), 67–77 (2013)
2. Battarbee, K., Koskinen, I.: Co-experience: user experience as interaction. CoDesign **1**(1), 5–18 (2005)
3. Leonard, D., Rayport, J.F.: Spark innovation through empathic design. Harv. Bus. Rev. **75** (75), 102–113 (1997)
4. Sanders, B.N., Dandavate, U.: Design for experiencing: new tools. School of Industrial Design Engineering Department of Industrial Design (1999)
5. Hunsdahl, J.: Concerning einfuhlung (empathy): a concept analysis of its origin and early development. J. Clin. Psychol. **3**, 180–191 (1967)
6. Pigman, G.W.: Freud and the history of empathy. Int. J. psychoanal. **76**(Pt 2), 237 (1995)
7. Lipps, T.: Einfühlung, inner nachahmung, und organumpfindungen (empathy, inner imitations, and sensations). Arch. für die gesamte Psychol. **2**, 185–204 (1903)
8. Breithaupt, F.: Kulturen der Empathie. Suhrkamp, Berlin (2009)
9. Håkansson, J.: Exploring the phenomenon of empathy. Doctoral dissertation, Department of Psychology, Stockholm University, pp. 1–140 (2003)
10. Titchener, E.B.: A text-book of psychology. Br. Med. J. **2**(3268), 3285–3286 (1923)
11. Köhler, W.: Gestalt psychology. Psychologische Forschung **31**(1), XVIII–XXX (1967)
12. Piaget, J.: The Moral Judgment of the Child. Kegan Paul, Trench, Trubner, London (1932)
13. Piaget, J., Elkind, D., Tenzer, A.: Six Psychological Studies. Random House, New York City (1967)
14. Mead, G.H.: Mind, self, and society from the standpoint of a social behaviorist. Philos. Rev. **179**(September), 272–273 (1934)

15. Kohut, H.: Introspection, empathy, and psychoanalysis: an examination of the relationship between mode of observation and theory. In: Ornstein, P.H. (ed.) The Search for the Self, vol. 1, pp. 205–232. International University Press, New York (1959)
16. Rogers, C.R.: The necessary and sufficient conditions of therapeutic personality change. J. Consult. Clin. Psychol. **60**(6), 827–32(1957, 1992)
17. Stotland, E.: Exploratory investigations of empathy 1. Adv. Exp. Soc. Psychol. **4**, 271–314 (1969)
18. Hoffman, M.L.: The contribution of empathy to justice and moral judgment. Reach. Out Caring 47–80 (1987)
19. Batson, C.D.: The Altruism Question: Toward a Social-Psychological Answer. Lawrence Erlbaum Associates, New Jersey (1991)
20. Wispé, L.: The distinction between sympathy and empathy: to call forth a concept, a word is needed. J. Pers. Soc. Psychol. **50**(2), 314–321 (1986)
21. Ickes, W.: Empathic accuracy. J. Pers. **61**(4), 587–610 (1993)
22. Davis, M.H.: Measuring individual differences in empathy: evidence for a multidimensional approach. J. Pers. Soc. Psychol. **44**(1), 113–126 (1983)
23. Mehrabian, A., Epstein, N.: A measure of emotional empathy. J. Pers. **40**(4), 525–543 (1972)
24. Shamay-Tsoory, S.G., Aharon-Peretz, J., Perry, D.: Two systems for empathy a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain A J. Neurol. **132**(Pt 3), 617 (2009)
25. Gladstein, G.A.: Empathy and counseling outcome:An empirical and conceptual review. J. Couns. Psychol. **6**, 70–79 (1977)
26. Bons, D.M.A., Scheepers, F.E., Rommelse, N.N.J., et al.: Motor, emotional and cognitive empathic abilities in children with autism and conduct disorder. In: International Conference on Methods and Techniques in Behavioral Research. ACM (2010)
27. Visser, F.S., Kouprie, M.: Stimulating empathy in ideation workshops. In: Tenth Conference on Participatory Design, PDC 2008, Bloomington, Indiana, USA, October DBLP, pp. 174–177 (2008)
28. Silberstein, L.K.: Human-directed empathy and childhood history of pet ownership and attachment. Dissertations & Theses – Gradworks (2013)
29. Greenson, R.R.: Empathy and its vicissitudes. Int. J. Psychoanal. **41** (1960)
30. McDonagh, D.: Empathic research approaches to support the designer: a supra-qualitative research for designing model. Des. Issues (2006)
31. Riess, H., Kelley, J.M., Bailey, R.W., et al.: Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. J. Gen. Intern. Med. **27** (10), 1280 (2012)
32. MacIsaac, D.S.: Empathy: Heinz Kohut's contribution. In: Bohart, A.C., Greenberg, L.S. (eds.) Empathy Reconsidered: New Directions in Psychotherapy, pp. 245–264. United Book Press, Baltimore (1997)
33. Bozarth, J.D.: Empathy from the framework of client-centered theory and the Rogerian hypothesis, pp. 81–102 (1997)
34. Batson, C.D., Sager, K., Garst, E., Kang, M., Rubchinsky, K., Dawson, K.: Is empathy-induced helping due to self–other merging? J. Pers. Soc. Psychol. **73**(3), 495–509 (1997)
35. Koppen, E., Meinel, C.: Knowing people: the empathetic designer. Des. Philos. Pap. **10**(1), 35–51 (2012)
36. Steen, M., Kuijt-Evers, L.: Early user involvement in research and design projects - a review of methods and practices. In: 23rd EGOS Colloquium (European Group for Organizational Studies), 5–7 July, Vienna (2007)

37. Ho, D.L., Ma, J., Lee, Y.: Empathy @ design research: a phenomenological study on young people experiencing participatory design for social inclusion. CoDesign **7**(2), 95–106 (2011)
38. Mattelmäki, T.: Probing for co-exploring. CoDesign **4**(1), 65–78 (2008)
39. Visser, F.S., Stappers, P.J., van der Lugt, R., et al.: Contextmapping: experiences from practice. CoDesign **1**(2), 119–149 (2005)
40. Buchenau, M., Suri, J.F.: Experience prototyping. In: Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 424–433. ACM (2000)
41. Hargraves, I.: People matter: care in design and healthcare. Doctoral dissertation, School of Design, Carnegie Mellon University (2012)
42. Battarbee, K., Koskinen, I.: Co-experience: user experience as interaction. CoDesign **1**(1), 5–18 (2005)
43. Kouprie, M., Visser, F.S.: A framework for empathy in design: stepping into and out of the user's life. J. Eng. Des. **20**(5), 437–448 (2009)

A Trial of Intellectual Work Performance Estimation by Using Physiological Indices

Shutaro Kunimasa(✉), Kyoich Seo, Hiroshi Shimoda, and Hirotake Ishii

Kyoto University, Yoshida-honmachi Sakyo-ku, Kyoto 606-8501, Japan
{kunimasa, seo, shimoda, hirotake}@ei.energy.kyoto-u.ac.jp

Abstract. In order to evaluate the intellectual productivity quantitatively, most of conventional studies have utilized task performance of cognitive tasks. Meanwhile, more and more studies use physiological indices which reflect cognitive load so as to evaluate the intellectual productivity quantitatively. In this study, the method which estimates task performance of intellectual workers by using several physiological indices (pupil diameter and heart rate variability) has been proposed. As the estimation models of task performance, two machine learning models, Support Vector Regression (SVR) and Random Forests (RF), have been employed. As the result of a subject experiment, it was found that coefficient of determination (R^2) of SVR was 0.875 and higher than that of RF ($p < 0.01$). The result suggested that pupil diameter and heart rate variability were effective as the explanatory variables and SVR estimation was also effective in task performance estimation.

Keywords: Intellectual productivity · Machine learning · Physiological indices · Pupil diameter · Heart rate variability

1 Introduction

Recently, mental works such as intellectual works have occupied most of office works in companies and have become more and more valuable in our society. Therefore, economic and social benefits can be larger by improving intellectual productivity such as the efficiency and accuracy of performing intellectual works. In order to achieve this, the quantitative evaluation of intellectual productivity is required, and several studies have been conducted [1]. However, a number of cognitive tasks used in experiments for the evaluation are different from actual office works, because the tasks have been designed in order for experimenters to collect operation logs easily and accurately. In order to evaluate intellectual productivity in actual office, it is desired to measure it when conducting actual office works. It is, however, difficult to collect and evaluate most of their logs. On the other hand, there are various studies which focused on human internal states (e.g. psychological states and cognitive load). Especially, Cognitive Load Theory study (CLT) [2–5] is one of the examples which have a deep relationship with working memory and cognitive load, which have an influence on several intellectual abilities. The goal of CLT is to estimate human internal states by

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_29

measuring cognitive load in order contribute the design of education and office environment. In this point, CLT is highly compatible with the evaluation of the intellectual productivity. In order to evaluate cognitive load, various conventional studies [6–9] employed physiological indices (e.g. heart rate and pupil diameter). And they revealed the relationship between these indices and cognitive load. Furthermore, several studies [10, 11] performed the estimation of the intellectual work performance by using physiological indices. It is expected that these studies contribute the evaluation of the intellectual productivity of office workers. However, the objective and quantitative evaluation method has not been established yet.

In this study, so as to develop the quantitative evaluation method of the intellectual productivity by physiological response, the estimation method of task performance is considered by using physiological indices measured when performing a cognitive task. As the physiological indices for task performance estimation, heart rate variability and pupil diameter are employed, which are supposed to reflect human cognitive states (e.g. stressed and relaxed). These indices are also expected to be measured by contactless measurement method without giving any load. As an estimation model, two machine learning models, Support Vector Regression (SVR) and Random Forests (RF), have been considered because they are typical methods in the conventional studies which classify several cognitive states by using physiological indices. And the estimation accuracy of these models was compared by using coefficient of determination (R^2) as accuracy indicators.

2 Physiological Indices

It is known that there is a close relationship between cognitive activities and physiological responses. In this study, contactless-measurable physiological indices were employed in order to estimate task performance because it is undesirable that the equipment of physiological indices measurement gives stress to a worker when measuring these indices. Concretely, pupil diameter [12] and heart rate variability [13] were employed because it is expected that contactless and accurate measuring methods will be developed soon.

2.1 Heart Rate Variability

The spectral analysis of heart rate variability has been conducted in various studies because it is supposed to reflect an autonomic nerve response against stress and cognitive load [14, 15]. Mulder and Mulder [16] found that the power of 0.06–0.14 Hz of heart rate variability was reduced by difficulty of cognitive task and high load on working memory. Moreover, low frequency (0.04–0.15 Hz) of heart rate variability (LF), high frequency (0.15–0.50 Hz) (HF) and the ratio of LF and HF (LF/HF) reflect a sympathetic nerve, a parasympathetic nerve and the balance of these nerves, respectively [17]. In this study, these features, LF, HF, and LF/HF, were employed because it is assumed that the change of cognitive load according to task performance gives an impact on the autonomic nerve and it changes these three features (LF, HF, and LF/HF).

In order to extract these feature values, the time window was set to 5 min (5-minute timeframe) because it should be long enough to contain at least 15–30 cycles of LF to make it accurate to extract 3 feature values of heart rate variability. LFs and HFs were extracted by calculating low (0.04–0.15 Hz) and high (0.15–0.50 Hz) frequency wave of heart rate in 5-minute timeframe with shifting it every 1 min. Moreover, LF/HF was also calculated by dividing LF by HF. These feature values were defined as explanatory variables for task performance estimation.

2.2 Pupil Diameter

In the field of psychophysiology, it is known that pupil diameter changes under various cognitive states (e.g. stressed and relaxed). There are various conventional studies which proved the validation of pupil diameter as the index of load on working memory [6–9]. On the other hand, there is a study which suggests a relationship between speed of information processing and the size of pupil diameter [18, 19]. Moreover, there is another study which showed that difficulty of sound classification has a relationship with the size of pupil diameter [21]. In this study, pupil diameter, therefore, was employed as an effective feature for task performance estimation. Mean pupil diameter (MPD) was extracted as a feature value of pupil diameter by calculating the average of pupil diameters in 5-minute timeframe with shifting it every 1 min as well as the heart rate variability.

3 Estimation Method

A cognitive task was employed in this study in which a task worker performs repeatedly problems which can be solved in several seconds. The task performance was measured by counting the number of solved problems as well as the above physiological feature value extractions. Then task performance was estimated by regression analysis with using physiological features (i.e. LF, HF, LF/HF, and MPD) and they were compared with the measured task performance. There are various methods of regression analysis and multiple regression analysis is one of the famous methods. However, it is assumed that regression line cannot be calculated accurately by the analysis because of multicollinearity. On the other hand, there are various studies where machine learning methods are employed as the physiological analysis [20–25]. In this study, therefore, SVR and RF were employed because these methods are supposed to avoid multicollinearity problem and have high generalization capability.

3.1 Support Vector Regression (SVR)

SVR is one of the non-linear regression analysis methods based on Support Vector Machine (SVM), which is effective for 2-class classification and was employed so as to estimate emotions by using several physiological indices [21, 22] and to detect stress [23].

The method to generate SVM classifier is described below. If n features $\boldsymbol{x_i}(i = 1, \cdots, n)$ belong to class $y_i \in \{-1, 1\}$, in order to classify them correctly, the hyperplane $\boldsymbol{w^T x} + b = 0$ ($\boldsymbol{w^T}$ is a coefficient vector, b is a bias term) is calculated which maximizes the distance between the hyperplane and the support vector $\boldsymbol{x}^*$ which is the closest to the hyperplane of all $\boldsymbol{x_i}$. The SVM model (the above hyperplane) can be calculated by solving this problem. The SVR model can be also calculated by using this hyperplane and solving regression problem.

3.2 Random Forest (RF)

RF is the machine learning algorithm proposed by Breiman [26] and is used to solve classification or regression problems as well as SVM and SVR. RF is the application method of the classification method called "bagging", which generates a number of weak learners by using a part of training data and integrate these weak learners by the majority rule. When estimating cognitive states by several physiological indices, it is said that RF and AdaBoost, which is similar to RF, are as accurate as SVM [24, 25].

The method to generate RF regression model is described below. First, n data is extracted with allowing duplication from n training data $\boldsymbol{x_i}(i = 1, \cdots, n)$ ($\dim \boldsymbol{x_i} = d$) and these n data is defined as new training data (bootstrap sampling). Next, a decision tree is generated by using this training data, and $d^{'}$ ($< d$) features are selected randomly in each non-terminal node of the tree. If several decision trees are generated as above, these trees have a weak correlation. After generating m decision trees, RF training finishes. The majority or the average of outputs from these trees is defined as the estimation result in the case of classification or regression problem solving, respectively. In training phase, the decision of the size of tree node, $d^{'}$, and m are required before training. It is known that the node size and $d^{'}$ are recommended to be 5 and $d/3$, respectively in the case of regression problem [26]. On the other hand, m may be large enough because overlearning is not caused even though m is too large. In this study, these values were employed when employing RF as the estimation method.

3.3 Model Calculation

In training phase, SVR and RF models were calculated by using physiological data of one task worker in order to deal with individual difference of physiological responses. Moreover, MATLAB ® [27] was used to calculate models and LIBSVM and randomforest-matlab were used as machine learning library.

4 Experiment

4.1 Purpose

The purpose of this experiment was to validate heart rate variability and pupil diameter as explanatory variables of task performance estimation and to evaluate the accuracy of the proposed methods.

4.2 Participants

31 healthy volunteers participated in this experiment and their native language was Japanese. In this experiment, only those who do not wear glasses participated because the accuracy of eye tracking system gets lower when wearing glasses.

4.3 Measurement of Physiological Indices

In the case of measurement of heart rate, ECG electrodes were pasted on a left rib and a right clavicle. Ground and reference electrodes were pasted on right and left earlobes, respectively. In order to reduce noise signals, the time constant of high-pass filter and the cut-off frequency of low-pass filter were set to 3.0 s and 100 Hz, respectively. Furthermore, notch filter was set to 60 Hz so as to reduce hum noise from commercial power supply.

Pupil diameter was measured by an infrared eye tracking camera, faceLAB5 [12]. A participant sat on the chair while performing tasks on the desk. Then, the position and height of the chair were adjusted and the angle of the eye tracking camera was set to 36° against the desk so as to recognize the face of a participant correctly.

4.4 Cognitive Task and Instruction of Task Performing

In this study, Receipt Classification Task [28] was used as a cognitive task to evaluate task performance. The task was presented on PC display.

In order to avoid overlearning and generate regression models properly, various values of training data should be measured and provided to machine learnings as training data. It is therefore necessary to instruct participants to change task performance (i.e. the number of classified receipts) and collect their physiological indices while their task performance varies. In this study, the participants were instructed to change the speed of classifying receipts according to the indicator of color bar displayed on the left as shown in Fig. 1. Concretely, the slower they were instructed to change the classifying speed, the larger the ratio of blue in the color bar got. On the contrary, the faster they changed the speed, the larger the ratio of red got. At the beginning of the task, the color bar was red or blue. Then, the ratio of the other color got larger and larger at a constant rate. Finally, the ratio got 100% 30 min after the beginning. In this study, the case in which the color changes blue into red was called "Pace-up" Phase, while the opposite case was called "Pace-down" Phase.

4.5 Experimental Protocol

The experiment was conducted for 2 days. The purpose of the first day experiment was (1) to perform participant screening, (2) to get accustomed to experimental environment, and (3) to practice Receipt Classification Task and Pace-up and Pace-down Phase. Regarding (1), the accuracy of eye tracking could be low because the camera could not capture eyes of some participants enough if eyes of them were small. In this case, they don't participate in the second day experiment. In regard to (2), it was supposed that participants got stressed because they participated in the experiment and performed the cognitive task for the first time, which may affect their physiological responses. In order to reduce this psychological influence, the first day experiment was designed to make them adapt to this environment. Finally, regarding (3), if participants perform the cognitive task for the first time, cognitive load on them may be high even if the difficult of the task was low. However, if they get used to the task, cognitive load may be lower even though the difficulty was the same. This is generally called "Learning Effect". This effect should be reduced because it gives an influence on physiological indices. In order to remove this effect, they practiced the task, Pace-up, and Pace-down Phase in the first day experiment.

Table 1. Experimental schedule: second day

Duration (minutes)	Content
10	Setting of Electrodes and Eye Tracking Camera
10	Task Practice
30	Phase A*
10	Rest
30	Phase B*
5	Removal of the Instruments

*Measurement of Physiological Indices and Task log.

The experimental schedule of the second day is shown in Table 1. Phase A in this table was set to either Pace-up or Pace-down Phase at random while Phase B was set to the other in order to get counterbalance of ordering effect. Physiological indices and task log were recorded in these phases.

5 Result

5.1 Participant Screening

As the result of screening in the first day experiment, pupil diameters of all the participants could be measured correctly by the eye tracking system. Then all of them participated in the second day experiment. However, the heart rate variability data of 4 participants were not properly measured because strong artifacts appeared. Thus, the experimental results of 27 participants were analyzed, where 13 participants performed Pace-up Phase in Phase A, while the other 14 did it in Phase B.

5.2 Task Performance Estimation

After the experiment, SVR and RF models which estimate task performance (i.e. the number of classified receipts) were calculated every participant by using the physiological features such as LF, HF, LF/HF and MPD. Then, the accuracies of these models were evaluated by using 13-fold cross validation.

Average of coefficients of determination R^2 of each method are described in Table 2.

Table 2. Average of R^2 of each method.

Method	R^2	
	Average	SD
SVR	0.875*	0.097
RF	0.648	0.199

*R^2 of SVR was significantly higher than that of RF ($p < 0.001$)

As the result of paired and two-tailed t-test, R^2 of the SVR model was significantly higher than that of RF ($p < 0.001$).

An example of the SVR estimation is shown as Fig. 1 ($R^2 = 0.969$). The values of vertical axis in Fig. 1 was normalized to [−1,1]. "Measured Values" in Fig. 1 is the number of classified receipts and "Estimated Values" is the one estimated by SVR.

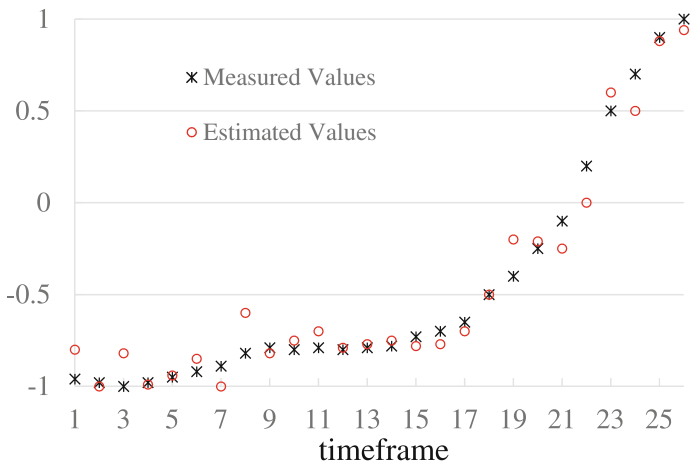


Fig. 1. An example of the SVR estimation (Pace-up Phase).

6 Discussion

6.1 Estimation Methods

As the comparison of the result between the SVR and RF models, the SVR model has significantly better estimation performance in terms of R^2. RF is the ensemble learning method which reduces generalization error by suppressing the variance with keeping the bias of model low. In order to achieve it, RF generates a number of decision trees which have weak relationship to each other by using bootstrap sampling. However, if the number of training data is small, RF cannot generate various decision trees enough to estimate response variables accurately. In this study, there were 52 feature data per a participant. Finally, only 48 training data per a participant remained because 13-fold cross validation was employed ($52 \times 12/13 = 48$). Regarding RF, the number of this sample, 48, was not enough to estimate task performance accurately. On the other hand, in the case of the SVR model, the number of tunable parameters is more than the RF model. Therefore, the performance of the SVR model is supposed to be significantly more accurate in this study.

6.2 Physiological Indices

In this study, the contribution of feature values (LF, HF, LF/HF and MPD) was considered in the case of the SVR model. These contributions can be discussed by evaluating each component of the coefficient vector $\boldsymbol{w}$ in the hyperplane $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ calculated by SVR training. Average coefficients of the feature values are shown in Table 3. It was found that the contribution of MPD was the highest among the feature values and that of LF was the second highest.

Table 3. Average coefficients of feature values.

	MPD	LF	LF/HF	HF
Average of coefficients	2.00	−0.98	−0.71	−0.27

Regarding the sign of each coefficient, MPD had positive relation to task performance. The result was supported by the study conducted by Poock [18] which shows the positive relationship between the size of pupil diameter and the speed of information processing. In the case of Receipt Classification Task, when the speed of performing task got faster, it is assumed that it requires parallel information processing in which participants should memorize several components (e.g. the date, amount, and name of company of a receipt). Then, it was supposed that their pupil diameter got larger [6, 7] because of the high cognitive load. On the other hand, LF had negative relation to task performance. According to Mulder and Mulder [16], the higher the difficult of a cognitive task gets, the lower the power of LF gets. This supports the result of this study.

6.3 Individual Difference

As mentioned above, the contribution of MPD was the highest, while MPD of some subjects had a weak correlation with task performance and heart rate variability of them had a strong correlation. Table 4, Figs. 2, and 3 show an example (EX1). The values of vertical axis in Figs. 2 and 3 were normalized to [−1,1].

Table 4. Coefficients of feature values in EX1.

	MPD	LF	LF/HF	HF
Coefficients	−0.02	−3.05	−2.06	−3.38

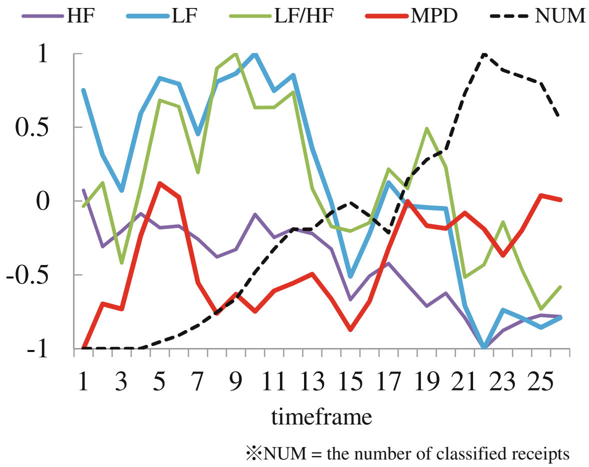


Fig. 2. Physiological features and task performance in EX1.

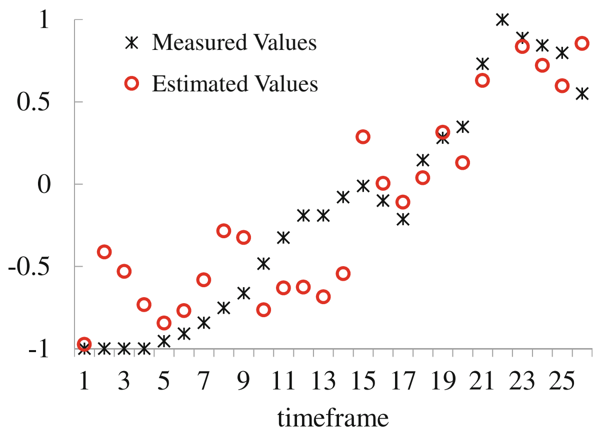


Fig. 3. An example of the SVR estimation in EX1 (Pace-up Phase).

Comparing Table 4 with Table 3, it shows coefficients of EX1 are different from the averages by individual difference. Therefore, if the estimation method such as SVR and RF employs a single explanatory value, the estimation performance can be low. In the case of the SVR model which employed these feature values, R^2 in EX1 was 0.878, respectively. The performance is so high, which suggested that plural physiological indices make task performance estimation accurate even if the individual differences are found.

7 Conclusion

In this study, the method which estimates task performance of intellectual workers by using several physiological indices (pupil diameter and heart rate variability) was proposed in order to develop the objective and quantitative evaluation method of the intellectual productivity. As the physiological indices for task performance estimation, heart rate variability and pupil diameter were employed, which were supposed to reflect human cognitive states (e.g. stressed and relaxed). These indices are also expected to be measured by contactless measurement method without giving any load. As the estimation model of task performance, two machine learning models, Support Vector Regression (SVR) and Random Forests (RF), have been employed. As the result of a subject experiment, it was found that coefficient of determination (R^2) of SVR was 0.875 which was higher than that of RF ($p < 0.01$). The result suggested that pupil diameter and heart rate variability were effective as the explanatory variables and SVR estimation was also effective in task performance evaluation.

However, the cognitive task, Receipt Classification Task, which was employed in this study, requires only a part of the cognitive processing used by office works. It is therefore necessary to consider the accuracy of the proposed method with various cognitive tasks. In the future, the authors are aiming at developing the more accurate evaluation method of the intellectual productivity.

Acknowledgements. This study was supported by JSPS KAKENHI Grant Number 23360257.

References

1. Indoor Health and Productivity Project, Lawrence Berkeley National Laboratory. https://iaqscience.lbl.gov/. Accessed Jan 2017
2. Brünken, R., Plass, J.L., Leutner, D.: Direct measurement of cognitive load in multimedia learning. Educ. Psychol. **38**(1), 53–61 (2003)
3. Paas, F., Tuovinen, J.E., Tabbersm, H.P., Gerven, W.M.: Cognitive load measurement as a means to advance cognitive load theory. Educ. Psychol. **38**(1), 63–71 (2003)
4. Paas, F., Merriënboer, J.J.G.: The efficiency of instructional conditions, an approach to combine mental effort and performance measures. Hum. Factors **35**(4), 737–743 (1993)
5. Ayres, P., Paas, F.: Cognitive load theory, new directions and challenges. Appl. Cogn. Psychol. **26**(6), 827–832 (2012)
6. Hess, E.H., Polt, J.M.: Pupil size in relation to mental activity during simple problem-solving. Science **143**, 1190–1192 (1964)
7. Yamamoto, H., Shimizu, Y.: Change of pupil size and blinking time in memorizing numerals. J. Inst. Telev. Eng. Jpn. **45**(3), 370–376 (1991)
8. Ishibashi, T., Ohtani, A., Miura, T.: Heart rate as an index of the mental load. Jpn. Soc. Occup. Health **10**(7), 377–379 (1968)
9. Kennedy, D.O., Scholey, A.B.: Glucose administration, heart rate and cognitive performance, effects of increasing mental effort. Psycho-pharmacology **149**(1), 63–71 (2000)
10. Nishihara, N., Wargocki, P., Tanabe, S.: Cerebral blood ow, fatigue, mental effort, and task performance in offices with two different pollution loads. Build. Environ. **71**, 153–164 (2014)

11. Kunimasa, S., Miyagi, K., Shimoda, H., Ishii, H.: Detection of temporary rest state while performing mental works by measuring physiological indices. Trans. Hum. Interface Soc. **16** (1), 41–50 (2014)
12. facelab, Seeing Machines, Inc. http://www.seeingmachines.com/. Accessed Jan 2017
13. Poh, M.Z., McDuff, D.J., Picard, R.W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt. Express **18**(10), 10762–10774 (2010)
14. Jorna, P.G.A.M.: Spectral analysis of heart rate and psychological state, a review of its validity as a workload index. Biol. Psychol. **34**(2), 237–257 (1992)
15. Acharya, U.R., Joseph, P.K., Kannathal, N., Min Lim, C., Suri, J.S.: Heart rate variability, a review. Med. Biol. Eng. Comput. **44**(12), 1031–1051 (2006)
16. Mulder, G., Mulder, L.J.M.: Information processing and cardio vascular control. Psychophysiology **18**(4), 392–402 (1981)
17. Saida, S., Akamatsu, M., Inukai, Y., Kuchinomachi, Y., Nakamura, K., Nagamura, N., Yoshioka, M.: Human Measurement Handbook. Asakura Publishing Co., Ltd., Tokyo (2003)
18. Poock, G.K.: Information processing vs pupil diameter. Percept. Mot. Skills **37**(3), 1000–1002 (1973)
19. Kahnemann, D., Beatty, J.: Pupillary responses in a pitch-discrimination task. Percept. Psychophys. **2**(3), 101–105 (1967)
20. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence, analysis of active physiological state. IEEE Trans. Pattern Anal. Mach. Intell. **23**(10), 1175–1191 (2001)
21. Takahashi, K.: Remarks on SVM-based emotion recognition from multi modal bio-potential signals. In: 13th IEEE International Workshop on Robot and Human Interactive Communication, pp. 95–100 (2004)
22. Katsis, C.D., Katertsidis, N., Ganiatsas, G.: Toward emotion recognition in car-racing drivers. A biosignal processing approach. IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum. **38**(3), 502–512 (2008)
23. Zhai, J., Barreto,A.: Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2006, pp. 1355–1358 (2006)
24. Zhao, C., Zheng, C., Zhao, M., Liu, J.: Physiological assessment of driving mental fatigue using wavelet packet energy and random forests. Am. J. Biomed. Sci. **2**(3), 262–274 (2010)
25. Miyaji, M., Kawanaka, H., Oguri, K.: Driver's cognitive distraction detection using physiological features by the adaboost. In: Proceeding of the 12th International IEEE Conference on Intelligent Transportation Systems, pp. 90–95 (2009)
26. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
27. Inc. MathWorks: MATLAB. https://www.mathworks.com/products/matlab.html. Accessed Jan 2017
28. Miyagi, K., Oishi, K., Uchiyama, K., Ishii, H., Shimoda, H.: Proposal of intellectual productivity model based on work state transition. In: 15th International Conference on Human-Computer Interaction, Engineering Psychology and Cognitive Ergonomics, vol. 16, pp. 335–343 (2013)

General and Systemic Structural Activity Theory

The Model of the Factor of Significance of Goal-Directed Decision Making

Alexander Yemelyanov(✉)

Department of Computer Science, Georgia Sothwestern State University,
800 Georgia Southwestern State University Drive, Americus, GA 31709, USA
Alexander.Yemelyanov@gsw.edu

Abstract. From the position of systemic-structural activity theory, this paper presents a critique of prospect theory as a descriptive model of decision-making under risk and uncertainty and develops an alternative model, which is called the significance model. This model is psychologically determined and reflects the emotionally evaluative mechanism of goal-directed decision making where attractiveness of the goal and the uncertainty (subjective difficulties) of attaining it should be taken into consideration. Based on the suggested significance model, a mobile decision making software application – Express Decision, was constructed. The paper illustrates how Express Decision can be implemented in the analysis of a pilot's decisions within the context of an aviation accident.

Keywords: Systemic-structural activity theory · Decision making · Factor of significance · Goal · Motivational force · Mobile decision support

1 Introduction

Decision theories and motivational theories present two relatively prominent theoretical approaches to the analysis of decision making under conditions of risk and uncertainty. Decision theories assume that people's preferences depend on two factors: the value people attribute to the outcomes of different courses of action and the probability that each of the outcomes will happen. With this in mind, the normative approach places emphasis on how to make the best decisions by deriving an algebraic representation of preference from abstracted behavioral axioms, whereas the descriptive approach uses this algebraic representation with the incorporation of people's preferences for safety or risk. The motivational theories assume that people are not just concerned with the objective consequences of decisions, but also with the affective consequences of decisions and the motivational states that enhance them. Considering this, the main motivational factors influencing choice remain being presented by a simple combination of algebraic symbols.

While decision and motivational theories contribute considerably to the understanding of human decision making, in this paper, we argue that systemic-structural activity theory (SSAT) suggest more fruitful and promising approach for describing and predicting choices in both laboratory and applied settings. According to the activity approach, decision-making is a goal directed system in which cognition, behavior and motivation are integrated and organized by goals and the mechanisms of self-regulation.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_30

The factor of significance retains an important place in emotionally motivated processes of activity regulation, specifically in the evaluation of decision options and goal formation.

This paper presents results of psychological studies of significance, determines its main components, and proposes the model of significance. Such a significance model, as is demonstrated below, presents not only the significance of the decision alternative but also its motivational force and, therefore, can serve as a mechanism for decision making. This article ends with a description of the implementation of the significance model in a prominent aviation accident that took place on January 15, 2009, when the pilot had to arrive at a decision to land on the Hudson River after loss of thrust in both engines.

2 Decision Theories

The Decision theories have usually limited their consideration to cognitive phenomena such as choice and risk. The two main theories of choice are expected utility theory and prospect theory. Although both of these theories have different assumptions about subjective representations of objective outcomes and probabilities, they present decision rule as an algebraic function of the probabilities and values of the outcomes of each alternative.

Expected utility theory [1] – major normative theory of decision-making under risk is constructed based on a set of axioms, for example, transitivity of preferences, which provide criteria for the rationality of choices. According to this approach, decisions can be abstracted and represented as the selection of a single course of action X described by the value of the possible outcomes $\{x_1, x_2, ..., x_n\}$ and the associated probability $\{p_1, p_2, ..., p_n\}$ that each outcome would occur if the action were selected. The option X that has the highest expected utility $EU(X)$ will be selected:

$$EU(X) = p_1 \cdot x_1 + \ldots + p_n \cdot x_n. \quad (1)$$

For example, consider a prospect with two options: (A) a certain outcome valued at \$100, and (B) a risky option with an 80% chance of \$100, a 15% chance of \$200, and a 5% chance of receiving nothing. The expected utility calculation recommends that one should take the second option, because $EU(B)$ = \$110 > \$100 = $EU(A)$.

Although expected utility theory allows for individual differences in attitudes about risk, it does not describe any sort of psychological mechanism that underlies risk attitudes, such as satisfaction with success and fear of failure, and individual differences in motivation, such as need for security and avoidance of failure [2].

Descriptive theories of decision making accept the algebraic representation of decision rule, but incorporate known limitations of human behavior. Kahneman and Tversky [3, 4] noted that people exhibit patterns of preference, which appear incompatible with expected utility theory. They produced a descriptive model, called prospect theory, which modifies expected utility theory so as to accommodate these observations. Prospect theory is a behavioral economic theory that describes the way people choose between probabilistic alternatives that involve risk, where the probabilities of outcomes are known. Prospect theory introduced the notion of reference dependence, in which outcomes are not evaluated absolutely but relative to some benchmark or reference point. In this context, relative to the reference point, outcomes were evaluated differentially based on whether they were seen as gains (utility function *UG*(*x*)) or losses (utility function *UL(x)*). It was also proposed the concept of loss aversion, that the marginal utility of a constant change is greater for losses (a $10 loss is more aversive than a $10 gain is pleasant). Once the decision maker evaluates all prospects, he or she then chooses the prospect that offers the highest overall value *V,* which is calculated in a manner analogous to the expected value *EU(X)*:

$$V(x,p) = w(p_1)v(x_1) + \ldots + w(p_n)v(x_n). \tag{2}$$

However, decision weight $w(p_i)$ replaces probability p_i, and subjective value functions $v(x_i)$ replace value x_i. It is particularly worth noting that functions *v(x)* и *w(p)* are individual for each decision maker and serve as adjustment factors of that individual's perception of value and probability of outcomes. In view of this, as the authors of the prospect theory point out, decision weights are not probabilities, because "they do not obey the probability axioms and they should not be interpreted as measures of degree or belief" [3, p. 280]. With the incidence of uncertainty in the occurrence of outcomes, as it can be assumed from further analysis, the role of these weights escalates, since they present a subjective difficulty of the alternative and the subject's emotional reaction to this difficulty. Then again, decision weights in Prospect Theory remain psychologically undetermined [5]; therefore, for the occurrence of situations with uncertainties (unknown event probabilities), prospect theory as a theory for decisions under risk (known event probabilities) loses its descriptive properties, becoming misleading and logically contradictive.

As an illustration of the aforementioned theory, let us analyze the following decision-making problem in a hypothetical life and death situation presented by Kahneman and Tversky [4, 6] and widely discussed in several different papers on risk psychology [7], etc.

Participants were asked to choose between two alternative programs to combat an unusual Asian disease, which is expected to kill 600 people. This decision-making problem was presented to participants with positive framing (survival format), i.e. how many people would live, and with negative framing (mortality format), i.e. how many people would die. Results are presented below in the form of prospects (the percentage who chose each option is indicated in parentheses).

- Positive framing:
  - If Program A is adopted, 200 people will be saved. (72%)
  - If Program B is adopted, there is a 1/3 probability that 600 people will be saved and 2/3 probability that no people will be saved. (28%).
- Negative framing:
  - If Program C is adopted 400 people will die. (22%)
  - If Program D is adopted, there is 1/3 probability that nobody will die and 2/3 probability that 600 will die. (78%).

Kahneman and Tversky claim that positive and negative framing result in different descriptions of the same problem, where programs C and D are undistinguishable in real terms from programs A and B [6, p. 343]. At the same time, experimental results show that most participants chose programs A and D, despite the fact that in terms of consequences, these choices are contradictory. Because programs A, C and B, D are equivalent. This is a violation of the logical principle of extensionality in decision-making which states that making a decision in a problem should not be affected by how the problem is described [8]. Tversky and Kahneman, admitting the fallacy of such a choice, explained it by employing the general properties of people's attitude towards a risk: people are expected to show a risk-seeking preference when faced with negatively framed problems and risk aversion when presented with positively framed ones. It is clear that such an explanation is superficial and does not provide the answers to many questions, such as why participants were being illogical with their answers and why programs B and D – with the same statistical outcomes (200 living and 400 deceased) – were differently evaluated by participants.

The explanation for this can be garnered from the position of the SSAT approach, in decision making is considered as a goal-directed process where the decision choice is determined not only by people's attitude toward the outcomes, but also by their attitudes toward the expected difficulties in attaining positive and avoiding negative outcomes. In reality, in the analysis of the Asian disease, participants made decisions in two different problems with two different goals: Problem #1 (positive framing) has one specific goal – "save all 600 lives" and Problem #2 (negative framing) has another specific goal – "do not allow any living patient out of 600 to die". Considering this, program B motivated to "save lives", which is primarily associated with the medicinal effects of treatment to combat a disease, while program D motivated to "prevent loss", which is often associated with prophylactic measures to prevent illness. This conclusion also comes about because the given information regarding survivors is typically present in programs regarding treatment, while information regarding the deceased is generally found in programs related to preventive measures. And since with identical expected outcomes (200 survivors and 400 deceased) present in both programs, prevention is generally associated with fewer efforts than any sort of treatment undertakings, it can be suggested that program D is more preferable than program B. Such a conclusion is also consistent with the answers obtained from the participants of the experiment: 78% > 28%.

By this method, the application of prospect theory is mostly limited to decision-making problems with monetary outcomes and stated probabilities. Prospect theory fails when it is applied to problems with uncertainties and where task complexity and subjective difficulties should be taken into account. The absence of a goal in the problem being investigated leads to the situation that the same decision can elect different choices depending on how the problem is framed (edited).

3 Motivational Theories

The motivational approach assumes that when people making decisions, not only the subjective values of outcomes should be considered, but the affective components of decisions and the motivational factors that enhance them in achieving a goal should be considered, as well. The motivational theories propose that the choices people make are often the result of a compromise between two competing desires. On the one hand, people want to choose the option that maximizes their outcomes (in terms of the value of the outcome); on the other hand, they want to avoid making poor decisions that may cause feelings of failure or disappointment. The tension between these two motives pushes decision makers toward a specific level of risk-seeking or risk-averse behavior. In Lewin's field theory [9], the motivational force (S) is considered as a function of two basic indicators of the theory – the valence (V) of the event (the strength of its attractiveness or repulsion) and the expected probability (P) of occurrence of an event with this valence: $S = f(V,P)$. This field theory position became widespread throughout various models that formed the basis for developing new branches of field theory, such as cognitive valence theory, expectancy theory of motivation, achievement motivation theory, etc., all of which found various practical applications. For example, according to the Vroom's expectancy theory of motivation [10], when deciding among behavioral options, individuals select the option with the greatest amount of motivational force $MF = Expectancy \times Instrumentality \times Valence$. Atkinson [11], in analysis of achievement-oriented activity, proposed that preferences for different probabilities of achieving success or avoiding failure are related to individual differences in motivation. The several variables are combined multiplicatively to obtain a value termed "resultant motivation strength", which indicates tendencies to either approach success or avoid failure:

$$(M_s \times P_s \times I_s) + (M_f \times P_f \times I_f). \tag{3}$$

This model incorporates six variables: motive to achieve success (M_s) and motive to avoid failure (M_f); subjective probabilities of success (P_s) and failure (P_f); and incentive values of success (I_s) and failure (I_f).

It must be noted, however, that neither Lewin nor his successors could clarify specific functional dependences that existed among the indicators. They considered motivation to be a simple product of valence(s) and probability(s), without any proof of how these dependences are psychologically determined.

4 Systemic-Structural Theory of Activity

Based on the activity and apply activities theories, G. Z. Bedny and his associates [12–14] developed the Systemic-Structural Activity Theory (SSAT), which allow to provide a systematic view to decision making under risk and uncertainty.

From the position of SSAT, goal operate as a *reference point*, dividing outcomes into two areas – *positive* and *negative*. With each decision alternative comes its own related group of positive and negative outcomes with corresponding subjective expectations of their occurrence, which (expectations) motivate the decision maker to select one of the alternatives.

The level of motivational force of the alternative serves as the criterion for such a selection. The motivational force of the alternative depends on its *significance* (the level of emotional reaction to the sense of the alternative [15, p. 37] and its *difficulty* (the level of subjective complexity of the alternative). The relationship between difficulty and significance determines the *level of motivation*: if the alternative is very difficult for the individual and significant for him or her (subjectively important), the alternative can be accepted; if the alternative is very difficult but not significant, it can be rejected [13, p. 65].

The significance of the alternative has two components: *positive significance* measures the attractiveness of positive outcomes and *negative significance* measures the aversiveness of negative outcomes.

The alternative's difficulty includes two types of difficulties: *positive difficulties* and *negative difficulties* that related to attaining positive and avoiding negative outcomes, respectively. Positive (negative) difficulties are measured by the level of subjective likelihood determining whether positive (negative) outcomes can be attained (avoided).

5 The Significance Model

The significance function (S) is presented as a composition of two functions – *positive* (S^+) and *negative* (S^-) *significances:*

$$S(I^+, L^+, I^-, L^-) = S(S^+(I^+, L^+), S^-(I^-, L^-)), \tag{4}$$

where

- *the positive significance function* (S^+) is considered as a composition of *positive intensity* (I^+) (subjective value of the positive outcomes in the context of attaining the goal) and *positive likelihood* (L^+) (subjective expectations of I^+);
- *the negative significance function* (S^-) is considered as a composition of *negative intensity* (I^-) (subjective value of the negative outcomes in the context of attaining the goal) and *negative likelihood* (L^-) (subjective expectations of I^-);
- $I^+, I^- \in$ {zero, extremely weak, very weak, weak, not weak-not strong, strong, very strong, extremely strong, maximum};
- $L^+, L^- \in$ {never, extremely seldom, very seldom, seldom, not seldom-not often, often, very often, extremely often, maximum};
- $S, S^+, S^- \in$ {very low, low, middle, high, very high}.

In determining the significance functions S, S^+, S^-, Kotik [16–18] experimented with individuals who used soft (vague) *verbal characteristics* in decision making. Considering the capacity of such words to express people's emotions, attitudes, and judgments, Kotik used words such as "weak– strong" to determine the valence of an event (its *intensity*), and words such as "seldom – often" to establish the *likelihood* of its occurrence; he also measured the event's significance by using words such as "low – high". Kotik conducted a series of experiments with pilots, factory workers, and students and experimentally created graphs and that illustrated the dependence of positive significance and negative significance from the likelihood and intensity of expected outcomes (Fig. 1), as well as created graphs that combined the values of positive and negative significances into one integral value of significance (Fig. 2).

Overall, Kotik's experiments were capable of revealing and graphically demonstrating the general principles of people's attitudes towards emotion-inducing circumstances in the process of decision making, regardless of the physical, social, or material nature of

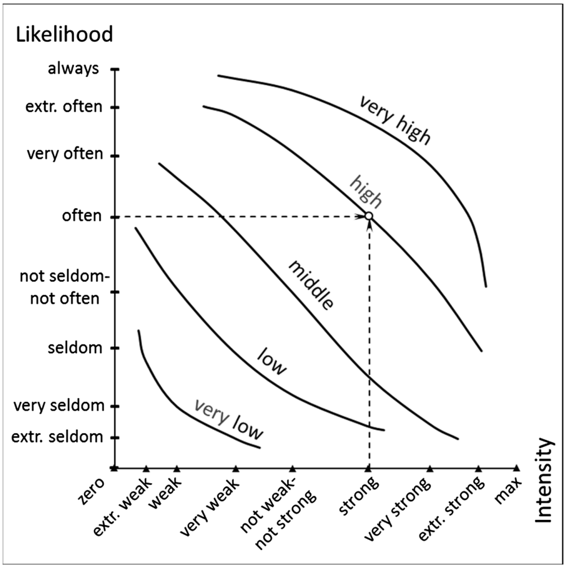


Fig. 1. Levels of positive (negative) significance as a function of intensity and likelihood.

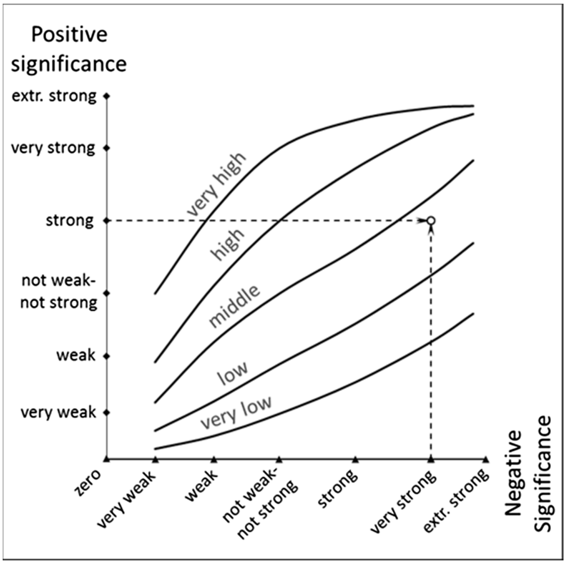


Fig. 2. Levels of attractiveness of emotive events as a function of their positive and negative significances.

these situations. To better establish the validity and reliability of the findings, additional experimental illustrations and diagrams were provided. The advantages of employing the method suggested by Kotik include its ability to evaluate not only conscious emotional responses, but also the unconscious preferences of people, with a simplicity of implementing it into real-life scenarios and analyzing the results achieved by way of this method.

According to the significance model, significance of the alternative is determined from the tetrad (I^+, L^+, I^-, L^-) of soft verbal characteristics by using the following aggregation scheme (Fig. 3).

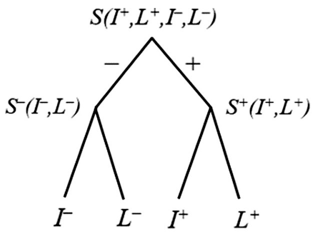


Fig. 3. Significance of the alternative.

For example, pair (I^- = "strong", L^- = "often") determines the "high" level of the negative significance (Fig. 1) and pair (S^+ = "strong", S^- = "very strong") – the level of significance between "middle" and "low" (Fig. 2).

It is worth pointing out that the suggested significance model includes not only an individual's emotional reaction, but also positive (L^+) and negative (L^-) difficulties. This allowed us to assume that significance model presents not only significance of the alternative but also its motivational force and, therefore, can serve as a mechanism for decision making. To verify the validity of this assumption, a special experiment, with 100 participants, was created, which allowed to determine that in 77% of cases, significance model correctly determined the best option for students, while in 66% of the cases, this model indicated the exact same priority ranking as selected by the students [19].

From the description of the significance model, it follows that its practical application presents *certain requirements* to the individual and the problem that must be solved by him. The individual must have *one conscious goal* that he is keen to attain due to his being motivated by its potential (positive) outcomes. These outcomes can remain unconscious separately or may be only partly unconscious, but despite this, according to the goal, they *must be separated* (consciously or unconsciously) into positive and negative categories. The *motivational force* is determined by the *attractiveness of the goal* in the context of attaining positive and avoiding negative outcomes, as well as by the *subjective difficulties* related to these two processes.

6 Express Decision

Based on the significance model, a mobile decision making software application - *Express Decision* (ED), was designed with the purpose of supporting individuals who need to make quick decisions (in as little as 10 s to a couple minutes) in situations of both uncertainty and risk, in which individual biases and both conscious and unconscious emotional factors may substantially influence decision-making. For this application, the significance functions were extrapolated, and the value of these functions were presented numerically in the range from 0 to 1: $0 = S_{\min} \leq S, S^+, S^- \leq S_{\max} = 1$.

When using ED, the user, first of all, should determine the number of decision options (alternatives) available to him, proceed to name them, and then, for each option, sequentially define two pairs of characteristics on a verbal scale with a jumping slider, one of which describes the intensity (weak – strong) and likelihood (seldom – often) of the alternative's pros (+), while the other one describes the intensity (weak – strong) and likelihood (seldom – often) of the alternative's cons (–). After all data has been entered, ED ranks the user's alternatives from best to worst by means of scores. If the user's perception of the problem changes, he or she should indicate this by altering the previously inserted characteristics, which in turn may change the priority of his decision option.

Below is a demonstration of how ED can help the investigator in the analysis of the following aviation accident, in which, as it was stated in the NTSB Accident Report [20, p. 123], the decision-making of the flight crewmembers was one of the key contributing factors to the survivability of the accident.

US Airways Flight 1549 which, three minutes after takeoff from New York City's LaGuardia Airport on January 15, 2009, struck a flock of Canada geese just northeast of the George Washington Bridge and consequently lost all engine power. Unable to reach any airport, pilot glided the plane to a ditching in the Hudson River off midtown Manhattan. All 155 people aboard were rescued by nearby boats and there were few serious injuries.

During the analysis of the flight crew's decision options, investigators were interested in the process of the flight crew's decision making in response to the bird strike for the last 3 min before the landing occurred, including the crew's final decision to land in the Hudson River. ED might be capable of providing investigators with support in the process of arriving at their final decision regarding the landing location between the Hudson River and the Teterboro airport (TEB).

It is worth noting that the verbal characteristics of the likelihood and intensity of anticipated consequences that ED uses in its analysis are presented exclusively for illustrative purposes and are not actual expert conclusions on the observed scenario.

6.1 Decision Making with ED: "Hudson" vs. "TEB"

Which place (the Hudson River or TEB) was more preferable for landing at 3:29:28, considering that successful landing was a goal and taking into account the following flight information and FAA transcript of the communications.

Flight Information. Altitude of 1200 feet, airspeed of 194 knots (359 km/h), and a distance of about 5.5 miles east-northeast of the approach end of runway one at TEB.

FAA Transcript of the Communications between the Captain and Air-Traffic Control Groups from New York Tracon.

3:28:50, Flight 1549:	"I am not sure if we can make any runway. Oh, what's that over to our right in New Jersey may be Teterboro"
3:29:02, New York Tracon:	"Do you want to try and go to Teterboro"
3:29:03, Flight 1549:	"Yes"
3:29:21, New York Tracon:	"You can land runway one at Teterboro"
3:29:25, Flight 1549:	"We can't do it"
3:29:28, Flight 1549:	"We're gonna be in the Hudson"

Option 1. "Hudson" (Ditch on the Hudson River).

Pros (+). Due to the following positive conditions: the aircraft was EOW equipped for an extended overwater flight and had good visibility on water; vessels were immediately available to rescue the passengers and crewmembers, and the flight was executed by a very experienced flight crew, we can conclude that the pilot had a very good chance ("very often" - L_1^+) to land successfully on the water, so that the subsequent emergency event could be deemed a "not weak-not strong" (I_1^+) positive result.

Cons (–). It was still unclear as to how well the ditching on the Hudson River in the icy waters would go, as well as whether the passengers would have enough time to prepare for it; also, there was a concern that some unexpected obstacles on the water could deter the ditching (there was a conflict alert between the accident airplane and the helicopter at 1529:10). Then, we can conclude that although "often" (L_1^-), the pilot still anticipated "strong" (I_1^-) negative consequences.

Option 2. "TEB" (Land at TEB).

Pros (+). Considering the fact that TEB was relatively close and ready to let the plane land (the runway was cleared for landing), as well as that landing on ground is more preferable than on water, for the pilot to land successfully at TEB was an "strong" (I_2^+) positive event that was also difficult ("very seldom" – L_2^+) to execute when both engines experienced an almost total loss of thrust.

Cons (–). Because the flight to TEB would have taken place over residential areas while the airplane was descending quickly, the pilot had very high odds ("very often" – L_2^-) that he would not land successfully at TEB, with "very strong" (I_2^-) negative consequences for the surrounding community.

Decision Making with ED. ED aggregated two pairs of verbal characteristics of option 1 (I_1^+, L_1^+) and (I_1^-, L_1^-) in its significance level S_1 and two pairs of verbal characteristics of alternative 2 (I_2^+, L_2^+) and (I_2^-, L_2^-) in its significance level S_2 and finally determined that option 1 has twice higher level of significance than option 2: $S_1 = 0.52$ and $S_2 = 0.26$.

It means that option 1 is twice more preferable than option 2, which is consistent with NTSB conclusions: "the captain's decision to ditch on the Hudson River rather than attempting to land at an airport provided the highest probability that the accident would be survivable" [20, p. 89].

Additionally, ED allowed to compare levels of positive significance and negative significance related to the pros and cons of each option and thus conclude that ditching on the Hudson River presented a greater expectation for successful landing ($S_1^+ = .67 > S_2^+ = .43$) with lesser risk for passengers and the surrounding community ($S_1^- = .67 < S_2^- = .86$).

The positive, negative, and resultant motivational forces for "Hudson" ($\vec{S}_1^+, \vec{S}_1^-, \vec{S}_1$) and "TEB" ($\vec{S}_2^+, \vec{S}_2^-, \vec{S}_2$) can be illustrated graphically as vectors which magnitudes (lengths) present significances of the corresponding motivational forces (Fig. 4).

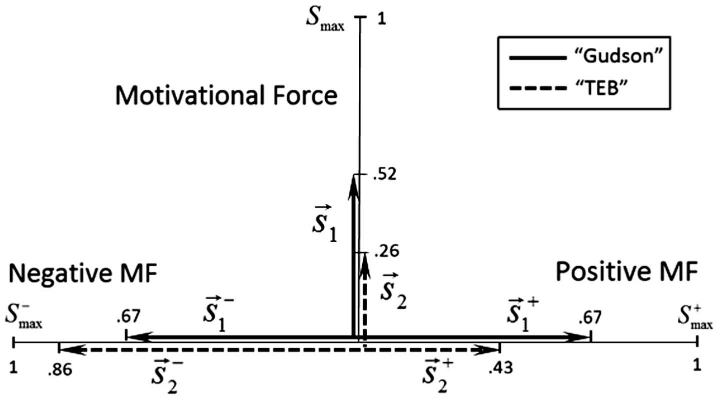


Fig. 4. Graphical comparison of motivational forces of "Hudson" and "TEB".

References

1. von Neumann, J., Morgenstern, O.: The Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944)
2. Larrick, R.P.: Motivational factors in decision theories: the role of self-protection. Psychol. Bull. **113**(3), 440–450 (1993)
3. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–291 (1979)
4. Tversky, A., Kahneman, D.: The framing of decisions and the psychology of choice. Science **211**(4481), 453–458 (1981)
5. Johnson, J.G., Busemeyer, J.R.: Decision making under risk and uncertainty. Wiley Interdisc. Rev.: Cogn. Sci. **1**(5), 736–749 (2010)
6. Kahneman, D., Tversky, A.: Choices, values, and frames. Am. Psychol. **39**(4), 341–350 (1984)
7. Bless, H., Betsch, T., Franzen, A.: Framing the framing effect: the impact of context cues on solutions to the "Asian disease" problem. Eur. J. Soc. Psychol. **28**(2), 287–291 (1998)
8. Bourgeois-Gironde, S., Giraud, R.: Framing effects as violations of extensionality. Theory Decis. **67**(4), 385–404 (2009)
9. Lewin, K.: Field theory and experiment in social psychology. Am. J. Sociol. **44**(6), 868–896 (1939)
10. Vroom, V.H.: Work and Motivation. Wiley, New York (1964)
11. Atkinson, J.W.: Motivational determinants of risk-taking behaviors. Psychol. Rev. **64**, 359–372 (1957)

12. Bedny, G., Meister, D.: The Russian theory of activity: current application to design and learning. Lawrence Erlbaum Associates, Mahwah (1997)
13. Bedny, G.: Application of Systemic-Structural Activity Theory to Design and Training. CRC Press, Boca Raton (2015)
14. Bedny, G., Karwowski, W., Bedny, I.: Applying Systemic-Structural Activity Theory to Design of Human-Computer Interaction Systems. CRC Press, Boca Raton (2015)
15. Kotik, M.A.: Psychology and Safety, 3rd edn. Valgus, Tallinn (1989)
16. Kotik, M.A., Yemelyanov, A.M.: A Method for Evaluating the Significance-as-Anxiety of Information. Studies in Artificial Intelligence. Scientific Notes of Tartu University, vol. 654, pp. 111–129. Tartu University Press, Tartu (1983)
17. Kotik, M.A.: A Method for Evaluating the Significance-as-Value of Information. Studies in Artificial Intelligence. Scientific Notes of Tartu University, vol. 688, pp. 86–102. Tartu University Press, Tartu (1984)
18. Kotik, M.A.: Developing applications of "field theory" in mass studies. J. Russ. East Eur. Psychol. **2**(4), 38–52 (1994)
19. Kotik, M.A., Yemelyanov, A.M.: Emotions as an indicator of subjective preferences in decision making. Psychol. J. **13**(1), 118–125 (1992)
20. Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River: US Airways Flight 1549. Accident Report NTSB/AAR-10/03, PB2010-910403, May 4 2010

Applying SSAT in Computer-Based Analysis and Investigation of Operator Errors

Alexander Yemelyanov(✉) and Alla Yemelyanov

Department of Computer Science, Georgia Southwestern State University,
800 Georgia Southwestern State University Drive,
Americus, GA 31709, USA
{Alexander.Yemelyanov,Alla.Yemelyanov}@gsw.edu

Abstract. Through the vision of the systemic-structural activity theory, a method of modeling human operator performance in complex systems, realized as a DSS, is presented. It allows to conduct a detailed analysis of an operator's performance to observe the nature of his errors in different stages of his control action: the perception of a problem, the motivation for solving it, the evaluation of variants of a solution; then, making a decision and subsequently implementing it. This method provides a deeper analysis of underlying factors using the general model of activity self-regulation, along with the proposed frame descriptions for errors with logical, decision-making, and classification algorithms. The specific feature of the suggested approach is that it presents the opportunity to provide a comprehensive functional analysis of erroneous actions and their underlying factors in the process of collecting data on them, instead of following the traditional approach of drawing conclusions from investigation reports.

Keywords: Human operator · Error analysis · Systemic-structural activity theory · Modeling · Algorithms · Decision support system

1 Introduction

Human errors are an unavoidable component of any human performance. They are most explored for human-operator activity in complex systems such as aviation, nuclear power, military and hospital systems, etc., in which they are associated with severe consequences. Without the deduction of the root causes of operators' erroneous actions, it is ultimately impossible to create a reliable and safely functioning human-machine systems.

Error taxonomy plays an important role in error analysis. For example, in the aviation domain, Wiegmann and Shappell [2] proposed a taxonomy – Human Factors Analysis and Classification System (HFACS) – to apply to the investigations of human errors in aircraft accidents. Specifically, HFACS describes four levels of failure, unsafe acts, which present active errors, and preconditions for unsafe acts, unsafe supervision, and organizational influence, which present latent conditions. Each of these components has been classified by its own taxonomy. For example, preconditions for unsafe acts include physical and technological environmental factors, personal factors (crew recourse management and personal readiness), and different conditions of operations (adverse mental

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_31

and psychological states, physical and mental limitations); unsafe acts include two categories of errors (failures): errors (skill-based, decision, perceptual) and violations (routine, exceptional). It should be emphasized that whereas the characteristics of latent conditions of these taxonomies are enough precise, the characteristics of active errors remain very general and their underlying cognitive sources are relatively poorly understood. Moreover, depending on the scientific school, active error characteristics such as unsafe act, error, and violation can be interpreted differently [3]. The sources of information for these types of taxonomies are mostly safety (accident) reports. For example, the system HFACS used over 300 naval-aviation accident reports from the US Naval Safety Center [2]. At the same time, according to ICAO, in aviation, accident reports regularly do not really contain "causes on which safety recommendations can be made" (why an accident occurred), but rather, merely "brief descriptions of the accident" (what caused it) [4]. So, while HFACS successfully classifies the existing causal factors contained within NTSB and other accident reports, it is fairly limited in the field during an actual accident investigation.

The model for operator performance, which will be described below, includes characteristics for both active errors and latent conditions and, along with the existing taxonomies, facilitates the determining of those causes on which safety recommendations can be made. This model will be presented without those details which are non-essential for understanding its main idea.

2 Preliminary

As the unit of analysis of an operator's errors, a *control action* is chosen – a conscious goal-directed social act. Analysis of the error that was committed begins from the moment of inception of an operator's problem (regardless of whether he was able to perceive it) because all stages of a control action are directed towards finding a solution for it. The *moment of concluding an analysis* must be considered as the occurrence of the discussed consequences of the error. The *operator* himself must also be present in the description of the model, as he is the one who committed the erroneous action and influenced the affected object, as well as the *means* with which this action can occur, and the *material* and *close social environment* in which he exists and acts. Also highlighted in the model are the influences on all its components of a *higher-level social environment*. It is worth paying particular attention to the importance in the model of such a factor as *society* (naturally, a specific aspect of it). It is exactly this factor that gave rise to the given human-machine system and developed the rules and norms of system implementation, which predetermine human-operator activity, and that those errors which we analyze turn out to be nothing more than a violation of these norms. For this reason, the model must demonstrate an entire complex of various *norms* – technical, juridical, social, – which function to determine human-operator activity in the control system. Without the consideration of these norms, it is essentially impossible to discuss the errors performed by the operator. From this, it follows that the performance (activity) of the human-operator, relative to the norms that restrict his actions in the control system, is the primary indicator of this model. This is why the influences of the system's substructures (operator, control system and object, and their environment) are realistically described by a language indicating a relationship between the interactions of the established norms – and if they are violated, this language rationally describes the levels of deviation from these norms. In the given model, aimed at determining the causes of an operator's errors, all the main stages of a control action must also be considered: *the perception of a problem, the motivation for solving it*, and *the evaluation of variants of a solution*, followed by *making a decision* and subsequently *implementing it* and the *obtained results*, in which there could lie the causes of the error. During the analysis of the operator's behavior at each one of these stages, it is necessary to consider his *obligations*, his *awareness of these obligations*, *the ability to foresee* the changes in deviation from these norms in the interactions (influences) of various system substructures, and his *desire to avoid/diminish* these deviations. Thus, the model must contain means by which the operator's obligations, abilities, foresights, and desires could be anticipated. It is apparent that during the analysis of the operator's errors, it is insufficient to use only realistically existing situations in which he has acted; in this analysis, it is also necessary to include evidence of subjective reflections of the operator regarding these situations, transformed by the *meaning* and *significance* that he perceived in them. In this way, at each stage of solving the task and realizing the control actions, the model must consider not only the *objective situation* at hand, but also its *subjective perception* in the consciousness of the operator, in order to gain a clear insight on his errors from these interconnected viewpoints. Finally, the model must reflect the vital fact that humans are *goal-directed self-regulative systems*.

3 Model of Operator Error

In the model, the human operator error was considered as the commission or omission of human action that lead the controllable parameters of the system outside of its acceptable limits, or is prohibited by rules of the system [5–7]. This is one of the most general definitions for an error, which is not necessarily associated with guilt, consequences of the error, or the presence or absence of intention, and corresponds to the Reason's "unsafe act" [8] and to the HRA erroneous action.

The modeling language for operator performance is similar to a unified modeling language, which is widely used in the field of human-computer interaction engineering [9]. It includes the following building blocks [3]:

1. Three types of diagrams: *action diagram,* which shows the stages of operator performance arranged in a time sequence; *inception diagram,* which shows the basic schematic behavior of a control system and can be extended or displayed in details; and an *elaboration diagram,* which is built by adjusting the inception diagram to an actual operator's performance: pilot, controller, etc.
2. Relationships: interaction, aggregation, realization, evaluation, etc.
3. Specifications for relationships (provide cognitive backplane of the model): extent and likelihood of a deviation, operator's intent to prevent the deviation, operator's obligation to predict the deviation, operator's ability to predict the deviation, operator's awareness of his abilities and obligations.
4. Classifiers for different types of errors, violations, risk behavior, working conditions, information supplying, etc.

An action diagram provides the description of an operator‘s performance (his control action) and includes the stages of: perception of a problem, motivation for solving it, evaluation of variants of a solution, making a decision, its implementation, and obtained results [10]. An inception diagram presents each of the listed stages as a frame description while the entire operator’s performance is presented as a sequence of these descriptions.

The operator’s personality I is presented by “the material self” (I_{ph}), “the spiritual self” (I_{sp}), and “the social self” (I_{sc}), that are driven through interactions with the external environment, The external environment for the operator includes: the controlled object; means of control; material environment for operator, the controlled object and means of control; operator’s close social environment; and, higher-level social environment.

All of these components with corresponding interactions present a *frame of action* – a basic cognitive structure that guides the perception and representation of the operator’s performance [11]. The interactions allow to display different features of erroneous action. The specified substructures (slots) of the frame are represented by circles and the mutual interactions between the frame substructures which appear in the operator’s performance are marked by arrows.

The operator implements his decision by affecting a controllable object, either directly or by means of control. The frame shows the reverse effects of the external substructures at the operator and their mutual interactions.

Figure 1 presents the elaboration diagram, which was built by adjusting the general inception diagram to the pilot’s actual performance. As shown in this figure, the pilot (I) implements his decision via the control systems – (I, C_1), (I, C_2), and (I, C_3), and also via other members of the aircrew and the passengers – (I, P). The reverse effect for the pilot is provided by the aircraft and its functional systems – (E, I_{ph}), flight regulation documents – (D, I_{sp}), other members of the crew and the passengers – (P, I_{sp}), the reference group – (U, I_{sp}), and the flight supervisor – (S, I_{sc}).

The aircraft and its control systems – (D, E), as well as the crew members and the passengers, operate according to flight regulation documents – (D, P) and are exposed to outside environmental effects (N, E).

All interactions presented in the elaboration diagram for the pilot’s performance are subdivided into three groups:

- $W_1 = \{(I, C_1), (I, C_2), (I, C_3), (C_1, E_1), (C_2, E_2), (C_3, E_3), (I, P)\}-$ for the pilot’s control actions (presented by deviations);
- $W_2 = \{(E, I_{ph}), (CE, CE), (N, E), (E, P), (P, I_{sp}), (P, P), (D, E), (D, I_{sp}), (D, P)\}-$ for the state of the controllable system (presented by deviations);
- $W_3 = \{(S, I_{sc}), (U, I_{sp})\}-$ for social assessment (presented by absolute values, “positive” or “negative”).

Since norms are essential grounds for the functioning of control systems, the degree of interaction for the first and second groups are expressed in their deviations (“negative” interactions) from the established norms. The degree for social evaluation in the third group of interactions is expressed by their absolute values (“negative” or “positive”).

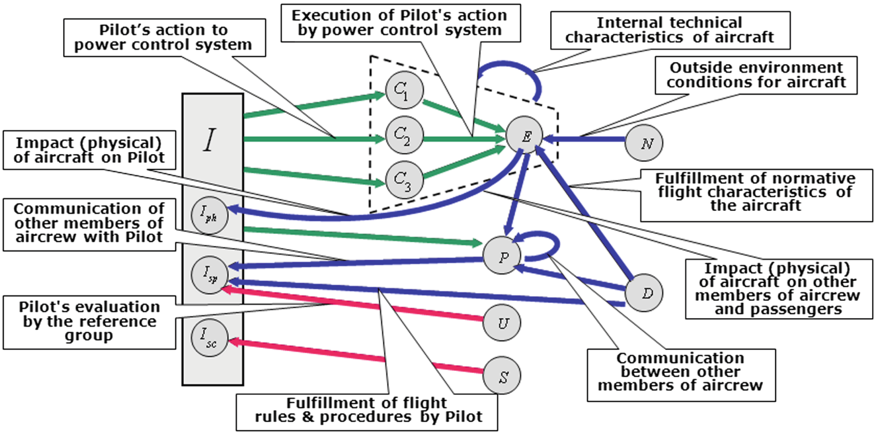


Fig. 1. Elaboration diagram for the pilot's performance (sample schematic for demonstration purposes only). *I* – pilot; *E* – aircraft and its functional systems; C_1 – flight control system; C_2 – power control system; C_3 – life support system; *N* – outside environment of aircraft (terrain, weather, hazards, etc.); *D* – flight regulation documents; *P* – other members of the aircrew, passengers; *U* – pilot's reference group (family, colleagues, etc.); *S* – flight supervisor (controller)

The deviations and absolute values are determined on a soft (vague) verbal scale of intensity ("weak"–"strong").

Besides the intensity of effects, the frame descriptions also contain a number of other *parameters*. These parameters characterize the likelihood for the occurrence of various effects and the operator's estimates regarding these likelihood ("seldom" – "often"); his/her intent to avoid undesirable negative consequences ("desire" – "no desire") as well as obligations ("oblige" – "not oblige") and abilities ("able" – "unable") to predict those; and the operator's awareness of his abilities and obligations ("believe" – "reject"). The logical and causal properties of these parameters, were used for creating logical rules and classifiers for errors. Specifically, to provide logical completeness of each frame description, to verify the logical consistence of all the frames integrated into the action diagram, and to classify them, a special mathematical apparatus was constructed [3]. The language of this logic turned out to be relatively productive in the description of operator actions and the classification of many vague types of operator behavior, such as risky, intentional, negligent, overconfident, etc. To provide a complete analysis of all interactions in the frame description through the vision of their preferences to the operator and the decision which he finally made, besides the logical classifiers, the experimental classifiers were created. These classifiers use the significance model [12] and can measure the levels of negative and positive significances of different interactions with their characteristics of *intensity* and *likelihood*. In other words, they are presented to answer questions such as: "Is the risk of getting injured more significant or less significant to the pilot (at the moment of decision-making) than the risk of not performing the task and therefore losing the respect of his colleagues?"

4 Algorithms for Error Analysis

To provide a comprehensive analysis of the operator's erroneous action, the different types of algorithms were developed: *elementary algorithms* for the analysis of particular interactions in the frame description; more general *algorithms for the analysis of the single stages of the action;* and the *general algorithm*, serving the purpose of analysis of the erroneous action as a whole unit [13].

For example, one of the elementary algorithms determines how the existing underlying factors influence the erroneous action (Fig. 2). It uses six performance shaping factors (PSFs) comprised of *poor information supplying*, provided by team members (factor F_1), audio (factor F_2) and video (factor F_3) devices, and *inappropriate working conditions*, consisting of external conditions – factor F_4 (load, climatic conditions, ergonomic factors, etc.) and internal conditions – factor F_5 (training, operator conditions: fatigue, hypoxia, medical illness, etc.), and factor F_6 – operator's means of control, that range their influence from "zero" to "high" on the scale: {"zero", "low", "not low-not high", "high"}. The influence of the PSF is not rated in relation to the whole action, but to the basic interaction (system parameter) of a considered stage of action (i.e. perception of a problem). For this basic interaction, the influence of the PSF is applied toward the ability of the operator, for instance, to correctly predict some level of deviation of this parameter from the existing norms. The algorithm, based on these influences, uses the classification rules to determine on the scale {"able," "somewhat able," "somewhat unable," "unable"} the operator's ability to correctly estimate the possibility of the violation.

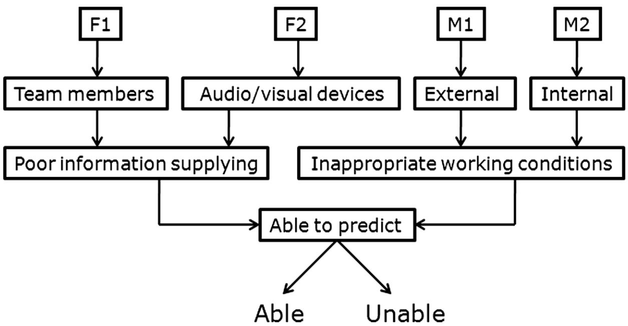


Fig. 2. Algorithm for determining of the underlying factors.

Another elementary algorithm (Fig. 3) determines the type of an operator's responsibility for the different outcomes of his erroneous action. This algorithm, when used at the decision-making stage, determines if the operator caused this interaction (outcome) intentionally, with overconfidence, or with negligence, or if it happened due to some other underlying factors. The error with indirect intention is a specific type of error that occurs when the operator makes a decision, anticipates its consequences, but then has no intention to avoid them. It may occur, for example, in a situation in which the operator recognizes the risk of a more severe consequence for his action and tries to avoid it first. In other words, the operator follows a principle of choosing "the lesser of two evils". If the operator did not anticipate the consequence of his action, it could either be his responsibility, or it could be caused by some underlying factors. The operator's responsibility is presented in the form of overconfidence, when the operator was aware of his obligation to anticipate the consequence, or in the form of negligence, when the operator was not aware, but was obliged and able to predict this consequence. However, if the operator anticipated the consequence and decided to avoid it, his responsibility can be determined by further analysis of the implementation of this action.

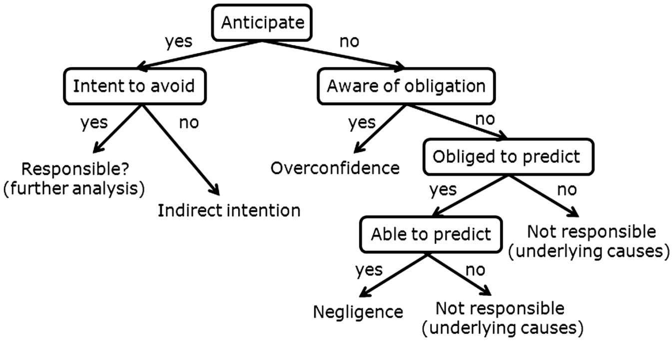


Fig. 3. Algorithm for determining the operator's responsibility.

Algorithms for stages, employing elementary algorithms for every stage of an operator's erroneous action, analyze all its interactions in the context of their direction, sign, and intensity.

For example, the algorithm for stage 1 (perception of a problem) functions by considering such mechanisms of self-regulation as meaning (objective significance) and sense (subjective significance) [1] for each deviation, verifying whether its subjective significance corresponds to the objective (i.e. if the operator correctly estimated the deviations) and, if not, by examining his information-supplying (*factors* F_1, F_2, *and* F_3) and working conditions (*factors* F_4 *and* F_5) to determine whether the operator was able to correctly provide this estimation. The algorithm for stage 2 (motivation for solving), for each deviation anticipated by the operator, determines if the operator was motivated to avoid this deviation; if not, it examines his working conditions (*factors* F_4 *and* F_5) to determine whether the operator was able to have the proper motivation to do so. This algorithm implements the inducing mechanism of the functional analysis of activity self-regulation [1].

The general algorithm, using the general model of activity self-regulation [1] provides an analysis of the erroneous action as a whole unit. It starts from the stage of obtained results and sequentially analyzes all stages, from the perception of a problem to the decision implementation. The analysis of each stage can contribute to the final conclusion regarding what the main cause of the operator error and its underlying factors were. The algorithm can stop and make a conclusion at any time, once the operator's error and its causes have already been recognized. It may happen when the perception of the problem was completely inadequate (the operator did not anticipate all expected deviations), when the operator had no motivation to resolve the problem (the operator did not try to avoid any deviations of those that he anticipated), when the operator's solution was not the best (or at least acceptable), or when making a decision, the operator intentionally caused the resulting deviation.

These algorithms, along with other algorithms, are used in the decision support system SAFE to logically analyze errors and their influencing factors in the process of data collection during error investigation.

5 Implementation

The suggested model of operator performance was implemented in project SAFE (System for Analyzing and Forecasting Errors) [3]. The purpose of this project was to provide a powerful decision support tool for investigating pilot and air traffic controller errors, analyzing their underlying causes, and making recommendations. SAFE provides support for two types of users: investigators and system analysts (Fig. 4).

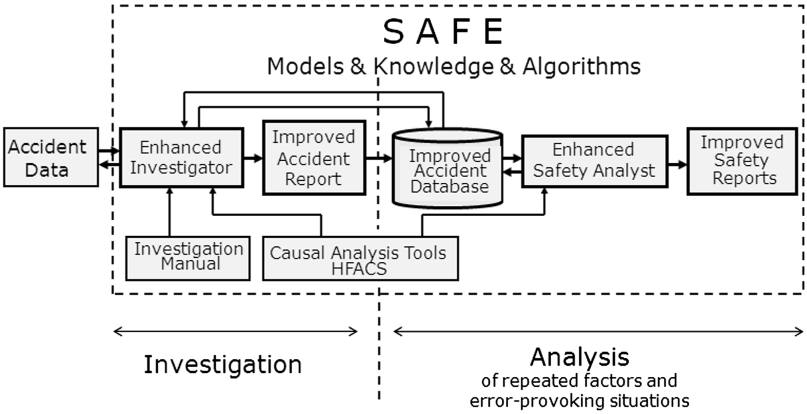


Fig. 4. Decision support in error analysis and investigation

It guides the investigator in the gathering and analyzing of the active error characteristics and related underlying factors (latent conditions) in the process of error investigation.

Using the embedded models and computational algorithms, SAFE provides a sufficient completeness and correspondence of all the collected data, according to existing regulations and taxonomies, and checks that important information is not to be ignored either unintentionally or deliberately.

This support should enhance the investigator's awareness and ensure the necessary qualitative level of the whole process of investigation. The core of SAFE is its database, which helps in error investigation and in producing recommendations, as well as making a large amount of statistical data available and helping in further analysis provided by the safety analyst, who determines repeated factors and error-provoking situations. All of this improves the function of both the investigator and safety analyst, which, in turn, improves the safety of the whole control system.

References

1. Bedny, G., Karwowski, W., Bedny, I.: Applying Systemic-Structural Activity Theory to Design of Human-Computer Interaction Systems. CRC Press, Boca Raton (2015)
2. Wiegmann, D.A., Shapell, S.A.: A Human Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System. Ashgate Publishing, Aldershot (2003)
3. Yemelyanov, A.M.: Modeling and decision support for analysis of operator errors in complex systems. J. Comput. (JCP) **4**(4), 319–329 (2009). Academy Publisher
4. International Civil Aviation Organization: Investigation of human factors in accidents and incidents. Human factors digest, no. 7, Montreal, Canada (1993)
5. Kotik, M.A., Yemelyanov, A.M.: Errors in control action. In: Psychological Causes, Method of Automated Analysis. Valgus, Estonia, Tallinn (1985)
6. Kotik, M.A., Yemelyanov, A.M.: The Origin of Human-Operator Errors. Transport Publishers, Moscow (1993)
7. Miller, D., Swain, A.: Human error and human reliability. In: Salvendy, G. (ed.) Handbook of Human Factors. Wiley, New York (1986)
8. Reason, J.T.: Human Error. Cambridge University Press, New York (1990)
9. Booch, G., Rumbaugh, J., Jacobson, I.: The Unified Modeling Language User Guide. Addison-Wesley, Boston (1999)
10. Norman, D.A.: Cognitive engineering. In: Norman, D., Draper, S. (eds.) User Centered System Design: New Perspectives on Human-Computer Interaction, pp. 31–62. Erlbaum, Hillsdale (1986)
11. Minsky, M.: A framework for representing knowledge. In: Winston, P. (ed.) The Psychology of Computer Vision, pp. 211–277. McGraw-Hill, New York (1975)
12. Kotik, M.A.: Developing applications of "field theory" in mass studies. J. Russ. East Eur. Psychol. **2**(4), 38–52 (1994)
13. Yemelyanov, A.: Analysis of influencing factors in the model of erroneous action: computer-based method and algorithms. In: 2011 IEEE International Conference on Systems, Man, and Cybernetics, pp. 2107–2112, Anchorage, Alaska, 9–12 October (2011)

Mediating Subjective Task Complexity in Job Design: A Critical Reflection of Historicity in Self-regulatory Activity

Mohammed-Aminu Sanda[1,2](✉)

[1] University of Ghana Business School, P.O. Box LG 78, Legon, Accra, Ghana
masanda@ug.edu.gh, mohami@ltu.se
[2] Luleå University of Technology, 97187 Luleå, Sweden

Abstract. This paper critically reflects on the influence of historicity in self-regulatory activity towards mediating subjective task complexity in job design. This is based on the growing interest of using the 'practice' approach in overcoming the gap between the theoretical understanding of what people do and the realistic understanding of what people actually do at the workplace. The paper argued that since an objective analysis of the number of alternatives presented in any given situation will not always coincide with subjective perceptions, and the individual's lack of knowledge about the external world may result in his/her inability to accurately predict the outcomes of his/her decisions, then it is important to understand how an individual acquire knowledge about his/her external world of work, and also if its consequential effect on the routine is acquired through the recollections of the subconscious mind's daily encounters at the work situations.

Keywords: Mediation · Historicity · Subjective task · Task complexity · Self-regulation activity · Job design

1 Introduction

There is a growing interest among both management researchers and practitioners in using the 'practice' approach in management literature to overcome concerns on the gap between the theoretical understanding of what people do at the workplace, and the realistic understanding of what people actually do at the workplace [1, 2]. The focus of the practice approach in organizations, as highlighted by [1], has been on the way that actors (or workers) interact with the social and physical features of context in the everyday activities that constitute their organizational practice. As such, the term 'practice' is used here to imply the repetitive performance of activity undertakers at the workplace, and which activities leads them to attain recurrent, habitual or routinized accomplishment of particular actions. Arguing from the perspective of [1], practices embedded in organizational activities can be viewed as being influenced more by their economic and social histories (such as, institutional norms and technological choices) than in their rationalities. This implies that practices in organizational activities occur both in the macro contexts that provide commonalities of action, and the micro contexts

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_32

in which action is highly localized. The interaction between these contexts, as noted by [1], provides an opportunity for adaptive practice. Thus in the approach towards identifying factors that influence practices evolving from organizational activities in organizations, it is important to understand the practices that are entrenched in, and also pervades the workplace environment. This is because practice has to be seen as arising from the continual interactions among individual workers, as well as between workers and the workplace's internal environment.

According to [3], the complexity of the human work process is highlighted by the characteristics of its substructure. The basic components of such substructure have been categorized by [4] to include, motive-goal as a vector which demonstrates the directional and energetic aspects of the work activity, knowledge and skills which demonstrate the relevance past experience to the work process, abilities related to the tasks to be performed, as well as work actions (both cognitive and motor) which are organized into a structure representing method of work. As such, the structure of a work process, as argued by [4], is rooted in the concepts of knowledge and action, and by extension, the application of mental tools [4]. Therefore, in developing new organizational activities in an organization, the key actors involved in the practice development exercise must be seen as important learners of their newly developed work practices [3]. This is because, in the task interpretation process, workers will likely incorporate their personal prerequisites, such as experience, skills, and physical constitution in the workplace's social system [3]. There is also the likelihood that workers will engage in solution-seeking activities to problems that emerge during task undertakings, a situation that can lead to task uncertainty and its associated consequence on the worker undertaking the task.

It has been explained by [4] that the uncertainty in a task is dependent on both its objective and subjective characteristics. According to [4], while the number of possible alternatives available in any task is an example of an objective characteristic of uncertainty, an objective analysis of the number of alternative presented in any given situation will not always coincide with the individual's subjective perceptions. This is because, an individual's lack of knowledge about the external world may result in his/her being unable to accurately predict environmental events or the outcomes of his decisions [2, 4].

Arguing from the perspectives of [2, 5], the complexity of a worker's activity can be viewed as the number of rules inherent in the cognitive and motivational-emotional attribution of the activity. Thus by integrating both the cognitive and motivational-emotional aspects of an activity in the activity design process, a worker's physical activity can be successfully optimized. Such integration, as noted by [2] is useful to designers of activities in organizations and the required technological tools for the activity performances. This is because, workers use their generic skills, such as flexibility, technical intelligence, perceptive ability, technical sensibility, a sense of responsibility, trustworthiness, and independence [6] to mediate complexities arising from the cognitive and motivational-emotional attribution of their assigned activities [2]. Therefore, the specificity of a worker's memory workload [4] must be viewed as an important source of task complexity that needs to be understood and factored in organizational activity design. This paper therefore, argues for the need to understand ways of mediating constraints in the self-regulation system, in terms of continual reconsideration of activity strategies informed by changes in internal and external conditions or situations, since it sometimes results, not only in changes in the methods of achieving the goal, but a change in the goal itself. This is underlined by the notion [4] that self-regulation is an intrinsic self-organizing tendency of organizational activity whose underlying goal and evaluation criteria may change during an individual self-regulation process.

The purpose of this paper is to critically reflect on the influence of historicity in self-regulatory activity towards mediating subjective task complexity in job design. This is based on the realization in recent times that, there has been a growing interest among both management researchers and practitioners in using the 'practice' approach in management literature to overcome concerns on the gap between the theoretical understanding of what people do at the workplace, and the realistic understanding of what people actually do at the workplace.

2 Notion of Self-regulation

The very survival of organizations, as noted by [7], depends on their ability to adapt to their external environment. From the perspectives of science, and in terms of open systems theory, organizations require monitoring and feedback mechanisms to be able to follow and sense changes in their relevant task environments and a capacity to make responsive adjustments. A science, according to [8] can only base itself on something real and empirically given. But the existent thing must be captured as a concept because it is the starting point both for a real development and for development of understanding. Context, [8] further noted, is potentially an open-ended infinity of social, physical, cultural and historical circumstance. In a more general sense, [8] argues for the need to determine a concept of context which captures the teleological content of a person's action, just as the understanding of social situation of development entailed forming a concept of the situation, which captures the way in which the situation determines social interactions and psychological development.

Arguing from organizational perspective, the concept of self-regulation can be viewed as providing context for capturing the teleological content of a worker's action in an organizational activity. The major problem facing the self-regulation system, as identified by [9], is the process of continuing reconsideration of activity strategies when internal and external conditions or situations have changed. The bane of this problem, as highlighted by [9], is that it sometimes leads to changes in the goal as well as the methods of achieving the goal. The implication of this observation, from organizational activity theoretical perspective, is that self-regulation is a goal-directed process, with the goal entailing integrative systemic functions which is meaningful only when it is presented and defined in terms of functional mechanisms [9], as shown in Fig. 1 below. In this regard, it is important to take into account [4] the observation that the program of self-regulation, its criteria of evaluation, and even its goal may be changed during any individual self-regulation process. This is because self-regulation functions in time, and is based on the evaluation past experience, as well as present situation and anticipated future scenarios [4].

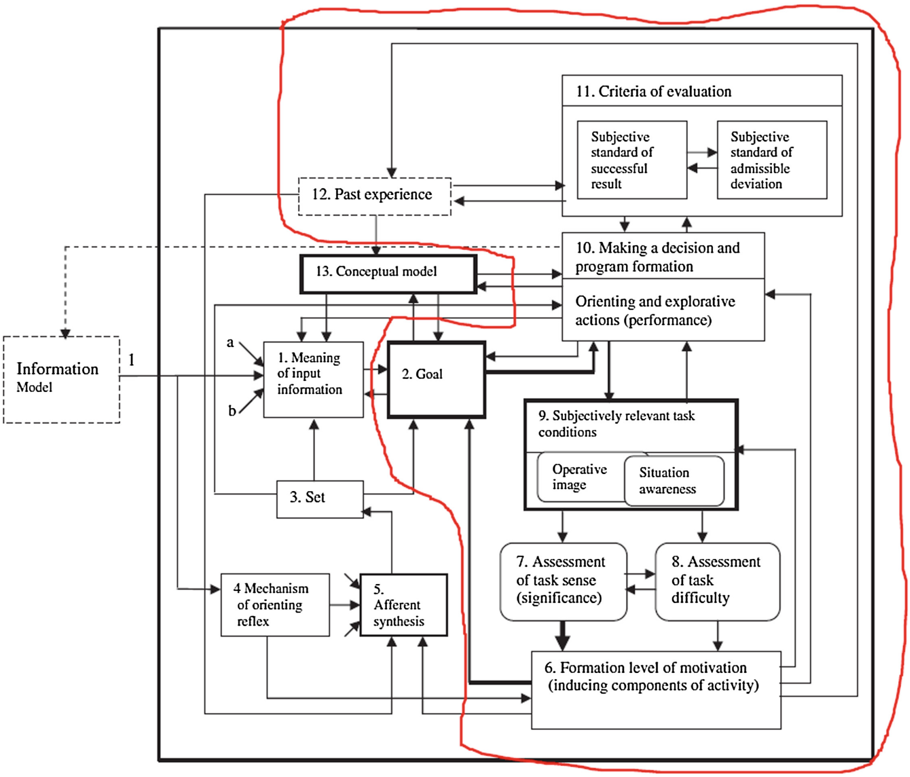


Fig. 1. The model of the self-regulation of orienting activity [9].

A reflection of the model shown in Fig. 1 above depicts it as characterizing three key elements. Firstly, all the function blocks are interrelated. Secondly, each block represents a functional subsystem directed towards achieving specific sub-goals of the activity. Thirdly, each function block is part of a flow of activities with multiple of entry and exits. Generally, the model describes how a person, in the conduct of an activity, creates a goal, a subjective mental model of the activity, the type of exploratory actions and operations utilized, the types of possible mental models developed, and how preferable mental models are selected, among others. Using [9] argument as a point of departure, the model manifest that the outcome of a worker's orienting self-regulation activity is informed by the formation of the worker's mental representation of the activity prior to its execution. This, as [9] explains, is because the cognitive processes which inform the worker's mental representation of the activity are integrated, and it is such integration that facilitates the formation of functional blocks or mechanisms of self-regulation required for achieving a specific purpose of organizational activity. Even though a number of function blocks in a self-regulation model are constant, the context of these blocks changes constantly [9]. The question here then becomes:

What are these contexts and why do they undergo such constant changes?

In answering this question, it is important to reflect on the type-classification of self-regulation processes. Two interconnected self-regulation processes classified by [9], namely, physiological and psychological processes with each having its own dynamics can be associated with the conduct of a worker's organizational activity. In the physiological self-regulation system, the tendency for reduction of physiological disturbances prevails with to the need to correct departures from optimal states in order to restore balance [4]. As [4] has noted, the structure of physiological self-regulation processes is wholly predetermined with many physiological imbalances corrected automatically. The psychological self-regulation system, which is posited by [4] as a goal-directed self-regulative process, is deemed as capable of changing its own structure based on its experience, as well as forming its own goals, sub-goals and own criteria for activity evaluation. The psychological self-regulation process entails the integration of cognitive, executive, evaluative and emotional aspects of activity [4]. Given the complex set of variables involved, a worker will be expected to exhibit an infinite diversity of actions in the performance of an organizational activity. In this regard, as noted by [4], internal changes in the psychological aspects of a worker's self-regulation may not emerge from experiences only, but also from memories that can be adapted to a situation. Such memory epitomizes historical recollections of adaptable practices that are manifestations of declarative knowledge.

3 History-Making in Task Performance

According to [10], since history is made in future-oriented situated actions, the challenge is to make situated history-making visible and analyzable. This implies that for studies of organizational activity, it is important to look for ways of capturing how workers discursively create new forms of activity and organization. Using [11] perspectives as point of departure, it is imperative that in a worker's performance of organizational activity in an organization, the worker's motive of the activity can be divided into two, namely, sense-formative motive and situational motive. Using [11]'s explanation, the sense-formative motive, which is connected with individual features of personality, is a relatively stable element that help determine the worker's general motivational direction, while the situational motive, which is more flexible and more involved with task performance, is connected with the worker's progressing activity and the solving of specific tasks. But unlike the sense-formative motive, the relative weight of the situational motive can change in an organizational activity, depending on the activity's character. Therefore, bids for the simultaneous realization of the two motives in an organizational activity, coupled with the uncertainty associated with such activity brings to the fore the issue of task complexity.

Uncertainty in a task, as noted by [4], depends on both the objective and subjective characteristics of tasks entailed in an organizational activity. Arguing from the perspectives of [4], the complexity that arises from activity uncertainty is deemed as an objective characteristic of the activity. Likewise, the difficulty that comes with the activity uncertainty can be ascribed as the worker's subjective evaluation of the effects of the activity's complexity. Thus depending on the skills and the individual features of the worker, the same complex task will be evaluated by him/her as relatively more or less difficult [4]. As such, when the complexity of an organizational activity increases, the probability of a worker exerting more cognitive effort with an increase in motivational mobilization will also increase [4]. But as [4] has clarified, complexity itself does not have a subjective component. As such, a worker cannot experience it directly, but can only perceive a subjective difficulty. The question here then becomes:

How does this individual acquire knowledge about his/her external world of work? Is it through action by doing of his/her everyday routinized activity for which experience and its consequential effect on the routine is acquired through the historical recollections of the subconscious mind's daily encounters at the work situations?

In answering the above question, it is important to understand the need for an organizational activity to be analyzed against the history of its local organization and against the more global history of the organizational concepts, procedures and tools employed and accumulated in the local activity [12]. This brings to the fore the relevance of [12] third principle of activity which is on historicity. It postulates that activity systems take shape and get transformed over lengthy periods of time. The problems and potentials of an organizational activity system can only be understood against their own history. According to [12], history itself needs to be studied as local history of the activity and its objects, and as history of the theoretical ideas and tools that have shaped the activity. History-making, according to [10], is embedded in everyday organizational activities. This postulation by [10] is based on the argument by [13] that history-making is based on the skill of uncovering the tension between standard, commonsense practices and what an individual actually does. According to [10, 13] has suggested three basic ways for resolving such tensions, namely articulation, cross-appropriation, and reconfiguration [10]. In articulation, the basic pattern of the organizational activity is not changed, but important practices or values that have become vague, confused or lost are recovered and a new coherence is thus achieved [10]. In cross-appropriation, which typically manifests itself in interpretive speaking and personal narratives, practices, ideas or tools are taken over from other activities or social worlds. This may or may not change the whole pattern of the organizational activity. In reconfiguration, which notion comes close to the concept of expansion elaborated in activity theory [14] and requires constant awareness of anomalies, a marginal aspect of the organizational activity becomes dominant and the entire pattern is radically transformed. As found by [15], in overcome the gap between action and imagination in history-making, it may be necessary to bring closer to one another, and occasionally merge articulative decision-making and reconfigurative modeling [10].

4 Mediation in Organizational Activity System

Activity theory, as explained by [16], was founded upon the premise that psychological development is a social process arising from an individual's interactions within particular historical and cultural contexts. This interaction is perceived to provide an interpretative basis from which individuals attribute meaning to their own actions as well as that of others. Thus an action and the way it is perceived are profoundly interrelated within an activity system that provides shared meanings. As [1] explains, meaning is shared through a process of mediation between actors (workers in an organization), their socio-cultural contexts, and the actions undertaken. The mediation can occur through the use of many different types of tools, material tools as well as mental tools, including culture, ways of thinking and language [16]. However, mediation may also occur through practical activity, from which perspective interpretative context is externalized from the individual, assuming a collective reality based in practice [17–20]. As such, transforming the objective of an organizational activity into an outcome motivates the existence of the practical activity. In this context, practical activity reflect the daily work in which actors engage, and that it is external to any single actor providing a context of shared meaning. [21] has also pointed out that while the elements of agents (human resources) and mediational means are always involved in mediated actions, one of them may take on special importance in particular cases, such as the design of organizational activities in organizations.

It has been established by [2] that in the design processes of organizational activities, it is important to identify and distinguish the complexities associated with the activity's cognitive attributions informed by the specificity of information processing, and its emotional-motivational attribution reflected by the energetic aspects of the activity. This, according to [2], will enable designers of organizational activities and associated technologies understand the performance enhancing strategies used by workers to mediate the cognitive difficulties and the emotional-motivational challenges inherent in their designed and assigned organizational activities (see Fig. 2 below). The transformations that a mediated action undergoes often involve changes in this mix

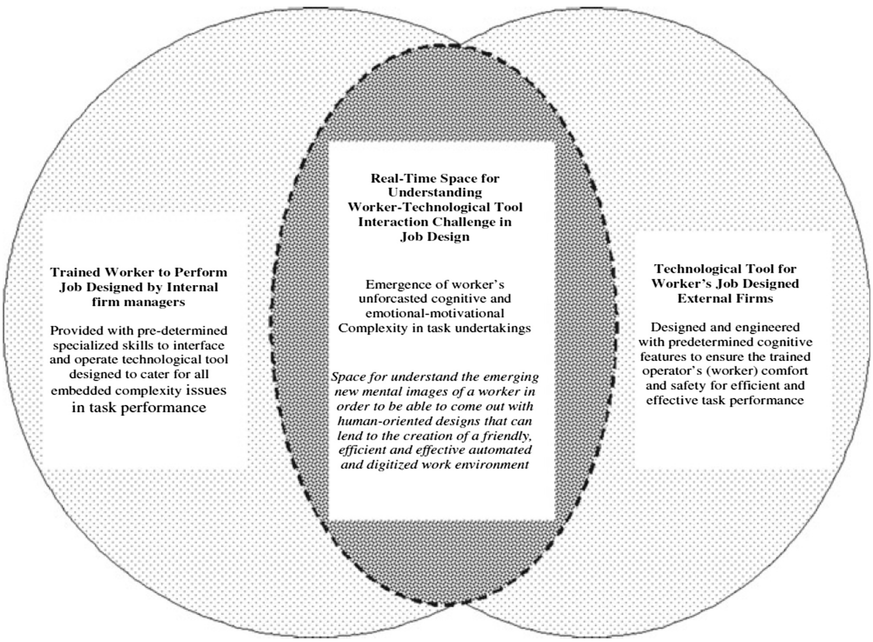


Fig. 2. Model for understanding a worker's emerging cognitive and emotional-motivational aspect of task complexity [2]

rather than the introduction of a new element. Hence, the understanding of such dialectical complexities requires the separation and comparison of elements. This therefore, calls for the relative need of giving consideration to the existence of multiple purposes and goals in mediated actions in an organization's activity system. These multiple purposes or goals of mediated action are often in conflict. What this means is that, in most cases, mediated action cannot be adequately interpreted upon the assumption that it is organized around a single, neatly identifiable goal. Instead, multiple goals, often in interaction and sometimes in conflict, are typically involved.

Thus the claim that mediated action is situated in one or more developmental paths is an elaboration of the general assertion that mediated action is historically situated. The inferred meaning here is that agents, cultural tools, and the irreducible tension between them always have a particular past and are always in the process of undergoing further change. Therefore, the outcome of interaction as a practical activity is the purposive, outcome-oriented work in which actors engage. This was by virtue of the fact that interaction is enabled through the technical and psychological tools that actors use to engage with their environments [10, 17]. In an organizational activity context, the use of such tools can be viewed to be practical, and directed towards constructing outcome-oriented activity. Since these 'tools' are used to establish practical activity, they may be defined as the practices through which activity is constructed. The tools and materials do not create the practice, but rather mediate its usage and outcomes in a given context. Thus mediation is a distinctive concept in activity theory that explains how individual actors, the social collective, and their shared endeavors are integrated in the pursuit of action [10, 18, 22]. Mediation occurs through structuring mechanisms, such as role, division of labour, tools, as well as implicit and explicit rules that enable interaction between actors and context. In agreement with [22–24], these instruments of mediation can be viewed as socio-cultural artifacts imbued with meaning particular to the context in which they are used, thus enabling shared interpretation among participants. These structuring mechanisms are reflected correctly in the organizational context by [1] who considered them as the practices through which strategic activity is mediated. It has been pointed by [1] that out that mediation may be further explained by two concepts, namely 'situated' and 'distributed'. As [1] explains 'situated' indicate that interaction is embedded within and derives meaning from the historical and cultural context in which it occurs, permitting shared interpretations to arise. Thus situatedness may be applied with a wider or narrower focus in accordance with the topic of research. Situated actions are inherently tension-laden, unstable and open-ended [15], and that instead of just retrospectively asking why an action or an utterance occurred, one should also ask of what dynamics and possibilities of change and development are involved in a given action. This is from the understanding that every cluster of actions offers a window into the emerging zone of proximal development of the collective activity system in which the actions occur [14, 16].

According to [1], the term "distributed" refers to actors and also to the context and action in which they are engaged. In principle, it refers to the overlapping knowledge involved in action. Thus widely distributed interaction between the actors, structures and actions of an organization provides systemic knowledge of how to act strategically [25]. In this respect, [1] views "distribution" to be a key factor in mediation that enhances shared interpretations and collective action. Within this context, tensions are viewed to be conceptualized as a generative process of continuity and displacement between old and new players. New participants learn from continuing members how to use the practices of the system, and in the process re-socializing the continuing players and reinforcing the existing practices. However, due to their low socialization to the system, new members also question the practices, so creating the potential for their re-evaluation and adaptation [22, 26, 27]. Distribution enables collective action because individual actors have sufficient knowledge of the other immediate components of activity to conceptualize and respond to actions that are outside but connected to their own task [26, 28, 29]. As noted by [30], the process followed for successfully completing any human activity is constantly evolving. In this respect, a worker, or group of workers, given responsibility for a particular task, will naturally seek to minimize the amount of work necessary for completing the task by deriving new, more efficient methods. In this respect, the following nine basic claims outlined by [21] as characterizing mediated action (i.e. action that can occur through the use of both material and cultural tools) are of immense significance in understanding mediation in an organization's activity system;

- Mediated action is characterized by an irreducible tension between agent and mediational means. Mediational means are material;
- Mediated action typically has multiple simultaneous goals;
- Mediated action is situated on one or more developmental paths.
- Mediational means constrain as well as enable action.
- New mediational means transform mediated action.
- The relationship of agents toward mediational means can be characterized in terms of mastery.
- The relationship of agents toward mediational means can be characterized in terms of appropriation.
- Mediational means are often produced for reasons other than to facilitate mediated action.
- Mediational means are associated with power and authority.

Therefore, the essence of examining agent and cultural tools in mediated action is to examine them as they interact, and that any attempt to reduce the account of mediated action to one or the other of these elements runs the risk of destroying the phenomenon under observation. This argument is of great relevance when one takes into account the mediation that is to occur among the different agents in an organization's activity system.

5 Conclusion

By recognizing the importance of human interpretive activity to organizational life, the role of organizational as well as non-organizational influences in conditioning all such interpretive activity in the organizational arena might need to be understood. The acquisition of such broad understanding will therefore, require a prior understanding of the issues of mediation and synergism in the human interpretive activity system. Thus, building on the notion that self-regulation in activity theory is not a homeostatic, but rather a goal-directed process where the goal has integrative systemic functions, and is also not a psychological notion, but rather a cybernetic one [9], the model of goal formation and acceptance underlined by the systemic-structural theory of activity [4] must be considered as one of the major principles of activity functioning. Also, since an objective analysis of the number of alternatives presented in any given situation will not always coincide with subjective perceptions, and the individual's lack of knowledge about the external world may result in his/her inability to accurately predict the outcomes of his/her decisions [4], then it is important to know how this individual acquire knowledge about his/her external world of work. By implication, it is important to know whether the acquisition of such individual knowledge is through action by doing of his/her everyday routinized activity for which external influence and its consequential effect on the routine is acquired through the recollections of the subconscious mind's daily encounters at the work situations. Thus, based on the premise that an individual's self-regulation system takes shape and gets transformed over lengthy periods of time, with its problems and potentials being understood only against its own history, this paper situates that if the major problem facing the self-regulation system is the process of continuing reconsideration of activity strategies when internal and external conditions or situations have changed, then it is important to understand how such problem is remedied by the mediation that occurs through the individual's use of mental tools or practical activity imbued with historical meaning.

References

1. Jarzabkowski, P.: Strategy as social practice: an activity theory perspective on continuity and change. J. Manag. Stud. **40**(1), 23–55 (2003)
2. Sanda, M.A. Cognitive and emotional-motivational implications. In: Hale, K.S., Stanney, K.M. (eds.) The Job Design Of Digitized Production Drilling In Deep Mines. Advances in Neuroergonomics and Cognitive Engineering, Advances in Intelligent Systems and Computing, vol. 488, pp. 211–222. Springer, Switzerland (2016)
3. Sanda, M.A., Johansson, J., Johansson, B., Abrahamsson, L.: Using systemic structural activity approach in identifying strategies enhancing human performance in mining production drilling activity. Theor. Issues Ergon. Sci. **15**(3), 262–282 (2014)
4. Bedny, G.Z., Karwowski, W.: A Systemic-Structural Theory of Activity: Applications to Human Performance and Work Design. Taylor and Francis, Boca Raton (2007)
5. Kieras, D., Polson, P.G.: An approach to the formal analysis of user complexity. Int. J. Man-Mach. Stud. **22**, 365–394 (1985)
6. Abrahamsson, L., Johansson, J.: From grounded skills to sky qualifications: a study of workers creating and recreating qualifications, identity and gender at an underground iron ore mine in Sweden. J. Ind. Relat. **48**(5), 657–676 (2006)
7. Hendrick, H.W., Kleiner, B.M.: Macroergonomics: An Introduction to Work System Design. HFES, Santa Monica (2001)
8. Blunden, A.: An Interdisciplinary Theory of Activity. Koninklijke Brill, Leiden (2010)
9. Bedny, G.Z., Karwowski, W., Bedny, I.S.: Complexity evaluation of computer-based tasks. Int. J. Hum-Comput. Int. **28**(4), 236–257 (2012)
10. Engeström, Y.: Managing as Argumentative History-Making. In: Boland, R.J., Collopy, F. (eds.) Managing as Designing, pp. 96–101. Stanford University Press, Stanford (2004)

11. Leontiev, A.N.: The problem of activity in psychology. Sov. Psychol. **13**(2), 4–33 (1974)
12. Engeström, Y.: Expansive learning at work: toward an activity theoretical reconceptualization. J. Educ. Work **14**(1), 133–156 (2001)
13. Spinosa, C., Flores, F., Dreyfus, H.L.: Disclosing New Worlds: Enterpreneurship, Democratic Action, and the Cultivation of Solidarity. The MIT Press, Cambridge (1997)
14. Engeström, Y.: Learning by Expanding: An Activity-Theoretical Approach to Developmental Research. Orienta-Konsultit, Helsinki (1987)
15. Engeström, Y., Engeström, R., Kerosuo, H.: The Discursive construction of collaborative care. Appl. Linguist. **24**(3), 286–315 (2003)
16. Vygotsky, L.S.: Mind in Society: The Psychology of Higher Mental Functions. Harvard University Press, Cambridge (1978)
17. Kozulin, A.: Vygotsky's Psychology: A Biography of Ideas. Harvard University Press, Cambridge (1999)
18. Lave, J.: The Practice of Learning. In: Chaiklin, S., Lave, J. (eds.) Understanding Practice: Perspectives on Activity and Context, pp. 3–32. Cambridge University Press, Cambridge (1993)
19. Leontiev, A.N.: Activity, Consciousness, and Personality. Prentice-Hall, Englewood Cliffs (1978)
20. Zinchenko, P.: The problem of involuntary memory. Sov. Psychol. **22**, 66–70 (1983)
21. Wertsch, J.V.: Mind as Action. Oxford University Press, Oxford (1998)
22. Lave, J., Wenger, E.: Situated Learning: Legitimate Peripheral Participation. Cambridge University Press, Cambridge (1991)
23. Brown, J.S., Duguid, P.: Organizational learning and communities-of practice: toward a unified view of working. Learning and Innovation. Organ. Sci. **2**, 40–57 (1991)
24. Engeström, Y., Middleton, D.: Introduction: studying work as mindful practice. In: Engeström, Y., Middleton, D. (eds.) Cognition and Communication at Work, pp. 1–14. Cambridge University Press, Cambridge (1996)
25. Spender, J.C.: Making knowledge the basis of a dynamic theory of the firm. Strateg. Manage. J. **17**, 45–62 (1996)
26. Hutchins, E.: Learning to navigate. In: Chaiklin, S., Lave, J. (eds.) Understanding Practice: Perspectives on Activity and Context., pp. 35–63. Cambridge University Press, Cambridge (1993)
27. March, J.G.: Exploration and exploitation in organizational learning. Organ. Sci. **2**, 71–87 (1991)
28. Blackler, F.: Knowledge, knowledge work and organizations: an overview and interpretation. Organ. Stud. **16**, 1021–1046 (1995)
29. Weick, K.E., Roberts, K.H.: Collective mind in organizations: heedful interrelating on flight decks. Admin. Sci. Q. **38**, 357–381 (1993)
30. Bardram, J.E.: Plans as situated action: an activity theory approach to workflow systems. In: Proceedings of the 1997 European Conference on Computer Supported Cooperative Work, Lancaster, U.K, pp. 17–32 (1997)

The Emerging Cognitive Difficulties and Emotional-Motivational Challenges of Ghanaian Air Traffic Controllers: Implication for Improved Job Design

Mohammed-Aminu Sanda[1,2(✉)] and Emmanuel Owusu-Asiedu[1]

[1] University of Ghana Business School, P.O. Box LG 78, Legon, Accra, Ghana
masanda@ug.edu.gh, mohami@ltu.se
[2] Luleå University of Technology, 97187 Luleå, Sweden

Abstract. This study explored the cognitive difficulties and emotional-motivational challenges associated with air traffic control activity in Ghana. This based on the observation that air traffic controllers at the main airport in Ghana operate under very stressful situations in managing the high traffic and movement of aircrafts. Guided by the systemic structural theory of activity, it was found that the air traffic control activity entailed a complex set of tasks that required operators to have very high levels of knowledge and expertise in the practical application of specific skills pertaining to cognitive domains. It is concluded that even though a complex motor task in the air traffic control activity can be performed by the workers, it demands significant cognitive effort, with consequential effect on the emotional-motivational aspects of the activity they perform. This understanding can be incorporated in designing an operator-efficient and effective work system for air traffic controllers in Ghana.

Keywords: Cognitive difficulties · Emotional-motivational challenge · Air-traffic controllers · Air-traffic control activity · Job design

1 Introduction

Air Traffic Controllers are generally considered as one of the working groups that have to cope with highly demanding job situations. The job entails a complex set of routinized activities that require very high levels of knowledge and expertise, as well as practical application of specific cognitive skills pertaining communication and human relations. Six main tasks in the air traffic control activity have been identified by [1]. These include situation monitoring, resolving aircraft conflicts, managing air traffic sequences, routing/planning flights, assessing weather impact, and managing sector/position resources. In all, these tasks entail up to 46 sub-tasks and around 348 distinct operations. Therefore, the cognitive attributes required for high performance at radar workstations include; spatial scanning, movement detection, image and pattern recognition, prioritizing, visual and verbal filtering, coding and decoding, inductive and deductive reasoning, short-term and long-term memory, as well as mathematics and probabilistic reasoning. As such, the main sources of stress reported by Air Traffic

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_33

Controllers are associated with the operative aspects of the job as well as the job's organizational structures [2]. The stress associated with the operative aspect of the job emanates from the peaks of traffic load, time pressure, having to bend the rules, limitations and the reliability of equipment [2]. The stress that emanates from the organizational structures is associated with shift schedules (night work in particular), role conflicts, unfavorable working conditions and the lack of control over work [2]. According to [3], an Air Traffic Controller's work performance can be impaired at certain hours of the day by an excessive workload. It can also be lowered during the night by a decline in mental and physical functions, in spite of a reduced external load [3]. Studies by [4] further suggests that Air Traffic Control takes place in a complex environment and many studies have been conducted to determine the activities involved in controlling air traffic and its complexity. Air Traffic Controllers' job involves cognitive demanding processes, such as constantly processing changing information, keeping the mental picture of the air traffic situation, dividing attention among different situations, solving conflicts, planning ahead, working under time pressure and constantly adapting to changing circumstances [5].

The jobs of air-traffic controllers constantly involve the use of radar, which determines the distance to an object by measuring the time required for a radio signal to travel from a transmitter to the object and return. Such measurements can be converted into lines of position (LOP's) comprised of circles with radius equal to the distance to the object. Since radars use directional antennae, they can also determine an object's bearing. However, due to its design, radar's bearing measurement is less accurate than its distance measurement. Therefore, Air Traffic Controllers are exposed to critical accidents, such as air accidents, near-collisions or loss of control due to overload [6], which situations ignite strong emotional reactions.

In Ghana, The Kotoka International Airport is currently the main international airport which serves both international and local flights, causing a lot of traffic at its premises. This, consequently, comes with high mental demand on all workers within the work environment, especially persons in charge of managing the movement of aircrafts. This study, therefore explored the cognitive difficulties and emotional-motivational challenges associated with air traffic control activity in Ghana, especially, the perceived increase in cognitive workload of Air Traffic Controllers at Kotoka International Airport. This is based on the observation that air traffic controllers at the main airport in Ghana operate under very stressful situations in managing the high traffic and movement of aircrafts. Such a high traffic volume and frequency congestion represent the leading factors that influence their task complexity, characterized by unplanned demands and having to dynamically re-plan in response to weather changes, traffic management initiatives, airport construction, maintenance activities, and other associated air traffic control activities. As such, the number of decisions to be made by air traffic controllers is very enormous and highly stressful, especially when the controllers' decision-making capacity is stretched to the maximum.

2 Literature Review

The complexity of the air traffic control activity is influenced by the pattern of the strategic traffic situation [5]. As such, the physical framework within which air traffic has to be controlled is set by the strategic traffic situation [5]. The strategic situation, either permanent or temporal, can vary in terms of airspace structure and sectors, airport layout and runways used, traffic volume and density, as well as flight planning and flow management [5]. High traffic volume and frequency congestion represent the leading factors influencing complexity [4]. The tactical traffic situation comprises the actual traffic situation, the aircraft positions and clearances given, the mix in traffic and aircraft performances, the traffic flows and routings, the actual weather situation and emergencies and exceptional situations [5]. It is characterized by the dynamic nature of the situation, with continuous changes in the information and situation and factors interacting with each other [5]. These often result in unexpected situations that can be cognitively demanding [5]. As it is observed by [4], much of the complexity is characterized by unplanned demands, and dynamically re-planning in response to weather, traffic management initiatives, airport construction, maintenance activities, and other events. According to [4], unplanned tasks make it difficult for controllers because they cannot prepare for them and may increase controller workload.

Air traffic control work is performed in teams consisting of air traffic controllers, their supervisors and assistants, as well as other professional colleagues from adjacent departments and units, including aircraft pilots [5]. It has been noted by [5] that due to changes in the strategic traffic situation, working positions can be combined or separated, resulting in changes in team configuration and operating procedures. These changes increase the complexity of controlling air traffic. According to [5], the procedures describe the formal operating procedures within which air traffic has to be handled, and comprises standard operating procedures and rules for conducting air traffic control activity, as well as its associated work-methods.

Cognitive complexity is influenced by the number of procedures used, the complexity of procedures, diversity in working methods and the increase of legal liability issues [7]. Some of the procedures that are established for the control of air traffic include the communicating of multiple or lengthy instructions to every aircraft [5] within the precinct of the covered airspace. Air traffic controllers use several systems in performing their tasks, and they include communication systems, planning systems and navigation and surveillance systems [5]. According to [8], air traffic control requires a lot of cognitive processing in the synthesis and analysis of significant amounts of information, the mastery of complex procedures, real time problem solving, and the listening and speaking skills necessary for effective information transfer.

According to [8], the rapid and continued growth in air traffic globally is outpacing and putting strain on air traffic control systems to handle the traffic volume and its inherent complexity [8]. Because of this rapid growth, the workload of controllers also increases [8]. The cognitive skills required to control air traffic include the following as outlined by [8]:

- Perception for sensing and reacting to visual and aural information. For example, detecting and resolving deviations from planned flight paths [8].
- Being attentive/vigilante for prolonged periods of intense activity, and sometimes for prolonged periods of relative inactivity [8].
- Learning to master the procedures, practices and peculiarities of the position as well as from day to day operational experience [8].
- Memory to interpret evolving situations correctly and quickly. Short-term for dealing with situations in real time, and long-term for integrating knowledge and procedures [8].
- Information processing to synthesize many diverse pieces of changing data about traffic, weather, aerodrome conditions, and navigation-aids into a coherent picture and to manage that picture in accordance with existing plans and procedures [8].
- Situational awareness to successfully integrate all relevant information into a coherent and current picture [8]. This includes knowledge of the present, past and pending situation, system functioning, human roles and tasks, as well as air traffic controllers roles, procedures and objectives [8]. Losing this "picture" is the worst nightmare for air traffic controllers [8].
- Planning to integrate the time element by extrapolating from the controller's picture to develop expected aircraft sequencing and spacing in accordance with established procedures and objectives [8].
- Communicating for both the reception and correct interpretation of information as well as for sending information and instructions, often through the barriers of language and radio noise [8]. Effective communication also requires a feedback mechanism to confirm understanding [8]. For example, pilots have to read back the instructions given by controllers and controllers must listen to the read back [8].
- Problem-solving to resolve deviations from plans and cope with unforeseen circumstances such as system outages or aircraft emergencies [8].
- Decision-making for the timely selection of the best alternative course of action for a particular situation, and appreciating how such decision will affect subsequent traffic [8]. Not only must the traffic flow safely and expeditiously, it must continue to be orderly [8].

3 Methodological Issues

As it has been outlined by [9], various practitioners have attempted to develop suitable methods for task complexity evaluation, including the use of various units of measure, such as the number of controls and indicators, or the number of actions [10, 11]. Many of the studies regarding workload in Air Traffic Control were conducted in the laboratory. According to [12] laboratory studies generally focus on relatively weak acute stress. It is also noted by [13] that experiments regarding workload in air navigation have been performed under very simplified simulated conditions, often far from the reality. During such experiments, controllers are exposed to simulated stressful events that might not necessarily be the true reflection of the actual events. But task complexity, as argued by [9, 14] cannot be successfully evaluated by such methods, principally because they employ incommensurable units of measure.

The quantitative method of task complexity evaluation, according to [9], suggests a requirement for units of measurement and measurement procedures that permit the comparison of different elements of activity. However, this important issue has not yet been resolved [9]. Thus, task complexity can be evaluated both experimentally and theoretically [9]. The experimental evaluation is based on criteria, such as the evaluation of probability of errors, the measurement of time performance, the evaluation of duration of skill acquisition, and the measurement of mental fatigue [9]. According to [9, 17], expert judgments, such as the use of a five-point scale for complexity evaluation, and the subjective opinion of the task performer can also be taken into consideration. But as it is pointed out by [9, 15], the motivational aspects of an activity are usually ignored in the design process. These functional blocks of self-regulation are connected with the evaluation of task difficulty and significance, ad play a central role in integrating the cognitive and motivational aspects of activity [9, 14].

Thus arguing from the perspectives of [9, 14], it is deduced that the fundamental notions of task complexity and cognitive challenge and significance in the Air Control Activity, and the concentration of attention on the conduct of the Air Control activity will permit the job designer to take into consideration, not only the cognitive and behavioral aspects of the activity, but also its motivational aspects. In this regard, the subjective opinions of the task performers were taken into consideration [14].

3.1 Sampling Technique

Due to the limited number of Air Traffic Controllers in Ghana, and the difficulty of having access to them, the snowballing technique [15, 16] was used to get as many Air Traffic Controllers to complete the questionnaires as possible. In all, out of a total 85 operational staff (i.e. air controllers) working at the Kotoka International Airport, 50 of them volunteered to participate in the study. As such, quantified qualitative data was collected from these 50 participants using a questionnaire.

3.2 Data Collection Procedure

A standardized self-completion questionnaire entailing five sections was used to collect data from the fifty Air Controllers. Section A of the questionnaire collected information on the controllers' demography. Section B of the questionnaire collected information on the controllers' measure of their work stress. Section C of the questionnaire collected information on the controllers' measure of their workload. Section D of the questionnaire collected information on the controllers' measure of their works' cognitive demand.

Since the study participants' responses are representative of their expert judgments (i.e. subjective opinions) on the complex tasks they perform, it is appropriate to use a five-point scale for such complexity evaluation [9, 14]. As such, the response ratings in sections BC and D of the questionnaire followed the five-point Likert scale, ranging from very low (1) to very high (5).

In the data collection process, 50 questionnaires were delivered to the management of the Air Traffic Control unit at the Kotoka International Airport in Accra. The management then facilitated the administering of the questionnaire to the 50 Air Traffic Controllers identified from snowballing [15, 16] exercise and who volunteered to participate in the study. In all 50 questionnaires administered were fully completed and returned for analysis yielding a response rate of 100%.

3.3 Data Analysis Procedure

Using the systemic analytical approach [9], the cognitive aspect of complexity that depended on the specificity of workload and stress in the air-control activity, and those emotional-motivational aspects of complexity that reflected the energetic aspects of the air-control activity were analyzed descriptively.

4 Results Analyses

4.1 Demographic Assessment of Study Participants

The demographic characteristics of the respondents showed that there were more male respondents (80%) than female respondents (20%). In relation to the number of years spent in the air-control activity, majority of the respondents (76%) are quite experienced, having been in practice for periods of 4 years and above. A significant number of them (20%) have been in practice for less than a year. Only a few of the respondents (4%) have been practicing for a period between 1–3 years.

The implication of the demographic distribution is that, majority of the air-controllers (i.e. study participants) surveyed were duly qualified and experienced to provide the needed information requested in the questionnaire administered.

4.2 Analysis of the Cognitive Aspect of Complexity in Air-Control Activity

The tabulation of the number of respondents out of the 50 study participants who attributed their stress to the cognitive demand of the air-traffic control activity is shown in Table 1 below. As it is highlighted in the table, 47 (90%) of the respondents agree that in the air-traffic control activity, assigning them cognitively demanding tasks increases their mental workload. Almost all the respondents (98%) share the perception that the air-traffic control activity entails a complex set of tasks requiring very high levels of knowledge and expertise, as well as the practical application of specific skills pertaining to cognitive domains (e.g. spatial perception, information processing, logic reasoning, decision making), communicative aspects and human relations. The respondents agreed that meeting all these task demands adds to the complexity of their jobs, and the excessive cognitive demand it entails.

Similarly, all the respondents (100%) agree that the air-traffic control activity involves cognitive demanding processes, such as constantly processing changing

Table 1. Tabulation of the number of respondents attributing stress to the demanding nature of the air-traffic control activity

Cognitive factors	No. agreed
Since automated systems performs many cognitively demanding tasks faster and more accurately than humans, assigning cognitively demanding tasks to humans increases mental workload	47
The very high levels of knowledge and expertise, as well as the practical application of specific skills pertaining to cognitive domains (e.g. spatial perception, information processing, logic reasoning, decision making), communicative aspects and human relations, required in the job makes a complex activity	49
The cognitive demanding processes involved in the job, such as constantly processing changing information, keeping the mental picture of the air traffic situation, dividing attention among different situations, solving conflicts, planning ahead, working under time pressure and constantly adapting to changing circumstances, adds to its complexity	50

information, keeping the mental picture of the air traffic situation, dividing attention among different situations, solving conflicts, planning ahead, working under time pressure and constantly adapting to changing circumstances is confirmed. The respondents agreed that the incorporation of all the tasks in the activity performance adds to its complexity and cognitive demand.

4.3 Analysis of the Workload Aspect of Complexity in Air-Control Activity

The tabulation of the number of respondents out of the 50 study participants who attributed their stress to the workload of the air-traffic control activity is shown in Table 2 below.

As it is shown in Table 2 above, 49 (98%) of the respondents agree that in the air-traffic control activity, decision is made in a dynamic environment involving many

Table 2. Tabulation of the number of respondents attributing stress to the demanding nature of the air-traffic control activity

Workload factors	No. agreed
Making decision in a dynamic environment involving many factors, such as constantly updating relevant information and resolving conflicting goals, and often making difficult decisions with incomplete information, under time pressure, and high workload increase the job complexity	49
High traffic volume and frequency congestion represent the leading factors influencing the task workload and complexity	49
Much of the complexity in the activity is characterized by unplanned demands and having to dynamically re-plan in response to weather, traffic management initiatives, airport construction, maintenance activities, and other events	49

factors, such as constantly updating relevant information and resolving conflicting goals, and often making difficult decisions with incomplete information, under time pressure, and with high workload. The respondents agreed that all these issues increase their mental workload.

Almost all the respondents (98%) share the perception that the air-traffic control activity entails high traffic volume and frequency congestion, which situations adds to the complexity of the activity. Similarly, all the respondents (100%) agreed that much of the complexity in the air-traffic control activity, and its accompanying workload demand is characterized by unplanned demands and having to dynamically re-plan in response to weather, traffic management initiatives, airport construction, and maintenance activities. They agreed that incorporation of all these tasks in the activity performance adds to its complexity and cognitive demand.

4.4 Analysis of the Stress Aspect of Complexity in Air-Control Activity

The tabulation of the number of respondents out of the total 50 study participants who attributed their stress to the demanding nature of the air-traffic control activity is shown in Table 3 below.

Table 3. Tabulation of the number of respondents attributing stress to the demanding nature of the air-traffic control activity

Stress factors	No. agreed
Because the activity is extremely stressful, due to its characteristics, extensive responsibilities is shouldered by controllers	49
Due to the high number of decisions involved in the task performances, it becomes stressful when the controllers' decision-making capacity is stretched to the maximum	48
The information content of radio transmissions (e.g. instruction, information, question, report, and inquiry), their frequency and duration also contributes to the stress level of the activity	48

From Table 3 above, 49 (98%) of the respondents agree that the air-traffic control activity is a very stressful and demanding job. Almost all the respondents (96%) share the perception that the air-traffic control activity becomes a stressful when the Controllers' decision-making capacity is stretched to the maximum. Similarly, all the respondents (96%) agree that in the air-traffic control activity, handling the information content of radio transmissions (e.g. instruction, information, question, report, and inquiry), their frequency and duration also contributes to their stress.

5 Discussion

The results analyses have shown that in the air-traffic control activity, the principal actor is the individual (controller) employee who is simultaneously confronted with the challenges of handling the consequences of both the cognitive and motivational-emotional aspects of the air-traffic control's complexity. For the air-traffic control activity, one activity type is the operator's engagement in a complex set of tasks requiring very high levels of knowledge and expertise, as well as the practical application of specific skills pertaining to cognitive domains (e.g. spatial perception, information processing, logic reasoning, decision making), communicative aspects and human relations. As found in the analysis, the other activity type, which occurs simultaneously with the physical task, is the operators' engagement with cognitive (mental) task through interaction with digitized communication models, such as, constantly processing changing information, keeping the mental picture of the air traffic situation, dividing attention among different situations, solving conflicts, planning ahead, working under time pressure, and constantly adapting to changing circumstances. Since this routinized activity entails various characteristics of decision-making process at the verbal-thinking level, it signifies an important factor that influences the complexity of a task [14]. Therefore, the increase in the cognitive aspect of complexity in the controllers' could be related to the need for them to alter the stereotypical actions of the designed tasks in the air-traffic control activity. This is a realization that while the decision-making process is more complicated when it is determined by information extracted from memory, it becomes easier to perform when it is predominantly determined by external stimuli or by information provided from external source [9].

The implication from the results analyses is that the air-traffic controllers' objective and subjective activities are interdependent and shaped based on the mechanisms of self-regulation. The cognitive aspect of complexity in the air-traffic control activity is found to depend on the number of task elements and the specification of interactions of the different task elements. As found from the results, this include decision-making in a dynamic environment involving many factors, such as constantly updating relevant information and resolving conflicting goals, and often making difficult decisions with incomplete information, under time pressure, and with high workload. It also included the handling of information content from radio transmissions (e.g. instruction, information, question, report, and inquiry), their frequencies and duration, a task is very stressful and demanding, and which stretches the controllers' decision-making capacity to the maximum. Similarly, the mental or cognitive efforts expended by the operators could be said to have been influenced by the goals that their object-oriented activity aimed to achieve [14]. Thus, the complexity in the air-traffic control activity undertaking by the controllers was defined by the number of static and dynamic components of the task and interaction among these components [9]. As found, these include the high traffic volume and frequency congestion, air-traffic control activity which is characterized by unplanned demands, with controllers having to dynamically re-plan in response to weather, traffic management initiatives, airport construction, and maintenance activities. In this regard, by using the argumentation of [9, 14], the degree to which the tasks entailed in the air-traffic control activity could be viewed as unpredictable, and the uncertainty that could be associated with such unpredictability must be seen as an important influencing factor that adds to the activity's complexity. Thus arguing from the perspectives of [14, 17], the number rules in the air-traffic control activity that the controllers have to learn must serve as the major criterion of determining the level of the cognitive and motivational-emotional aspects of the activity's complexity. This understanding will help the designers of both the "activity" and the "procedures" for its performance to optimize the complexity of the air-traffic control activity and successfully integrate its cognitive and motivational-emotional characteristics.

6 Conclusion

The study has shown that the air-traffic control entailed several operator-oriented challenges with the complexity of the activity as an influencing factor. Based on the findings as discussed above, it is concluded that even though a complex motor task in the air traffic control activity can be performed by an operator, it demands significant cognitive effort and concentration from the operators which has consequence effect on the emotional-motivational aspects of the activity they perform. By implication, ameliorating the cognitive demanding processes of the air traffic control activity is critical when the complexity associated with both the cognitive aspect and the emotional-motivational aspects of the tasks entailed in the activity is to be understood. This is because, as these complexities increase, the emotional tension and motivational force associated with the operators' work also increase.

Thus designers of the task environment for the air-traffic control activity need to understand the emerging new mental images of the operators in order to be able to come out with human-oriented designs that can lend to the creation of a friendly, efficient and effective work environment for the air-traffic control. The work environment must also have good work and social conditions that enhance the workers 'emotional-motivational orientation. By implication, air-traffic control job designers' understanding of the controllers' cognitive difficulties and the emotional-motivational challenges associated with their tasks can result in the designers' ability to design an operator-efficient and effective work system for air traffic controllers in Ghana.

Acknowledgements. We are grateful to the management of the air-traffic control unit at the Kotoka international airport for the appreciable support provided us in the conduct of this study. We are also thankful to all the fifty air-traffic controllers who volunteered as study participants.

References

1. Ammennan, H.L., Bergen, L.J., Davies, D.K., Hostetler, C.M., Inman E.E., Jones G.W.: FAA Air Traffic Control operations concepts, vol. VI. AKTCC/HOST En route controllers, report number DOT/FAA/AP/86-01. Federal Aviation Administration, Washington (1987)
2. Smith, R.C.: Stress, Anxiety and the Air Traffic Control Specialist. Federal Aviation Report No. FAA-AM-80-14, Washington (1980)
3. Folkard, S., Monk, T.H.: Hours of Work: Temporal Factors in Work Scheduling. Wiley, Chichester (1991)
4. Koros, A., Rocco, P.S.D., Panjwani, G., Ingurgio, V., D'Arcy, J.: Complexity in Air Traffic Control Towers: A Field Study (2003). hf.tc.faa.gov/technotes/dot-faa-ct-tn03-14.pdf
5. Schuver-van Blanken, M., Huisman, H., Roerdink, M.: The air traffic controller cognitive process and operational situation model: a model for analyzing cognitive complexity in ATM. In: Proceedings of 29th Conference of the European Association for Aviation Psychology, Budapest, Hungary (2010)

6. Costa, G.: Occupational Stress and Stress Prevention in Air Traffic Control. International Labour Office, Geneva (1996)
7. Pawlak, W.S., Brinton, C.R., Crouch, K., Lancaster, K.M.: A Framework for the Evaluation of Air Traffic Control Complexity, pp. 1–11. American Institute of Aeronautics and Astronautics (1996)
8. International Civil Aviation Organization: Human Factors Guidelines for Safety Audits Manual (Doc 9806). ICAO, Montreal (2002)
9. Bedny, G.Z., Karwowski, W.: A Systemic-Structural Theory of Activity: Applications to Human Performance and Work Design. Taylor and Francis, Boca Raton (2007)
10. Mirabella, A., Wheaton, G.R.: Effects of Task Index Variations on Transfer of Training Criteria. Report NAVTRAEQUIPEN 72-C-0126-1. Naval Training Equipment Center, Orlando, Florida. January (AD 773-047/7GA) (1974)
11. Venda, V.F.: Engineering Psychology and Synthesis of Informational Sources. Manufacturing Publishers, Moscow (1975)
12. Bourne Jr., L.E., Yaroush, R.: A Stress and Cognition: A Cognitive Psychological Perspective (2003). http://psych.colorado.edu/~lbourne/StressCognition.pdf
13. Villena, J., Acevedo, J.M., Cruz, J.A., Le Gal, S.: Is it possible to observe air traffic controller (ATCOs) mental workload? In: Proceedings of 28th Conference of the European Association for Aviation Psychology, Valencia, Spain (2008)
14. Sanda, M.A. Cognitive and emotional-motivational implications. In: Hale, K.S., Stanney, K.M. (eds.) The Job Design of Digitized Production Drilling in Deep Mines. Advances in Neuroergonomics and Cognitive Engineering, Advances in Intelligent Systems and Computing, vol. 488, pp. 211–222. Springer, Switzerland (2016)
15. Patton, M.Q.: Qualitative Evaluation and Research Methods. Sage, Newbury Park (1990)
16. Sanda, M.A.: Impact of value-based transformational leadership in privatizing government institutions in a developing economy: a case study. Bus. Manag. Q. Rev. **1**(3), 1–13 (2010)
17. Kieras, D., Polson, P.G.: An approach to the formal analysis of user complexity. Int. J. Man-Mach. Stud. **22**, 365–394 (1985)

Studying Thinking in the Framework of SSAT

Gregory Bedny[1(✉)], Inna Bedny[1], and Waldemar Karwowski[2]

[1] Evolute, LLC, Wayne, NJ, USA
gbedny@optonline.net
[2] University of Central Florida, Orlando, FL, USA
wkar@ucf.edu

Abstract. Analyzing literature in ergonomics and psychology one can see that now attention is mostly given to mechanisms of memory while thinking is not considered enough when studying human work. Thinking is a vitally important cognitive process for problem-solving tasks. In this work, we consider thinking as a self-regulative system with function blocks being the main units of activity analysis. Such blocks are subsystems with specific regulatory functions within the structure of the self-regulatory system. In SSAT such approach of activity analysis is called functional analysis. Work psychology and ergonomics are not so much involved in studying isolated cognitive processes but rather in considering their combination in specific types of work activity. Content of each block can be described in terms of cognitive processes that are evolved in activity regulation or in term of cognitive and behavioral actions that are basic units of activity analysis. All models of activity self-regulation including the thinking model are not describing a homeostatic but rather a goal-directed self-regulative process. SSAT understands goal as a conscious component of activity that is connected with the motives. The self-regulative model of thinking is important for task analysis and specifically when analyzing tasks that emerge unexpectedly, performed in time restricted and dangerous conditions, accompanied by stress, has problem-solving aspects, and if the situation that is faced by the operator is of a stable structure with embedded dynamic elements. The thinking model includes nine function blocks and includes their interconnection. The function of each block is determined by theoretical data accumulated in the general, applied and system-structural activity theories. Some data from cognitive psychology is also taking into consideration as well as the opportunity to manipulate elements of the situation (operative units of thinking) that cannot be directly perceived but only mentally constructed. Such manipulation with operative units of thinking is provided by cognitive actions and specifically by thinking actions. Practical or behavioral actions during the thinking process are also considered in this model. Goal formation, development stable and dynamic mental models, formation of the program of performance, mental transformation of initial mental situation into new one, etc. as functions of various blocks are described. In general, presented model is considered as a loop structured system based on which various strategies of thinking during task performance can be described.

Keywords: Thinking · Self-regulative system · Function block · Self-regulation model of thinking · Cognitive thinking actions

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_34

1 Introduction

In this paper, we consider the concept of thinking from the perspectives of the systemic-structural activity theory (SSAT). Bedny has developed this new framework in psychology and ergonomics [1–4]. Thinking is a vitally important cognitive process when considering problem-solving tasks. Thinking is aimed at discovering new properties, relationships between phenomena and objects of reality that are directly given in the perceived situation or are unknown to the subject. It is also aimed at the transformation of data, including ideal objects in order to discover their properties and relationship. Thinking plays an important role in the study of human work. A worker may have to perform cognitive tasks that are ill-defined. Based on the information available to him/her from the work situation he/she may not be able to immediately formulate the task. This makes the task a problem and so it must be formulated as a problem. The process of task-problem formulation can often be more difficult than the actual task solution. Thus, in problem solving we can outline several stages.

At the first stage, there is only a problem-solving situation. It is something vague and not entirely conscious to a subject. At the second stage based on problem-solving situation subject can independently formulate task-problem. In work environment, the situation may be complicated because of time constraints. This is clearly seen in the study of pilots' tasks. One of specific features of the pilots' work is the fact that the performed tasks include instrumental and non-instrumental signals. Signals presented by equipment displays are instrumental signals, signals that are not presented by equipment displays are non-instrumental signals [5]. Such signals as vibration, engine noise, smell, etc. belong to the later. These signals should be correctly interpreted by the pilot. Non-instrumental signals are of two types: those that are interpreted based on learned instructions, and those that are understood based on independent learning from experience. Instrumental and non-instrumental signals can contradict each other, producing an incorrect task formulation. The ability to interpret non-instrumental signals specifically in contradicting conditions involve thinking and specifically its intuitive components.

Presently, ergonomists and cognitive psychologists pay much more attention to study of memory than thinking when studying human work. The other negative factor is the fact that practitioners pay much more attention to the decision making process. However, problem solving cannot be reduced to decision making that is a component of the thinking. Decision making is possible only when a task is understood by a worker. Thinking is especially important cognitive process for various problem-solving tasks. Any problem solving task includes givens or pieces of information that are presented to the subject; the goal or desired future result; cognitive and behavioral actions that should be performed to reach the goal. However, the concepts of goal and action have totally different meaning in SSAT in comparison with the interpretation of these concepts in the cognitive psychology. These two terms have no precise meaning in cognitive psychology. In problem-solving tasks, a goal often cannot be precisely formulated at the beginning of the problem-solving process but it is formed in an approximate manner, and as the solution of the problem gradually advances, the goal becomes clearer and more specific. Cognitive psychology ignores consciousness of the goal, specific interactions with motives, and in most situations simply integrates the goal with motives, which is incorrect. Cognitive psychologists often utilizes the term production instead of cognitive actions. They are basic units of procedural knowledge [6]. The term production has a mentalist orientation. Thoughts are considered to be held in the head of a person without its relevant environmental context. For example, Anderson considers human thinking as a system that manipulates internal rules in the person's mind [7]. These rules are triggered automatically in working memory. They define the character of practical actions that are performed by a person. However, human activity is object oriented. Object-oriented activity is performed by a subject using tools on a material or ideal objects. A tool can be material or mental. Object oriented activity is possible only by utilizing cognitive and behavioral actions. Study of thinking when major units of analysis are cognitive and motor actions is called morphological analysis of thinking.

2 Types of Thinking Process

In activity theory specialists distinguish various types of thinking such as theoretical and practical thinking depending on a type of a problem-solving task performed by a person. The latter is important when studying problem-solving task performance. Practical thinking in turn can be categorized into manual-manipulative, operative, and supervisory thinking. Manual-manipulative thinking is considered to be more suitable for blue-color workers. When an operator works with an automatic and/or semiautomatic systems the operative thinking process dominates. For such tasks an operator uses cues from instruments and manipulates controls. Sensory-perceptual and imaginative processes are extensively involved in this type of thinking. Supervisory thinking determines the style of leadership and influence on subordinates. Here, logical components and planning are important. Due to the supervisory planning, the problem-solving tasks have a social involvement. One of the components of this type of thinking is the formation and distribution of the tasks and clear formulation of their goals among personnel and coordination of task performance in time. For operative and supervisory thinking, a time limit is common. This often produces emotional tension in the thinking process. Erroneous solution of the problem in the work environment can also be dangerous. All these factors can involve emotional components in operative and supervisory thinking processes. The most important features of operative thinking are [8]:

1. Operative thinking is directed toward the solving of practical problems.
2. A practical cognitive and motor actions are formed based on operative thinking and are immediately analyzed in a practical situation. It is possible to immediately correct and change the thinking process.
3. Operative thinking is often performed under time constraints and can be accompanied by stress.
4. Operative thinking has three main functions: planning, control and regulation of practical thinking actions. It also can include intuitive thinking operations that might not be sufficiently conscious.

The following components of the thinking process are important: creation of mental structure of existing situation (development of enlarge meaningful units of thinking based on the connection between different elements of a situation), dynamic recognition, anticipation of the effects of the final situation, formation of algorithm of task solution (development of principles and rules of solutions), and sequence of cognitive and motor actions performance, required to perform the task. Usually visual type of information plays a leading role in task performance.

When eye movements are recorded in the studies of cognitive task performance, it is necessary to distinguish between perceptual eye movement and thinking eye movement. Perceptual actions have purpose of recognizing elements of the situation. In contrast the thinking actions have purpose of recognizing the meaning of elements of the situation based on the analysis of their functional relationship.

Utilizing sequential fixations a subject can extract from identical situation distinct essential characteristics that are germane to the solution of a problem. These features are not always conscious. According to Pushkin [8] this thinking eye movements are vital for extraction of “non-verbalized operational meaning” or “situation concept of thinking.” Therefore, thinking eye-movements are involved in the inner mental operations. The same external situation can constantly change in the mind of the subject. This external change of the situation in the mind of the subject in light of its external constancy is called “gnostic dynamic.” SSAT shows that the self-regulation process is the basis of gnostic dynamics. Thus, in the following section we consider thinking as self-regulative system.

3 Analysis of Thinking as the Self-regulative System

Work psychology and ergonomics are not so much involved in studying isolated cognitive processes but rather in considering their combination in specific types of work activity. Such principle of studying cognitive processes is provided by functional analysis of activity when activity is considered as self-regulative system with function blocks to be the main units of its analysis. Each function block includes various cognitive processes that are organized in various ways depending on specificity of function block and tasks performed by a subject. From the SSAT perspectives, we can describe the thinking process as a self-regulative system that is comprised of various stages presented as interdependent function blocks (see Fig. 1). Such blocks are subsystems with specific regulatory functions within the structure of the self-regulatory system that depict thinking process during task performance. Content of each block can be described in terms of cognitive processes that are evolved in activity regulation or in terms of cognitive and behavioral actions that are basic units of activity analysis. All models of activity self-regulation including the thinking model are not describing homeostatic but rather goal-directed self-regulative processes.

When a subject perceives information from a computer screen or a display panel, the essential data about a task is considered as an informational model. A subject selects information that is subjectively most important in a presented situation. He/she also can select required information directly from the environment. The selection of information also depends on emotionally-motivational aspects of the subject’s activity.

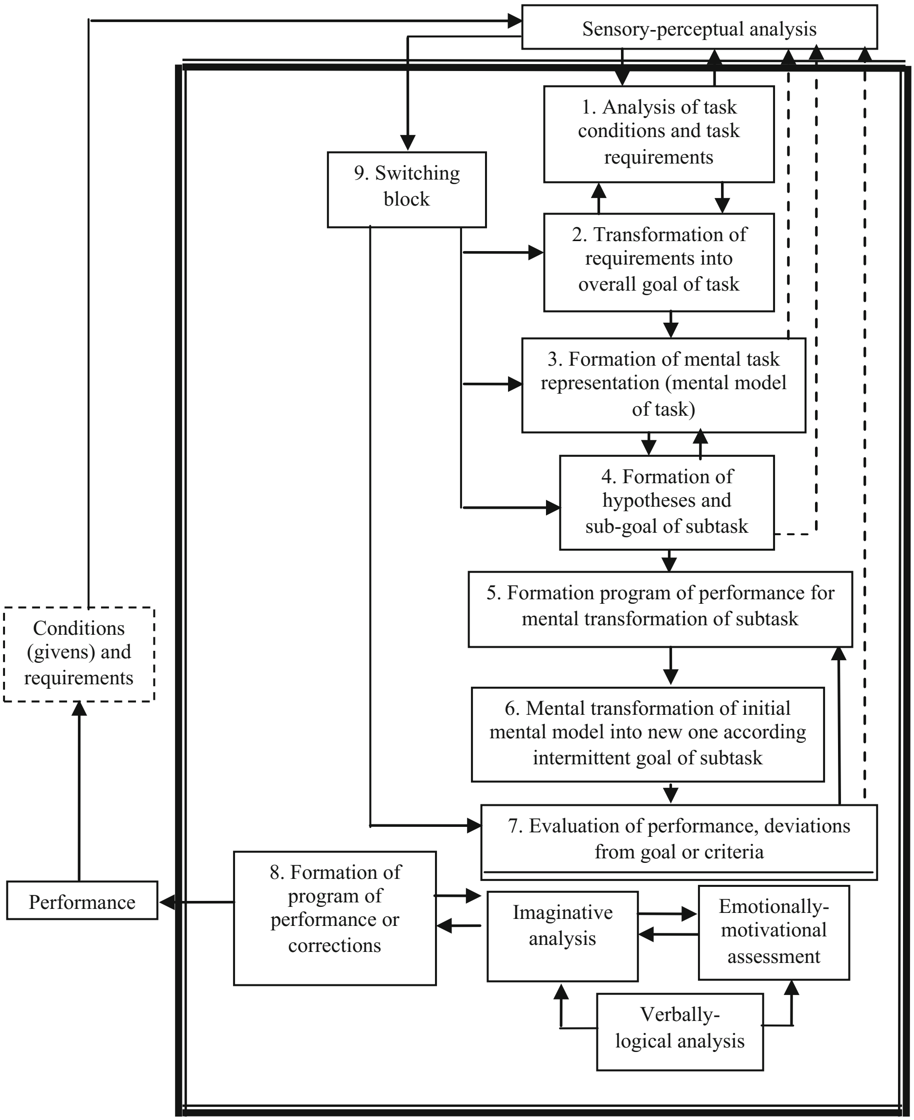


Fig. 1. Self-regulation model of the thinking process.

Therefore, not only externally presented data but also the internal mental state of the subject, her/his emotionally motivational state, is a key factor in selecting the required information. Thus, the process of receiving information involves not only sensation and perception but it is also closely linked to past experience or memory, attention, and thinking. At the first stage task conditions (givens) and requirements are received by a subject. This is sensory-perceptual analysis of problem-solving situation. This block is not directly involved in thinking process and is presented outside of thinking model.

At the stage of the sensory-perceptual analysis of a problem-solving task, verbal components of thinking are presented in a reduced form. Verbal processes are mainly connected with the mental categorization of elements of the situation, assigning them to a class of related objects.

The first functional block that is directly involved in the thought process is block 1 (analysis of task conditions and requirements). At this stage of task performance, a subject defines task conditions and task requirements that can be most relevant to solving a problem. A subject restructures a task and assesses the obtained results according to the subjective significance of the elements of a problem-solving task. Restructuring of a problem can be performed not just practically but also mentally by using thinking actions. Major operative units of thinking actions are concepts, propositions, and mental images. Thinking actions and operations are involved in the manipulation of these operative units of thinking.

The formation of mental task representation has a preliminary character because the goal of a task is not yet formed or accepted by the subject. This becomes possible only after block 2 (transformation of requirements into overall goal of task) is activated and the final goal of the task is formed and accepted. Blocks 1 and 2 have forward and backward connections and therefore these blocks are mutually adjusted (see Fig. 1). Based on the analysis of task conditions and accepted goal, a subject develops his/her mental model of the task. This stage of the thinking process is associated with block 3 (formation of mental task representation or mental model of task).

When a mental model of task is created, the *solution stage begins*. After understanding a task at hand, a subject divides a task into subtasks. He/she begins the formation of subtasks by formulating various hypotheses. Each hypothesis has its own potential goal. Based on comparison and evaluation of such hypotheses, a subject selects one and formulates the first sub-goal associated with the selected hypothesis (see function block 4; formation of hypothesis and sub-goal of task). Comparing a new sub-goal with an existing mental model of a task allows transforming an original mental model into a new one that is adequate for a new subtask. Therefore, this stage of thinking process is associated with block 5 (formation of program of performance for mental transformation of subtask) and block 6 (mental transformation of initial mental model into a new one based on intermittent goals of task).

The obtained result is evaluated in block 7 (evaluation of performance, deviations from goal or criteria) based on its correspondence to the goal or sub-goal of a task and a subjective criterion of success. The last one can deviate from the objective goal, and therefore a subject can evaluate her/his own result as a successful one even if it is lower than required. For example, a blue-collar worker can lower quality of performance in order to increase quantity. We have to distinguish goal from the subjective criteria of success. A goal is formulated in advance, but subjective criteria of success can change during task performance.

Function block 7 has three sub-blocks. The first one is involved in conscious and precise verbally logical analysis and evaluation of the thinking process. The sub-block on the left-hand side (imaginative analysis) is involved in not sufficiently conscious and therefore sufficiently precise evaluative process. The most imprecise is the third sub-block on the left-hand side called emotional-motivational assessment. Function block 7 has a feedback connection with block 5 because performance program for a subtask can be corrected and mental transformation of initial mental model into new one can be repeated based on such evaluation. This cycle can be repeated several times until evaluative stage (block 7) is perceived by the subject as positive.

After completion of the evaluative stage, a subject develops her/his program of performance that can be utilized not just for mental but also for real transformation of the problem-solving situation (see block 8, formation of program of performance or corrections). Block 8 can be corrected based on feedback from block 7. Results of an evaluative stage can be utilized in two ways:

1. Transmitting information into block 1 and then the whole cycle of thinking process can be repeated.
2. Switching block 9 to various blocks when the result of sensory-perceptual analysis can be used by blocks 2 or 3 or 4. Switching block 9 can also transmit information directly to block 7 (evaluation of performance, deviation from goal or criteria).

The actual performance result of a problem-solving task is evaluated in block 7 and the obtained data can be used for corrections of performance in block 8. If received data is evaluated positively, the thinking process cycle is completed. If the result is evaluated negatively, information can go from block 7 to block 8 or block 5 or all the way back to the sensory-perceptual block.

If information from block 7 goes to block 8, the program of performance can be corrected. When information from block 7 enters block 5, a subtask can be mentally corrected. If information from block 7 goes to the sensory-perceptual block, the whole task is re-evaluated. Hence, switching block 9 and block 7 can selectively influence various blocks. Connections between the blocks demonstrate that the model describes strategies of the thinking process very flexibly. Therefore, the circle of thinking regulation can be shorter or longer, depending on the specificity of a problem-solving task and the thinking process strategies selected by a subject.

4 Conclusion

Thinking should be considered as a complex self-regulative system with a variety of strategies for solving the same problem. Such a self-regulative process includes various stages or function blocks. Each stage is evaluated based on feed-forward and feedback influences between function blocks of self-regulation process. Function blocks or functional mechanisms are the important units of analysis of thinking. The more precisely we can describe functions of the thinking process at various steps of its regulation the more precisely we can define strategies of thinking during performance problem-solving tasks. Forward and backward connections between such blocks show relationships between stages of thinking and allow conducting a more effective analysis of strategies of the thought process.

References

1. Bedny, G.Z.: The Psychological Foundation of Analyzing and Design Work Processes. Higher Education Publishers, Kiev (1987)
2. Bedny, G.Z., Meister, D.: The Russian Theory of Activity: Current Application to Design and Learning. Lawrence Erlbaum Associates, Publishers, Mahwah (1997)
3. Bedny, G.Z.: Application of Systemic-Structural Activity Theory to Design and Training. CRC Press, Taylor & Francis Group, Boca Raton (2015)
4. Bedny, G.Z., Karwowski, W., Bedny, I.S.: Applying Systemic-Structural Activity Theory to Design of Human-Computer Interaction Systems. CRC Press Taylor & Francis Group, Boca Raton (2015)
5. Ponomarenko, V., Bedny, G.: Characteristics of pilots' activity in emergency situations resulting from technical failure. In: Bedny, G., Karwowski, W. (eds.) Human-Computer Interaction and Operators' Performance, pp. 223–254. CRC Press, Taylor & Francis Group, Boca Raton (2011)
6. Anderson, J.R.: Cognitive Psychology and Its Application, 2nd edn. Freemen, New York (1985)
7. Anderson, J.R.: Rules of the Mind. Lawrence Erlbaum Associates, Publishers, Hillsdale (1993)
8. Pushkin, V.V.: Construction of Situational Concepts in Activity Structure. In: Smirnov, A.A. (ed.) Problems of General and Educational Psychology, pp. 106–120. Pedagogy, Moscow (1978)

Concept of the Computer-Based Task in Production and Non-production Environment

Gregory Bedny(✉) and Inna Bedny

Evolute, LLC, Wayne, NJ, USA
gbedny@optonline.net

Abstract. In this work we conduct comparative analysis of concept of the computer-based task in production and non-production environment within the framework of the systemic-structural activity theory (SSAT) and cognitive psychology. These two approaches have significantly different understanding of the concept of task and of its basic elements and without the clear understanding of the concept of task it is impossible to accurately conduct task analysis. Today, computer-based technology opens a new wide area of nontraditional ergonomics design that includes recreational and nonproduction design. Computing technology with increasing frequency is now used for nonproduction purposes creating a field for designing human-computer systems for recreation activity, education, games, etc. By analyzing the task that is being performed by a user we can evaluate usability of HCI system in production and non-production environment. Classifications of activity such as play, learning and work and their relationship are considered in this work from perspectives of SSAT. In this work, we demonstrate that "pleasure-based" activity design as it currently exists in ergonomics covers only a very narrow area of recreational design. Such concepts as play and game are not considered in this area at all. Activity is classified into three categories: play, learning and work. The paper demonstrates process of transformation of play into game which is important for understanding the concept of task in recreation conditions.

Keywords: Systemic-structural activity theory · Production and non-production environment · Task · Goal of task · Play and game · Risky decisions

1 Introduction

Analysis of human-technology interaction usually begins with a description of the tasks. The concept of task and the techniques of task analysis are the most potent tools for the analysis of the human-technology relationship that determine the steps by which humans interact with technology. According to Bedny [1–3] who is the founder of systemic-structural activity theory, task analysis can be defined as the study of work activity strategies in terms of logical organization of cognitive and behavioral actions to achieve the goal of the task that is based on the demands of the system.

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_35

According to this definition, the goal of the task and the goal of the system are not the same. The overall goal of a task should be accepted or formulated by the performer based on the requirements of the system. Furthermore, there is never just one optimal way of task performance. Possible strategies of task performance should be developed based on constraint-based principles that depend on the system requirements. This understanding of task analysis eliminates contradiction between the so-called normative and constraint-based principles of task analysis. However, the nature of work constantly changes. Recently, the main change in human work is its computerization. This evokes new issues in the study of human work associated with informational technology. One of such issues is analysis of computer-based tasks in production and non-production environment. At present, we can distinguish two different approaches in the study of tasks in these various environments. One approach considers tasks from cognitive approach perspective and another one from of the systemic-structural activity theory (SSAT) perspectives. SSAT defines activity as a system that integrates cognitive and behavioral actions in the specific logical way that allows a subject to reach the goal of task. However, computers are utilized not only in work environment but also in recreation environment. These two types of environment bring specialists to conclusion that they have to distinguish between two types of tasks: computer-based tasks in production and non-production environment. Emotionally-motivational factors are important factor in task analysis. However, these factors in cognitive psychology and ergonomics usually are reduced to the study of the emotional or psychic tension of an operator in a stressful situations. This approach does not consider positive emotions in task performance, the relationship between emotions and motivation, and so on. Hence, in this work we will conduct the comparative analysis of concept of the computer-based task in production and non-production environment within the framework of SSAT and cognitive psychology. These two approaches have significantly different understanding of these concepts in task analysis.

2 Basic Characteristics of Task and Their Classification

Tasks itself is a basic component of activity. Our lives can be conceptualized as a continuing chain of various tasks. For example, people run a variety of tasks to maintain their house in order such as cleaning, washing dishes, and so on. This kind of tasks can be an object of ergonomic studies when working on designing or improving housekeeping related devices. The design principles for the development of machines, computers, and kitchen appliances have similarities. We can evaluate the adequacy of equipment, computer interface, or home appliances only by assessing them in the context of task performance. The whole diversity of tasks has two main types: skill-based tasks and problem-solving tasks. These two types of tasks require different levels of automaticity with which they are performed. Skill-based tasks require standardized methods of performance. An example of such tasks would be a simple deterministic production operation. These tasks very often are repetitive and have high level of task performance. The more complex problem solving tasks require discovering the unknown based on what we already know. These are purely creative tasks that are seldom in the performance of production processes. They sometimes can be caused by ill-defined task requirements.

Every task includes objective requirements (goal) of tasks and conditions (givens). Anything that is presented to a person and should be considered during performance is a task condition. What has to be achieved including finding a solution is known as task requirements. When the task requirements are given and accepted by an individual, they become a personal goal. Hence, goal recognition, goal interpretation, formulation, and acceptance of the goal are means by which the task's goal is formed. Thus, the objectively given goal (requirements) and the subjectively accepted goal are not the same, which in itself is an important aspect of task analysis.

Actions are one of the main units of task analysis. They can be cognitive and behavioral. The performance of actions leads to achievement of their goals. The performance of all actions required by a task leads to the achievement of the overall goal of the task. Therefore, cognition is not just storage for images, concepts, or propositions but also a system of mental actions and operations carried out with and upon them. Cognitive and behavioral actions are classified and described according to developed in SSAT standardized principle.

In cognitive psychology, task taxonomy is now based on Rasmussen's work [4] that considers skills, rules, and knowledge-based tasks. Each category of taxonomy has its own level of cognitive control of activity. This taxonomy is well known in literature. In SSAT task taxonomy is different. There are two initial classes of tasks: skill-based tasks and problem-solving tasks. Problem-solving tasks in turn can be divided into two basic groups: algorithmic and non-algorithmic tasks. Algorithmic tasks are performed according to some logic and rules. Algorithmic tasks are of deterministic and/or probabilistic types.

In deterministic algorithmic tasks, subjects make simple "if-then" decisions based on familiar perceptual signals and rules. Each decisions have only two outputs. For example, "if the red bulb is lit, then perform action A; if the green bulb is lit, then perform action B." Algorithmic tasks explicitly define the rules and logic of actions to be performed and guarantee successful performance if the subject follows the prescribed instructions. Deterministic algorithmic tasks according to Rasmussen terminology can be compared with rule-based tasks.

Probabilistic algorithmic tasks involve decisions with the possibility of two and more outputs, each of which possesses a different probability of occurrence. This probabilistic element significantly increases the operator's memory workload and the complexity of task performance in general. Probabilistic algorithmic tasks can be very complex and can present a significant workload not only for memory but also for thinking. Because of the inability to remember all possible rules for the performance of probabilistic algorithmic task and insufficient familiarity with probabilistic characteristics of the task, this task very often becomes semi-algorithmic and even semi-heuristic. According to Rasmussen terminology [4], probabilistic algorithmic tasks and semi-algorithmic tasks can be considered knowledge-based tasks. Purely heuristic tasks are seldom in production environment. The experience of a subject is critical for task classification. If the subject does not possess the required knowledge, even an algorithmic task can become a creative one. As we show below, the classification of tasks for production and non-production environment is the same.

3 Task Performance in Production and Non-production Environment

Specialists who study human work in activity theory distinguish between object-oriented and subject-oriented types of activity. However, there is another classification of activity: recreational activity, learning, and work activity. Such activity classification is important for understanding the concept of the computer-based task in production and non-production environment.

All three kinds of activities have similar structure and include goals, motives, cognitive, and behavior actions. Analysis of children at play has an important role in the study of HCI in a non-production environment. For our analysis, we consider the role of playing in mental development. Vygotsky's work gave the most wide-ranging account of psychological characteristics of game and its role in mental development [5]. When a child plays, it fulfills two functions: formation of a child's needs and formation of her or his cognitive functions. The purpose of play is not in achieving some useful result of activity but rather the activity process itself. However, this fact does not eliminate the goal formation and motivational aspects of activity. The actions of the child are purposeful and goal directed. For example, a child can often see how adult feed a child. Moreover, children have their own experience of being fed. When a child while he/she plays tries to feed a doll she/he try to utilize the prior obtained experience. Despite the fact that a child cannot really feed a doll, she or he still performs conscious goal-directed actions that are integrated in some logical way to achieve the goal that is formulated during the play. Moreover, imaginative aspects of play become critical when children operate with a variety of objects, which have a particular purpose and are associated with objects' meaning.

A child at play operates with meanings that are often detached from their usual objects [5]. For example, a child can take a stick, say that it is a horse, and start performing meaningful actions that are similar to a rider's actions. Play is associated with pleasure and it develops motivational components of activity. Gradually, play becomes more and more important and a child develops rules for her or his play. Subordination to rules makes playing more complex, and play is transferred into a game. The imposed rules force children to learn how to suppress involuntary impulses. Hence, the childhood games precede real adult activities.

In games, the same as in work, tasks can be formulated in advance by others or formulated by a subject independently. Independently formulated tasks exist even in jobs that have rigorous rules and requirements. For example, during a flight from New York to Moscow, a pilot performs not only the prescribed tasks but also multiple tasks formulated independently depending on the weather, information from the air controllers, and so on. Similarly, there are multiple self-initiated tasks during performance of computer-based tasks in productive and non-productive environment. In this type of task, goal formation and motivational aspects of activity are particularly important. Analysis of games and other types of activity demonstrates that goal is one of the central concepts in psychology, but in cognitive psychology, it is mixed with motivation, which is one of the main confusing factors in applied field. For example, Diaper and Stanton wrote [6]:

"The basic is that there is some sort of psychological energy that can flow, be blocked, deviated, and so forth. Goals as motivators of behavior would seem to be a part of this type of psychological hydraulics. Given that there is no empirical evidence of any physical substrate that could function in such a hydraulic fashion, perhaps we do not need concept of goals as behavior motivator."

This is an example of erroneous interpretation of goal that is common in cognitive psychology. In SSAT, a goal is a cognitive component that interacts with motives and creates a vector "motives → goal," or more specifically "motivation → goal."

Similarly, Karat et al. insist that a task in production environment has a clearly intended purpose or goal [7]. According to this author, the HCI field shifts its focus from production environment with its clearly defined tasks and goals to the non-production field, where the main purpose is communication, engaging, education, game, and so on. According to these authors, HCI field shifts its focus from production environment with its clearly defined tasks and goals to the non-production field, where the main purpose is communication, engaging, education, game, and so on.

Karat had wrote: "HCI professionals might say that people use technology because they have the goal of reaching a pleasurable state, but this is awkward and has not proven useful as a guiding approach in design. This is partly because of the difficulty in objectively defining the goal state, and without this there is not much the field can say about the path to the goal." [7].

Here, the authors mix the goal of the task with the motive and insist that there are no tasks and goals in non-production environments and particularly in games. However, in the above example, we have shown that the goal is to reach the desired future result of the game. The motive is to obtain a pleasurable state during a game and some satisfaction after the game. The goal of the game cannot be precise at the beginning of task performance in non-production and production environment. In designing a task in any environment, we should find out how the initially formed goal is gradually clarified and specified during further task performance.

In cognitive psychology, intermittent goals are known as sub-goals. Process of formatting the sub-goals is performed based on means-ends analysis, in which the desired sub-goal is considered to be the end state of the step. This end state is compared with the present state of knowledge. From the SSAT perspective, these data require some additional interpretation because sub-goals are mental representations of desired future results. Hence, a sub-goal is a cognitive and conscious entity. There is also a need for a general motivational state that creates an inducing force to produce this process of sub-goals formation. The comparison of a future hypothetical end state with an existing state is provided by feedback, which is performed in the mental plane. This demonstrates that thinking works as a self-regulative process. Moreover, there are well-defined and ill-defined problems. Well-defined tasks are those that have a clearly stated goal. The performance of ill-defined tasks in more complex situations begins with searching for and formulating the goal. A subject promotes hypotheses, formulates hypothetical goals, and evaluates them mentally or practically. Only after that can he or she formulate hypotheses to achieve a defined goal. Hypotheses formation process can include conscious and unconscious components. Unconscious hypotheses are not verbalized, and they can be performed, for example, in the visual plane. Some of these hypotheses can be later transferred into the verbalized, conscious plane. There are also hypotheses that are conscious during a short period of time, and then they are forgotten. People act as goal-directed systems in work environment or in game situations.

According to Karat et al. [7] and Diaper and Stanton [6], the advent of technology in the home environment and the everyday life of people eliminates task-oriented activities. Moreover, Karat et al. [7] wrote that "the science of enjoyment is not capable to define a goal-directed approach." First, we emphasize the fact that there is no such science as "enjoyment." In psychology, this term simply refers to a certain emotional state. Then, we note that the concept of "task" is very important even for entertainment. Our activity or behavior strives toward anticipated goals in production and non-production environments. The analysis of a variety of tasks that people perform in everyday life demonstrates that they attempt to break down the flow of activity into smaller segments or tasks, which are often self-initiated. Subjects' everyday life activity cannot be understood without referring to motivational and goal-formation processes. In contrast to cognitive psychology, in AT a task always has its desired final goal and motivational forces. Similarly, social interaction, learning, playing, and games are always, as any other type of human activity, motivated and goal directed.

Karat et al. [7] substitute complicated concepts such as motivation by the term "value," which simply has a common sense meaning in their discussion. Hence, the classification of HCI systems as communication driven, content driven and so on is questionable. The authors come up with a new "science of enjoyment," which, in their words, is not a goal-directed approach. It is hard to agree with such an interpretation of enjoyment. People can enjoy drugs, alcohol, work, sports, and so on, depending on the motivational factors. So, the study of motivation should be associated with enjoyment. Is there a need for a new "science of enjoyment" when there is psychology and motivation as its branch? Any technology is just a means or tool for human work or entertainment activity. Hence, we need to study the specifics of utilizing such tools or means of work in various kinds of activities. For instance, in order to design a computer-based system of person-to-person communication, such system should be adapted for social interaction activity, providing means for understanding the partners, prediction of their goal and motivational state, ability to formulate the general goal of subjects involved in such communication, ability to emotionally interact with each other, and so on. Usability engineers should work together with psychologists to improve the design of computer-based tasks in various environments. We can evaluate computer interface in non-production environment only in the context of task performance.

The concept of task and classification of tasks was first developed for production environment. However, the concept of task is critically important also for non-production environment. Classification of tasks for non-production environment is the same as for the production environment. The selection of tasks or development of tasks for non-production environment depends on their purpose in this environment. For tasks that facilitate social interaction for the wide audience, such tasks should be relatively easy. They can be deterministic, algorithmic or even skill-based tasks. In case of game, there is a trend to increase the task complexity or allow for regulating the task complexity based on the type of audience. Complexity of task in non-production environment is used as a motivational factor. Moreover, risk to lose a game is also utilized as a motivational factor.

Therefore, clear understanding of such concepts as task complexity and difficulty of the task is vitally important. Complexity is an objective characteristic of the task. Difficulty is its subjective characteristic. The more complex the task in non-production environment is the higher is the probability that it will be difficult for a subject. Entertainment tasks can be objectively sufficiently complex. However, a complex task might be perceived as a low difficulty task by a specific individual, which would reduce the subjective significance of this task for such individual and she/he would not be motivated to perform this entertainment tasks.

Motivational factors play a particularly important role in a game. Thus, let us consider the stages of motivational process in human activity. According to the concept of motivation developed by in SSAT [3, 8], there are five stages of motivation: (1) the preconscious stage, (2) the goal related stage, (3) the task-evaluative stage, (4) the executive or process-related stage, and (5) the result-related motivational stage. Based to the principle of self-regulation, these stages are organized as a loop structure, and depending on the specificity of the task, some stages can be more important than the others. Depending on task specificity, scientists should pay more attention to some of these stages and their relationship.

The preconscious stage of motivation predetermines motivational tendencies. This stage is not associated with a conscious goal but rather with an unconscious set that can be later transferred into a conscious goal and vice versa. The goal-related motivational stage is important for goal formation and acceptance. This stage can be developed in two ways: by bypassing preconscious stage of motivation or through the transformation of an unconscious set into a conscious goal. When the current task is interrupted and attention is shifted to a new goal, the previous goal does not disappear, but is transformed into a subconscious set. It helps a subject to return to an interrupted task, if necessary, through the transition of a set into a conscious goal.

The third motivational stage, the task-evaluative stage, depicts the evaluation of task difficulty and its significance. The fourth stage, the executive or process-related motivational stage, is associated with executive aspects of task performance. Goal formation, task evaluation (evaluation of task difficulty and its significance and their relationship), and process-related stages of motivation are particularly important for understanding risky tasks, games, and the development of recreational computer-based tasks. The result-related motivational stage is associated with evaluation of activity result (completion of task). All stages of motivation can complement or conflict each other.

Let us consider some examples. The relationship between process-related and result-related stages of motivation is important for production environment. In some cases, the work process itself does not produce a positive emotional-motivational state. This can be observed during the performance of a boring job when the work process–related stage of motivation is negative. In order to sustain a positive motivation during such task performance, commitment to the goal-related (stage 2) and result-related motivational stages (stage 5) should be positive to offset it.

In computer-based games, the process-related stage (stage 4) is critical and should be associated with the positive emotional-motivational state. The result-related stage (stage 5) should vary when positive results are combined with negative results, producing a combination of positive and negative emotional-motivational states. At the same time, the always-positive result in computer-based games can reduce interest in the game.

A simple game without a risk of losing can reduce the positive aspects of the process-related stage of motivation. Hence, the complexity of the task should be regulated depending on the previously obtained results. If the game is designed for children, the possibility to obtain a positive result should be significantly increased. Even in gambling, when people can lose their money, some relationship between success and failure is important. The strength of positive and negative emotions during different stages of motivation is also important. This is particularly relevant for the risk-taking addicted people, where manipulation of process and result related stages of motivation is critical. Of course, other stages of motivation also should be taken into consideration. In non-productive tasks, the simplicity or difficulty to obtain a desired result is an important factor. Understanding motivation as a sequence of interdependent motivational stages helps to create a desired motivational state in the production and non-production environment.

Sometimes, the goal of a task or game is not precisely defined, and at the beginning the goal is presented in a very general form. Only at the final stage of the game, the goal becomes clear and specific. This is not new. For example, when playing chess, the goal can be formulated only in a very general form "to win" or "to tie the game." A chess player also formulates some hypotheses about his or her possible strategies that are closely connected with the goal of the task in advance. For this purpose, a player uses some algorithmic and heuristic rules that he or she stores in his or her memory. When a chess player selects a possible strategy, he or she starts to formulate multiple intermittent goals that correspond to his/her strategy. The selected strategy can be corrected or totally abandoned depending on the strategies of the opponent. A clear and specific understanding of an overall goal is possible only at the final stage of the game, just before a checkmate.

Even when the goal of a task is externally given in a precise form, the subject can reach this goal by using a variety of strategies and various intermittent goals. Therefore, a goal cannot be considered as an end state of the system that the human or machine wishes to achieve, as has been stated by Preece et al. [9]. A goal of a system and a goal of a person are two totally different concepts.

4 Conclusion

Currently, there is no consensus on the understanding of task concept and the main task attributes. Task analysis is a multitude of independent techniques, sometimes not sufficiently grounded from theoretical point of view. There is a reasonable opinion that at this point task analysis is a mess. Moreover, some practitioners raise questions about the future of task analysis.

Hence, it is no coincidence that the concept of a task is rejected by a number of specialists working on non-production related projects. Specialist in ergonomics usually distinguish two types of human activity such as work and learning. The second one is usually considered in the context of training process. However, there is now another type of human activity that is also important with development of entertainment related industry. According to cognitive approach concept of task is redundant in the study of human work. From SSAT point of view concept of the computer-based task is important not only in production but also and non-production environment. Presented work clearly demonstrates that without clear understanding of the concept of task it is impossible to study human activity in non-production environment. Not only human work and learning are important topics in ergonomics. Playing and games are also important areas for studying human activity in the non-production environment.

References

1. Bedny, G.Z.: The Psychological Foundation of Analyzing and Design Work Process. Higher Education publisher, Kiev (1987)
2. Bedny, G.Z., Meister, D.: The Russian Theory of Activity. Application to Design and Learning. Lawrence Erlbaum Associates, Mahwah (1997)
3. Bedny, G.Z.: Application of Systemic-Structural Activity Theory to Design and Training. CRC Press/Taylor & Francis Group, Boca Raton/Abingdon (2015)
4. Rasmussen, J.: Skills, Rules, and Symbols, and Other Distinctions in Human Performance Models. In: IEEE Ttransaction on Systems, Man and Cybernetics, SMC-15, pp. 234–243 (1983)
5. Vygotsky, L.S.: Mind in Society. The Development of Higher Psychological Processes. Harvard University Press, Cambridge (1978)
6. Diaper, D., Stanton, N.: Wishing on a sTAr: the future of the task analysis. In: Diaper, D., Stanton, N. (eds.) The Handbook of Task Analysis for Human-Computer Interaction, pp. 603–619. Lawrence Erlbaum Associates, Mahwah (2004)
7. Karat, J., Karat, C.M., Vergo, J.: Experiences people value: the new frontier for task analysis. In: Diaper, D., Stanton, N. (eds.) The Handbook of Task Analysis for Human-Computer Interaction, pp. 585–602. Lawrence Erlbaum Associates, Mahwah (2004)
8. Bedny, G.Z., Karwowski, W., Bedny, I.S.: Applying Systemic-Structural Activity Theory to Design of Human-Computer Interaction Systems. CRC Press/Taylor & Francis Group, Boca Raton/Abingdon (2015)
9. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human-computer interaction. Addison-Wesley, Reading (1994)

Activity Approach in Management

Fred Voskoboynikov(✉)

Baltic Academy of Education, Dekabristov 35, 190121 St.-Petersburg, Russia
fredvosko@hotmail.com

Abstract. Regardless of the field of activity, managers work with people. Not only those, who are called managers by their title, belong to the category of managers. Business owners and plant foremen, supervisors and heads of departments, commanders of military units and coaches of athletic teams, and many others - all fall into the category of managers by their functions. In our work, we will describe the most important managerial functions of the psychological nature and make an attempt to shed light on a new understanding of these functions from the activity theory (AT) perspective.

Keywords: Management · Activity theory · Strategies of performance · Individual style of activity · Self-regulation of activity · Formal and informal groups · Compatibility factor

1 Introduction

Historically activity theory was developed in the former Soviet Union as a counterweight to behaviorism, a theory developed by the American psychologist Skinner [1]. By Skinner, human reactions are considered to be the results of only external influences, ignoring mediated functions of activity that provide a basis for personal development. In activity theory, a person interacting with a situation is considered the subject who performs conscious voluntary actions. The leading Soviet psychologist Sergey Rubinshtein was one of the main contributors into the development of activity theory. He wrote, "through the organization of the individual practice, society shapes the content of individual consciousness" [2]. His famous quotation "external acts through internal" emphasizes the dependence of activity on the subject's individual personality features. The social aspect depends on the individual, just as individual depends on the social aspect. In the same social environment different individuals act differently and are impacted by social environment in different way.

Activity theory distinguishes two types of activity: "object-oriented" and "subject-oriented". The former is referred to a subject using tools on material objects with the goal to complete the task and evaluate the results. The latter is referred to social interaction between people, which is the most important element in management. According to the systemic-structural activity theory (SSAT) human activity is considered to be a goal-directed self-regulative system [3]. It manifests itself in the way people through trials, errors, and feedback corrections create strategies of performance which are derived from the personality features. Activity consists of actions that could be cognitive/or internal/and behavioral/or external/ [4]. All actions are organized and

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_36

directed toward the achievement of conscious goals of the task. Activity can be defined as conscious, intentional, goal-oriented and socially formed behavior which is specific to humans.

In this brief presentation, we will consider some important factors of the psychological nature which managers should take into account in their work with people.

2 Management is a Multifaceted Human Activity

Most of attention in any organization is directed towards achieving financial goals, i.e. towards profitability. This is vital for the organization and well understood. However particularly for this reason people's interests are not often on the priority list in organizations' affairs. If that is the case, sooner or later such an approach will backfire and prevent the organization to function successfully in the long run. Hence, directing all possible efforts towards creating a positive psychological environment in the workplace is of a significant importance. To create such an environment without basic knowledge of psychology does not seem possible. To know people's individual personality features, their ability to work in a group environment as well as their values, goals and desires is just as necessary for managers as the technical knowledge in the chosen field of activity [5].

Activity approach in management manifests itself in many ways. Knowledge of subordinates' personality features and the ability to apply this knowledge in the process of management allows to bring out people's best. Understanding group dynamics and the distribution of roles in the group is as important for the effective implementation of managerial functions. To be aware of the presence of informal leaders, as the most influential group members, and communicating with them in the process of management helps maintaining a positive psychological climate in the work place. The compatibility factor among the group members and its direct effect on the group performance is another factor of the psychological nature which managers should work on in the process of management. Providing clear and systematic communication with subordinates keeps managers in the loop of the work progress in the ongoing projects and allows to anticipate possible missteps and make the necessary corrections in advance. Delegating of some managerial functions to subordinates enhances productivity and serves as an excellent tool for the subordinates' professional growth.

To work on creating a cohesive team is perhaps the most important thing for managers to do in order to achieve their teams' objectives. From the social psychology perspective a team is a small social group of persons who unite and cooperate to achieve the common goal. For the successful functioning of the team two factors is of the most importance: team members must possess the needed technical skills and experience in the field and complement each other. The first factor is usually well taken into account, but the second one is not always paid much attention to. To create the cohesiveness of the team is easier when each team member is at his or her respective place according to their best qualities. Each person has a unique personality; some people are good in some things, while others are good in something else. Hence, the golden rule in dealing with people is not to try to change people, but build on what they are and compensate for what they are not. To give a person the wrong role is like to ask him to be what he is not. When he is pressed to be what he is not, he does not feel good and does not perform well. When he is placed where he feels “in his shoes”, everything changes – he feels good, his productivity increases and all the rest comes with it. Then people around him are amazed about the changes in him. But he has not changed, he became himself. The late American psychologist Morris Viteles, who is considered as one of the fathers of industrial psychology and an enthusiast of taking the human element into the practice of management, wrote: “It is important that a man be kept out of a job for which he is not fitted. It’s even more important that he be placed in a job where he can be efficient and happy.” Such an approach should be a sort of the guiding star for managers in their work with people.

Managers should act as conductors, not as drill sergeants. Why is it important? The submission is quite rarely pleasant, any overbearing tone is perceived as a suppression of personality, as an encroachment on the individual freedom. Demands of blind obedience and underestimating of subordinates’ initiative and abilities is perceived as abnormal and determines the corresponding attitude towards the manager. Requirements expressed in the form of proposal perceive as more acceptable and tend to have a greater effect. By not showing the superiority, the discussion goes on an equal footing rather than direct criticism and instructions. By acting in this manner managers show respect to people, value their competence and psychologically put them on the level as they are. We have to note however, that it does not need to be construed as an endless idyll. When circumstances dictate, clear and direct instructions are justified in order to achieve the team’s objectives.

3 The Individual Style of Activity

There are two ways to ensuring the effectiveness of human performance. One is by professional selection, the so-called “screening out” of individuals with specific attributes. The other one is the individual style of activity which is based on personality features. All kinds of work, learning and athletic activity are characterized by the interaction of subjective personality features and objective requirements of activity. This interaction goes in two directions. The first one is the adaptation of objective requirements of activity to the subjective properties of the individual, the second – in the adaptation of subjective properties of the individual to the requirements of activity.

The concept of the individual style of activity was first introduced by the Soviet psychologists Merlin [6] and Klimov [7]. They were able to establish that different individuals can perform the same work equally efficient through the use of their own individual style of performance, which is more suitable to their personality features. That is, people attempt to compensate individual weaknesses with their personal strength in a given task situation. By implementing the individual style of activity they diminish the impact of negative personality features on performance and thus adapt to the situations easier and perform more efficiently.

Any kind of activity require a number of qualities from a person in order to perform. Some personality features better relate to the requirements of activity, others - not as much. It enables people to compensate their weak qualities by the more outstanding ones. It suggests that managers should rely on people’s strong qualities instead of insisting on fixing the weaker ones. As a result, managers will best benefit from what people are capable of and they will experience satisfaction by their performance. The strategy to rely on individuals' personal strength, as to compensate for individual weaknesses, occurs at the conscious and unconscious levels and is based on the principles of self-regulation [8]. Both levels are tightly interconnected and transform from one to another. The process of self-regulation manifests itself in a formation of desired goals and in developing a program of actions which correspondence with these goals, with conditions for achieving the goals, and with persons' individual abilities. Other words, people through trials, errors and feedback corrections create strategies of performance suitable to their individuality. For example, people with an inert nervous system develop a predisposition to organize and plan their work in advance and attempt to utilize a stereotyped method of performance.

Thus, the individual style of activity should be considered as strategies of performance deriving from the mechanism of self-regulation which depend on personality features [9].

4 The Group Environment and Individual Performance

Usually, people's activity unfolds in social environment. It has long been observed that people behave differently in a group setting as compared to the behavior in private. This is because an individual in a group appears in a new capacity – as a component of the system "individual – other individuals". Groups have properties of their own; they are different from the properties of the individuals who form the group. Just as two dozen of clear fragments of glass stacked one on another provide a rich blue color or a combination of copper and tin results an alloy, which hardness is neither of each of them, people in the group act and behave differently. In most cases, the presence of the group has a positive effect on the individual behavior and performance.

People's relationship in groups can be on the level of functional business contacts and on the levels of psychological human contacts. If a group is formed for a certain purpose – production, education, military, recreational activity and so on – they are called *formal* groups. Examples of such groups are a factory shop, a production unit, a school class, an aircraft or a ship crew, an athletic team and so on. Other groups are formed on the basis of personal relations, subjective feelings, sympathy, trust, common interests and so on. These groups are called *informal*. For example, a group of friends is a group of persons who are pulled together by common amateur interests. The structure of any informal group in general is as follows: leader, followers, and isolated (rejected by the group or rejecting the group).

From the point of view of management it is important to understand that psychological connections between people take place not only within informal groups. People built informal connections by functioning in formal groups, in the groups which are formed for certain organizational goals. That is, in every working unit there are two structures – formal (official) and informal (unofficial), and each structure has its respective formal and informal leaders. The formal structure is based on formal rules and written instructions, where the circle of obligations for each employee is clearly defined. The informal structure within the formal groups is a system of psychological connections between the group members. This structure is particularly important for our discussion. Every production unit is people, who are in the process of joint activity project psychological feelings towards each other – sympathy or antipathy, converge or diverge in tastes, personal interests, ethnic preferences, amateur affiliations, and so forth. The manifestation of these feelings result in either mutual attraction or repulsion. Precisely these factors affect emotional well-being of the team members which has a direct influence on their satisfaction in the workplace and, ultimately, on their productivity. People come to work expecting a colleague-friendly work atmosphere. The sense of belonging to the informal group gives a worker certain status and recognition, creates the feeling of the relation to others; the feeling that she is somebody, even though in the formal structure she is just one of many. The effectiveness of group and individual performances is largely determined by the conformity of formal and informal structures. Simply put, if people at work evoke positive perception of each other, the process of activity runs more effectively on the background of positive emotions.

In any group activity the question of compatibility arises. Not by chance someone said about ancient Egypt's pyramids, "Those are not just bricks put together; it's about *how* they were put together." The same approach should be used with the respect to the execution of group tasks. That is, not only the presence of needed specialists with their technical skills should be taken into account, but the degree of compatibility between them as well. Depending on the degree of compatibility the result of group performance may either be equal to the sum of the results of individual performances, or greater or lower than the sum. This suggests that the group is not the arithmetic sum of separate individuals, but rather a single organism – a whole – and the result of the group performance is not always a positive sum of the results of individual performances. An example of the incompatibility can be a working crew, where there is a significant difference in skills of workers requiring coordinated application of muscular effort or relatively accurate movements. This kind of incompatibility is called *physiological*. In this example such physical parameters as physical strength and motor skills are described. To note such differences in people is not that difficult and it's unlikely that anyone will instruct people with such differences to perform a task where these differences present hindrance. People always experience certain flow of feelings toward others within the group. These feelings are based on the commonality or differences of the psychological nature, such as temperament, character, social orientation, amateur interests, religious and ethnic peculiarities and others. They may be positive or negative, or neutral, they can be weak or strong in intensity; they can be mutual or non-mutual and therefore conflicting. The differences of this kind are not always obvious and apparent. However, particularly differences of this kind quite often have a decisive impact on compatibility, and, in turn, on the implementation of the group task. The incompatibility by the described differences is called *psychological*. The presence of psychological incompatibility is a major obstacle for the effective group performance.

Psychological incompatibility has its negative influence not only on the group performance, but also on the human health. Unfriendly uptight relationship between group members in the working environment calls up negative emotions. In mass professions where there are no expressed extreme conditions people can perform productively under the influence of negative emotions for a fairly long time. However, it's important to understand that it flows at the expense of the unnecessary stress, "until then, until the time" so-to-say. Many can recall the depressing mental state due to the incompatibility with colleagues or bosses at the current or previous job. Working activity on the background of negative emotions for a long period of time may cause pathological developments in the central nervous system, which could lead to various diseases of a neurotic order. People become irritable, experience headaches, insomnia, blood pressure disorders, dysfunction of gastrointestinal tract, and other deviations in health condition. Typical medical approach for the treatment of such conditions does not always give positive results. There are statistical data in different countries on the loss of a huge number of man-hours as a result of the nervous breakdown due to the psychological incompatibility.

Psychological compatibility is the most crucial factor when activity takes place in the extreme conditions. Examples of such conditions among others are: situations of danger, time limit, extreme cold or hot temperature, limited space of activity, and so on. In such conditions fuzzy coordination between the group members due to the insufficient compatibility may endanger people's lives. The activity of aircrafts and submarines' crews, polar explorers and mountains climbers flows in such conditions.

5 Clear and Systematic Communication

Communication is a two-way process where both parties want to communicate, otherwise it is not communication. Communication process takes place when each party makes an effort to understand what the other party is trying to communicate. In some activities the value of clear communication cannot be overestimated, for example in communication between pilots and air-traffic controllers. If a command or confirmation of the command is not understood correctly by either of them, it may lead to a serious and sometimes even tragic consequences. Here is a real-world example of how miscommunication between the air-traffic controller and the pilot has led to the tragic end.

On March 5, 1973, an aircraft Boeing 747-249F of the Flying Tiger Line was on the way to the Kuala Lumpur Subang airport (KUL). Air-traffic controller gave the command to the pilot of the approaching aircraft: "Get into echelon (go down) "two-four-zero-zero" (2400"). Because of the similarity in the pronunciation of numeral "2" and the preposition "to" in English language the pilot confirmed the command incorrectly.. He responded: "OK, to four-zero-zero" (that is "to 400", which sounds similar to "two-four-zero-zero"). The air-traffic controller did not catch the mistake and the aircraft crushed into a hill. Among recently introduced FAA rules and regulations some command were modified in order to prevent possible miscommunication. For example, they changed air-traffic controllers command "affirmative" to "affirm", because possible background noise in the microphone may cover first part of the word and pilot may heat the "tive" which may be understood as "negative". Some other commands were also modified [10].

The described is an example of the extreme situation, but in the ordinary everyday business affairs clear and proper communication is as important. In 2009, after the Financial Crisis of 2005–2007, Government launched the program to help homeowners with their mortgages. Under the program, banks reduced the interest rate to qualified borrowers and thus their monthly mortgage payments were reduced accordingly. The homeowners sighed with relief. But unfortunately that was not the end of the story. In some banks, the department which offered mortgage relief to homeowners failed to communicate with the department in charge of implementing foreclosure. As a result, there were cases when the foreclosure departments seeing that homeowners were making lower than their original monthly mortgage payments, assumed that they defaulted ... and seized their properties.

One of the most common blunders of communication in management is, when the manager assumes that everything was going well because he did not hear anything. Such an assumption is a clear evidence of the manager's rare and inefficient communication with subordinates. If subordinates are not getting regular check-ins of their work, they won't know if everything is going in the right direction. If the manager, on the other hand, is not getting regular feedback from subordinates, it will be difficult for her to track the progress of the work. Maintaining regular and accurate communication with all the people involved in the ongoing projects is essential for ensuring the smooth flow of the work progress. That, in turn, allows anticipating possible missteps and making the necessary corrections in advance. In order to maintain such a business environment managers must provide appropriate and timely information to subordinates, so they will know what they should do, when they should do it, and what is expected from them in general in the framework of the work requirements.

6 Conclusion

In the presented material we introduced the basic of activity theory as a counterweight to behaviorism, which portraits humans as reactive organisms. Behaviorists considers human reactions as the results of only external influences, such as punishment or reinforcement. Mediated functions of activity, that provide a basis for personal development, is completely ignored. In activity theory, a person who interacts with a situation is considered the subject; that is, we are talking about external behavioral and internal cognitive actions as components of cognition, and not about reactions to stimuli. Activity theory distinguishes two types of activity: "object-oriented" and "subject-oriented". The former is referred to a subject using tools on material objects with the goal to complete the task and evaluate the results. The latter is referred to social interaction between people, which is the most important element in management. We described some of the most important managerial functions, directed on implementing of the factors of the psychological nature. Based on activity approach these functions can be effectively materialized in the process of management. We also demonstrated, that according to the systemic-structural activity theory, human activity is considered to be a goal-directed self-regulative system. In particular, we noted that individual style of activity allows the subject to adapt to the situation more efficiently because it connects features of personality with mechanisms of self-regulation and strategies of performance. We discussed the influence of group environment on individuals' behavior and performance. And, finally we underlined the importance of active and systematic communication between the manager and subordinates.

References

1. Skinner, B.F.: About Behaviorism. Knopf, New York (1974)
2. Rubinshtein, S.L.: Foundations of General Psychology. Academic Science, Moscow (1946)
3. Bedny, G.Z., Karwowski, W.: A Systemic-Structural Theory of Activity Application to Human Performance and Work Design. Taylor & Francis Group, Boca Raton (2007)
4. Bedny, G., Karwowski, W., Voskoboynikov, F.: The relationship between external and internal aspects in activity theory and its importance in the study of human work. In: Bedny, G., Karwowski, W. (eds.) Human-Computer Interaction and Operators' Performance, pp. 31–59. Taylor & Francis Group, Boca Raton (2010)
5. Voskoboynikov, F.: The Psychology of Effective Management: Strategy of Relationship Building. Routledge, New York (2017)
6. Merlin, W.: Outlines of Integral Study of Individuality. Pedagogy, Moscow (1986)
7. Klimov, E.: Individual Style of Activity. Kazan University Press, Kazan (1969)
8. Bedny, G.Z., Voskoboynikov, F.A.: Problems of how a person adapts to the objective requirements of activity. In: Aseev, V.G. (ed.) Psychological Problems of Personality, vol. 2, pp. 18–30. Irkutsk University Press, Irkutsk (1975)
9. Voskoboynikov, F.: The influence of personality features on performance in work, study and athletic activity. In: Marek, T., Karwowski, W., Frankewicz, M., Kantola, J., Zgaga, P. (eds.) Human Factor of a Global Society: A System of Systems Perspective, pp. 187–192. Taylor & Francis Group, Boca Raton (2014)
10. Makarov, R., Voskoboynikov, F.: Methodology for teaching flight-specific English to nonnative English-speaking air-traffic controllers. In: Bedny, G.Z., Karwowski, W. (eds.) Human-Computer Interaction and Operators' Performance, pp. 277–304. Taylor & Francis Group, Boca Raton (2011)

Cognitive Computing and Internet of Things: Techniques and Applications

Personalized Instructions for Self-reflective Smart Objects

Jan Pascal Maas, Daniel Burmeister(✉), and Andreas Schrader

Institute of Telematics, University of Lübeck, Lübeck, Germany
jan.maas@student.uni-luebeck.de,
{burmeister, schrader}@itm.uni-luebeck.de

Abstract. Due to ever increasing integration of networking and processing capabilities into daily life objects, the Internet of Things (IoT) will affect evermore activities of daily living. These changes require new solutions for improved human-computer interaction. Especially the usage of device ensemble poses entirely new challenges for usability. Currently, we are developing a framework targeting the self-reflection of smart devices in order to counteract such challenges by generating and delivering usage instructions. In this paper, we extend the proposed architecture to consider user preferences during the process of instruction generation and delivery. Besides a user representation on the basis of stereotypes, multiple rule sets are used to identify parameters that are relevant for the whole process. The mapping from users to a matching stereotype relies on the usage of different similarity metrics, among which the Euclidean distance achieves the best results on the basis of a scenario-based evaluation.

1 Introduction

In recent years, networking technologies have become an immersive factor for technical development. Especially the interconnection of everyday devices leads to a new paradigm of creating a global infrastructure that provides users with additional services [1]. Furthermore, studies suggest that the Internet of Things (IoT) will affect evermore activities of our daily living, accumulating approximately 50 billion devices to be used [2]. Additionally, common Human Computer Interaction (HCI) will be characterized by the simultaneous use of multiple devices, so called *Ensembles* [3]. These advances in the area of the IoT are driving forces for the distribution of ubiquitous and pervasive computing, creating a variety of new opportunities in the development of smart objects and environments [4]. Concurrent to the arising possibilities, a consequence of these paradigms is the emergence of new challenges. The wireless interconnection of devices, the loose coupling of interaction possibilities and functionality and the growing usage of Natural User Interfaces (NUI) offer new opportunities for augmenting the environment with intuitive to use services. On the other hand, they are also increasing barriers for users for learning and remembering the interaction with smart objects and environments, since established concepts of usability such as *affordances* are based on the visibility of interaction capabilities [5, 6]. To address these challenges in smart environments, we've developed the approach of *self-reflection* which enables interconnected devices to receive knowledge about their specification as well as provided

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_37

interaction capabilities in combination with triggered changes of state [7]. In consequence, a self-explanation can be achieved, which may reduce the learning periods for users to interact with a device ensemble. In order to put this approach into practice, our institute currently develops the "Ambient Reflection" framework [7]. However, for improving the learning result, users' needs and preferences have to be considered during the generation and delivery of instructions about device ensembles. On this account, an additional component is proposed here, taking the users' needs into account. Therefore, we have analyzed multiple approaches to evaluate user needs alongside with different user profile approaches. Additionally, it is necessary to interpret a user's profile to extract relevant information and transform given needs into parameters which could be processed to personalize instructions. This conversion may be realized by using rule-based description languages such as the *Business Process Execution Language (BPEL)* [8]. In the following, the Ambient Reflection approach and the proposed framework are outlined. Afterwards, an analysis of various approaches for user profiles and rule-based evaluation engines based on a given context are presented. Section 4 introduces the *User Context Classifier* and *Delivery Coordinator* responsible for the personalization of instructions. Finally, both components will be evaluated by discussing different approaches of analyzing given user profiles and the results of a newly proposed delivery coordination process.

2 Ambient Reflection

To realize the approach of self-reflection, it is conceivable to relieve a user cognitively by utilizing autonomic *self-organization* of interconnected devices targeting the aforementioned challenges of usability [7]. Additionally, a *self-description* containing available interactions and corresponding device components based on the current interconnection state is necessary [7]. By transferring the programming concept of *reflection* [6] to self-descriptions, self-reflection goes beyond the distribution of a priori information on system reactions as a result of an interaction. Furthermore, it realizes a self-description of devices at runtime on the basis of provided interaction possibilities and their reactions. If such process take place in the ambience of a smart environment, the self-reflection of interconnected devices is called *ambient reflection* [7].

2.1 Ambient Reflection Framework

In order to realize the concept of ambient reflection, we presented a self-organizing system architecture [7]. Basically, the architecture consists of so called *Smart Objects* and a central control unit, named *Description Mediator*. Smart Objects are given a structured self-description containing knowledge about their functionality and associated interaction capabilities which is provided in the *Smart Object Description Language* [9, 10]. The Description Mediator acts in the sense of an Ambient Intelligence (AmI) to coordinate the exchange of information and thus is responsible for generating and delivering the instructions to a certain user [11]. It uses mainstream networking protocols to collect self-descriptions of current network participants. Afterwards, involved self-descriptions are aggregated based on the current interconnections state of devices and are further used as a generic reference for instructions. Based on user needs that are processed in the *User Context Classifier,* the process of generating instructions is adjusted. Additionally, if instructions are requested by a user or an external service (e.g., depending on the current context), the *Delivery Coordinator* determines ideal presentation devices within an ensemble on the same basis to ensure that the fitting instruction is presented in the best possible way to the user.

3 Analysis

To realize such user specific component, requirements based on the application's context should be met to ensure a correct user representation and transformation to personalized, need-based instructions [12]. To define such requirements, we discuss the usability context on the basis of relevant, but still flexible contextual factors for adaptive systems defined by [4]. This approach would also allow for a simple integration of new aspects.

3.1 Usability Context

Since the presented framework targets smart objects, it will be mainly used within heterogeneous and dynamic environments. Furthermore, the implementation of IoT features causes a fragmentation between conventional and smart devices. Within such environments, the availability and accessibility of information regarding the system usage is essential [5]. In order to display such information, adaptive applications are required, since devices are increasingly embedded and therefore not visible or recognizable to the user and former working concepts like perceived affordances are not working as intended [13]. These issues are also reinforced by the usage of NUIs, since the amount of possible interactions is increased largely and could lead to the irrational use of smart devices due to missing ways to learn the correct interaction paradigms [14, 15]. Additionally, the framework is entitled to support a heterogeneous user group containing different ages, levels of affinity to technology and diverse impairments which requires user information before adapting the system. Besides, the application's users can take several roles while using the system. As a *user*, they may be supported with instructions and set preferences. To map these to the personalization process, so called *rules* should be specified by an *expert*. It is expected that experts have knowledge in the area of media psychology to create the best fitting rules. Even without the extent proposed in this work, the Ambient Reflection Framework shall be used to support the (instrumental) activities of daily living as defined by [16]. As these are part of the daily living schedule, they become a routine and may not only be executed consecutively, but also in parallel [17]. In order to meet such heterogeneous requirements, the user should be able to decide, if the framework provides instructions either initiated by a user or proactively without requiring a user's interaction. This approach allows to fully retain control of the system on behalf of a user.

3.2 Requirements for Personalized Instructions

To create an adaptive application on the basis of this usability context, it is necessary to access user preferences and needs before transforming them into information and adapting the system. In the following, common *user profile* techniques are considered. Based on the usability context, it should be possible to model preferences of a user within the profile to consider them while generating personalized instructions. Also, the profile should be simple enough to be easily editable and enhance able whilst manageable for users with low affinity to technology. Furthermore, the proposed component shall lack demographic information to prevent users from mistrusting the application. Also, such information will not provide any advantages as the user audience is very diverse. To transform user information into parameters to generate personalized instructions, *rule-based description languages* are a conceivable approach. Such languages need to be capable of transforming a user's profile including all of its preferences into matching parameters in an unambiguous way. To easily enhance the system and adjust generated instructions, the languages shall provide "rules" that are easily editable and lack complexity. Another important issue is that most rule-based languages rely on an engine to be run. Hence, a local realization without requiring external suppliers is crucial to ensure, that no informational data is leaving the framework.

3.3 User Profiles

For creating user profiles, multiple approaches with different priorities are imaginable. To embed user profiles into various contexts, *ontologies* may be used. An approach by [18] creates a detailed user profile which is easily extendible to a class based structure. In its core, this ontology features a *Person* that is focused on demographic information about possible impairments, but may also contain user preferences. Since such profile is structured as an ontology, its complexity regarding the implementation is very high. Another approach by Casas et al. focusses on elderly and impaired people by creating a user profile based on personas on the basis of the user's characteristic attributes [19].

It mainly contains information about user levels and audiovisual impairments such as volume and magnification of displayed content without utilizing demographic measures. Still, the profile does not feature user preferences that exceed input and output devices. Every attribute of this profile is featured with detailed measures and multiple inquiries would be necessary to properly fill this profile. A third approach that may be realized easily in comparison to the previously discussed ones is based on the usage of stereotypes [20]. By utilizing these to describe users, no detailed user profiles but only raw characteristics are proposed that could be used for a basic personalization on the basis of few information about a user. Still, a properly detailed personalization requires a sufficient number of stereotypes and diverse criteria. This process requires different strategies in each context. The results of this analysis are shown in Table 1.

Table 1. Analysis of user profile approaches.

	Ontology	Stereotypes	Personas
Waiver on demographic information	✗	✓	✓
Mapping of preferences	✓	✓	(✓)
Complexity	high	medium	low to medium
Extensibility	✓	(✓)	(✓)

✓ = fully satisfied (✓) = partially satisfied ✗ = not satisfied

3.4 Rule-Based Description Languages

To transform user information into personalized instructions, multiple approaches from the field of rule-based description languages could be used. One evermore popularized strategy is the *Trigger-Action programming* like realized by the commercial service *IFTTT*[1] [21]. Due to the simplicity of used commands - a condition triggering a certain action -, even users with a low level of affinity to technology are capable of creating such rule sets. Still, this is one of the major reasons for the approach being limited according to completely considering given user information. With extending the Trigger-Action approach by intermediate configurations, the IoT middleware *Node-RED*[2] minimizes the limitations of user information consideration. Still, the barrier of complexity is not increased to a higher level due to the availability of a graphical user interface (GUI) that allows simple configuration access. Furthermore, the amount of configuration options in form of *nodes* may be increased nearly infinite by just creating new modules containing new possibilities. A different, more powerful approach are Business Process Modelling Languages like BPEL or the XML Process Definition Language (XPDL)[3] since they are capable of describing nearly every level of detail in very complex structures.

Such high complexity reduces the extensibility of rule sets. Still, the use of GUIs would lower the barrier and increase the amount of options available to create rule sets by hiding complexity. The results of this analysis are shown in Table 2 and consequences are discussed further.

Table 2. Analysis of rule-based description languages.

	Trigger-action	BPM-standards	Node-RED	Xtext
Completeness	(✓)	✓	✓	(✓)
Unambiguity	✓	✓	✓	✓
Complexity	low	high	low to medium	medium
Extensibility	(✓)	✓	✓	✓
Locally possible	✗	✓	✓	✓

✓ = fully satisfied (✓) = partially satisfied ✗ = not satisfied

[1] http://www.iffft.com, accessed: 10.03.2017.

[2] http://www.nodered.org, accessed: 10.03.2017.

[3] http://www.xpdl.org, accessed: 10.03.2017.

4 Concept and Realization

The given analysis provides knowledge to adjust the concept that basically consists of three parts: First, the user needs to input information about her or his needs in form of physical capabilities like ametropia or preferences of media types. Based on the given input, a user profile is created which can be used to derive parameters that are sufficient to create personalized instructions. Finally, the given user information is taken into account to decide which instruction should be delivered to what device in which specific way. Figure 1 shows the general process of gathering user information, creating stereotype and parsing the given input into a need-based instruction.

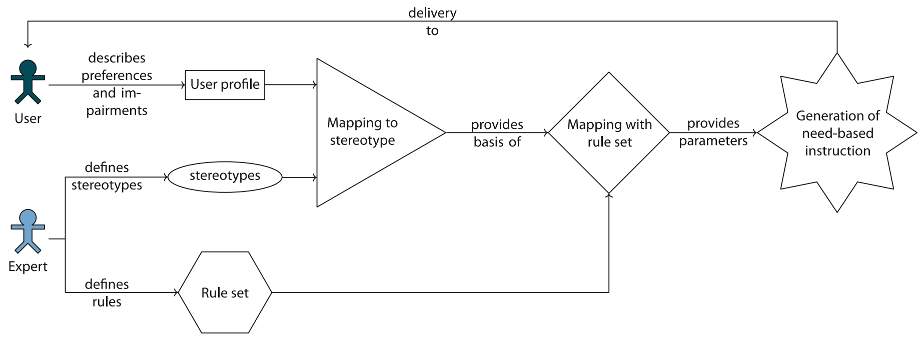


Fig. 1. Schematic representation of the generation process

4.1 User Profile

After considering multiple user profile approaches in Sect. 3.3, a hybrid approach based on stereotype and an adapted version of the user profile proposed by [19] will be used. An ontology provides too many features that are in need of being implemented and includes demographic information about the user. Therefore, it is not suitable for the given use case. The user should be capable of describing her or his preferences with the proposed user profile, therefore the profile by [19] is enhanced by an attribute *Preferences*. This contains information on how the instructions should be delivered on what devices in which matter. Regarding so called *device groups* and *output options*, multiple values can be used by defining an order in form of a sorted list. Additionally, the user is enabled to set the systems *initiative* to a proactive, reactive or interactive paradigm. It should be noted that a user may also ignore this preference by setting a "Don't care" option. Also, the usage of this attribute leads to dropping the attribute *Interface* proposed by [19] hence it is not detailed enough to contain required information. All other attributes are considered for creating and matching a number of stereotypes that are used to create personalized instructions. To enable this process, stereotypes have to be created beforehand by an expert as described in Sect. 3.1. After the creation of stereotypes, the user's needs have to be matched. To do so, a specific metric is used to compute a distance between the user information and the stereotype profile [22]. To be capable of utilizing these features, all profiles are normalized to numerical vectors on the basis of a fixed attribute scale. For evaluation purposes, four metrics have been implemented that may be used to compute the distance: The Cosine, Euclidean, Manhattan and the Chebychev distance. While the Chebychev distance was chosen in contrast to the others because it only considers one of the attributes, all other metrics were chosen because they considered all attributes in different ways that may end up in various matching results.

4.2 Transformation in Personalized Instructions

With regard to the generation of personalized instructions, the users' needs and capabilities represented by his profile were matched to a stereotype, defined by an expert. Additionally, experts should be able to enrich the existing information to tailor and therefore personalize instructions from an expert view. As analyzed in Sect. 3.4, rule-based languages to describe the appearance, point and form of presentation enable experts to define such information. Although, most rule engines have a high complexity and operations that are possibly hard to enhance, the Trigger-Action programming paradigm is easy and enhanceable in many simple ways. In combination with an easy-to-use GUI and intermediate configurations as provided by Node-RED, this approach can be very powerful and comprehensible for non-technically-minded users. Therefore, we decided to use the IoT middleware to provide the parameters needed for creating a personalized instruction based on expertise as described in Sect. 3.1. For this, Node-RED has been enhanced with multiple, newly developed nodes to map the attributes from a stereotype to a so-called *render configuration,* which is the basis to create instructions. Each node provides one input and output pipeline that is used to connect all nodes with one another to so-called *flows*. Additionally, only framework communication nodes are mandatory which means that the transformation may also be created by passing through the given stereotypical values to the render configuration. The communication with the Ambient Reflection framework is ensured by two mandatory nodes that provide REST endpoints inside the Node-RED environment. Furthermore, three optional node groups have been added that may be combined in any possible way as a toolbox for experts to adapt the generation and delivery process. The first group directly influences the system's initiative when interacting with the user. Therefore, the provided nodes allow to configure the initiative to a reactive, proactive or interactive way in which the user is treated accordingly by the system. Additionally, a second group of nodes focusses on the delivery target by influencing the device group that shall be used. Hence, a new device group shall be considered for delivery coordination if such node has been used. This behavior may be enhanced by the third group that is capable of adjusting the output option.

4.3 Delivery Coordination

Based on the aforementioned render configuration, the delivery of personalized instructions using available appropriate output devices in the same or via an alternative ensemble is coordinated. Firstly, under consideration of the given parameters, any device that is known by the Ambient Reflection framework and providing specific output options is considered as a valid device to display instructions. Secondly, the instructions are mapped based on devices that are capable of delivering it. Based on the device that should be supported by the use of instructions, a delivery device should be found. For this, the output devices capable of delivering the necessary instruction are filtered based on the information provided by the render configuration like preferred output options and device groups. If either of these criteria could not be fulfilled, the search for devices falls back to the remaining filters to consider as many as possible. As an example, one could imagine a spoken instruction delivered to the next available speaker set instead of a multimedia presentation on a display, if the user is blind.

5 Evaluation

Based on the provided implementation, multiple evaluation scenarios that may also contain a user-based approach are considerable. Still, such a procedure would suffer from a too small amount of configuration options that would be tested. To eliminate this flaw, we decided to use a scenario-based evaluation approach [23]. All implemented distance metrics, described in Sect. 4.1, were evaluated based on their subjective correctness of matching a stereotype to a given user profile. Additionally, an evaluation of the influences provided by different user preferences and information have been performed.

5.1 Comparison of Multiple Distance Metrics

To analyze multiple distance metrics, some different measures have been taken. First, the robustness of each metric was tested by creating the power set of all possible user profiles that were unique after normalizing and matching each of these to the power set of stereotypes. Based on this, each metric besides the Cosine is capable of matching all user profiles to the correct stereotype, respectively. The Cosine metric suffers from design issues if one of the profiles is equivalent to the zero vector. In the following, we increased the amount of user profiles that do not have an equivalent match inside the stereotypes and matched this newly created set to six user profiles defined in advance. These match common average, maximum and minimum values for each attribute. The distance between each stereotype was subsequently increased by only keeping every second to sixth stereotype inside the matching set. Since these measures are particularly unlikely to happen in real scenarios, three realistic scenarios with distances of 500; 1,000 and 2,000 stereotypes missing between each considered stereotype from the inaugural set of 31,104 have been used. It turns out, that all metrics except the Chebychev distance are nearly equally capable of matching a large number of user profiles correctly to a best-fitted stereotype. Most outstanding were the Manhattan and Euclidean distance which resulted in the identical set of matched stereotypes in the first, unrealistic sets. In the realistic sets though, the Euclidean distance emerged by a few percent which makes this metric the best in the current circumstances. However, it did not match the anticipated results, since the Manhattan distance is expected to be more

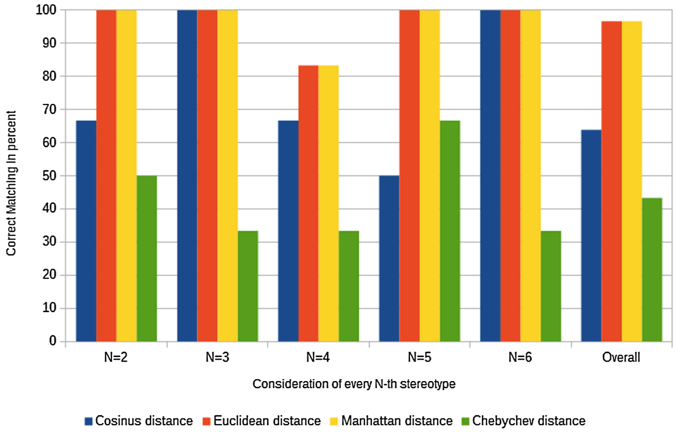


Fig. 2. Mapping results based on the evaluation of distance metrics with theoretical test cases

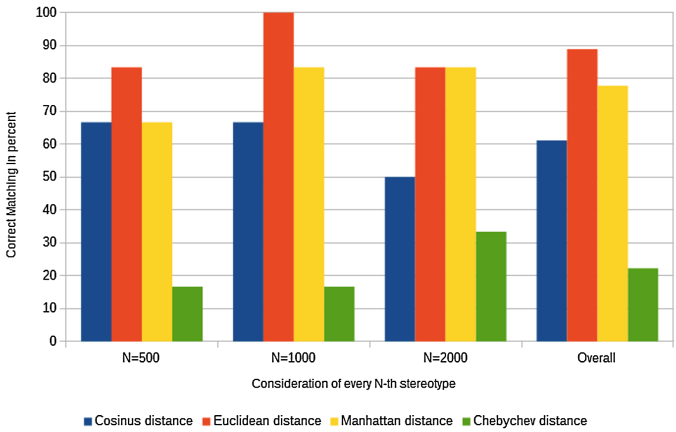


Fig. 3. Mapping results based on the evaluation of distance metrics with realistic test cases

accurate in higher dimensions [24]. The results of both, the realistic and theoretical evaluation steps, are shown in Figs. 2 and 3.

5.2 Delivery Coordination

In a second evaluation step, the correctness of the delivery coordination based on user preferences and information about available devices was analyzed. For this, multiple user profiles considering various device groups (stationary and ambient resp. embedded devices) and output options (video and audio) have been used to find the best matching delivery device from a fixed set of devices with different input and output capabilities. Also, the output devices are split up into two different device groups - both containing each possible output option - to discuss the influence of user preferences. Additionally, we provide instructions for six out of the given ten devices. Furthermore, the instructions are each of an output type such that the described device is not capable of presenting it. For example, a stationary display within a device ensemble may present image, text or video, but the interaction is described by an audio instruction. In that case, an alternative presentation device needs to be determined. The results (see Fig. 4) show, that the device selection process in general works as expected and the delivery coordination will find the device best matching a user's preference considering the generated instructions. Still, there exists some variability on device selection if the user preferences won't lead to a unique device selection, e.g., if the user prefers mobile devices with video output capabilities, all smartphones connected to the Ambient Reflection framework may fulfil such preferences. In such a case, the device that is known for the longest time by the framework may be chosen as delivery device. Hence, if no device matches the user preferred capabilities in any way, the longest connected device will be used for delivering the personalized instruction as a fallback device to at least deliver some kind of instruction.

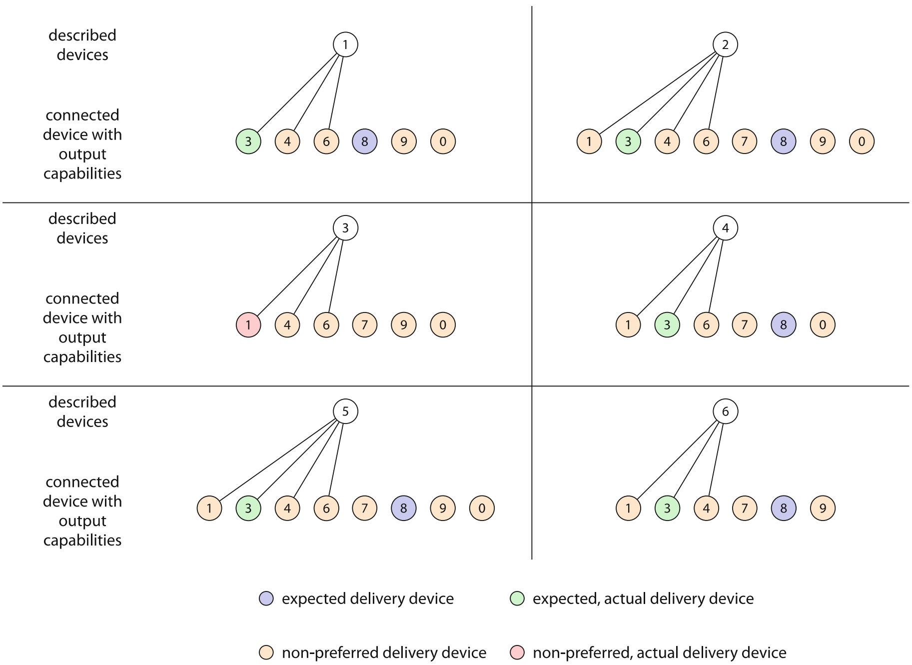


Fig. 4. Exemplary overview about the actual delivery devices. Each number indicates one specific device in the provided setup. An edge between two device nodes indicates, that both devices are in the identical device group.

6 Conclusion

This contribution provides an approach to enable the generation of personalized instructions by the Ambient Reflection framework based on the concept of self-reflective smart objects. Collected self-descriptions of smart objects are aggregated based on their current interconnection state and transformed to instructions to reduce the barriers in learning the interaction with such devices. By extending the existing framework using a developed user profile, a basic personalization process of instructions with stereotypes was realized targeting an advanced learning outcome. In order to ensure the best possible result, experts from disciplines like media psychology are able to define stereotypes and graphically adapt the belonging generation and delivery process by using a given set of rules. In order to assign user profiles to stereotypes and apply the belonging rule-sets, we've evaluated different distance metrics. Finally, the delivery process of instructions was evaluated based on a given scenario of a variety of devices.

References

1. ITU Telecommunication Standardization Sector: ITU-T recommendation Y.4000/Y.2060: overview of the internet of things. Technical report, International Telecommunication Union, June 2012. http://handle.itu.int/11.1002/1000/11559. Accessed 10 Mar 2017
2. Evans, D.: The internet of things: how the next evolution of the internet is changing everything. Technical report, Cisco Internet Business Solutions Group (IBSG), April 2011. https://www.cisco.com/c/dam/enus/about/ac79/docs/innov/IoTIBSG0411FINAL.pdf. Accessed 10 Mar 2017
3. Altakrouri, B.: ambient assisted living with dynamic interaction ensembles. Ph.D. thesis, University of Luebeck, Luebeck, Germany, August 2014
4. Ziegler, J., Lohmann, S., Kaltz, J.W.: Kontextmodellierung für adaptive web-basierte systeme. In: Stary, C. (ed.) Mensch und Computer 2005: Kunst und Wissenschaft – Grenzüberschreitung der interaktiven Art, pp. 181–189. Oldenbourg Verlag, München (2005)
5. Burmeister, D., Altakrouri, B., Schrader, A.: Ambient reflection: towards self-explaining devices. In: Proceedings of the 1st Workshop on Large-scale and Model-based Interactive Systems: Approaches and Challenges, LMIS 2015, Co-located with 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2015), Duisburg, Germany, pp. 16–20, June 2015
6. Norman, D.A.: Affordance, conventions, and design. Interactions **6**(3), 38–43 (1999)
7. Burmeister, D., Schrader, A.: Interusability durch ambiente Hilfestellung für Interaktion mit dem Internet der Dinge. In: VDE Kongress Internet der Dinge (2016)
8. OASIS Web Services Business Process Execution Language (WS- BPEL) TC: Web Services Business Process Execution Language Version 2.0, April 2007. https://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html. Accessed 10 Mar 2017
9. Burmeister, D., Burmann, F., Schrader, A.: The smart object description language: modeling interaction capabilities for self-reflection. In: The First International Workshop on Pervasive Smart Living Spaces 2017 (PerLS 2017), Big Island, USA, pp. 503–508, March 2017

10. Kallmann, M., Thalmann, D.: Modeling objects for interaction tasks. In: Arnaldi, B., H égron, G. (eds.) Computer Animation and Simulation 1998: Proceedings of the Eurographics Workshop in Lisbon, Portugal, 31 August–1 September 1998, pp. 73–86. Springer, Vienna (1999)
11. Raffler, H.: Other perspectives on ambient intelligence. Password Mag. **23**, 12–13 (2006)
12. Dey, A.K., Hamid, R., Beckmann, C., Li, I., Hsu, D.: A CAPpella: programming by demonstration of context-aware applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, pp. 33–40. ACM, New York (2004)
13. Preim, B., Dachselt, R.: Interaktive Systeme: Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung, chap. Die Interaktion mit Alltagsgeräten, pp. 135–161. Springer, Heidelberg (2010)
14. Norman, D.A.: THE WAY I SEE IT: simplicity is not the answer. Interactions **15**(5), 45–46 (2008)
15. Poppe, R., Rienks, R., van Dijk, B.: Evaluating the future of HCI: challenges for the evaluation of emerging applications. In: Huang, T.S., Nijholt, A., Pantic, M., Pentland, A. (eds.) Artifical Intelligence for Human Computing: ICMI 2006 and IJCAI 2007 International Workshops, Banff, Canada, November 3, 2006, Hyderabad, India, 6 January 2007, Revised Selected and Invited Papers, pp. 234–250. Springer, Berlin, Heidelberg (2007)
16. Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., Jaffe, M.W.: Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function. JAMA **185**(12), 914–919 (1963)
17. Barth, J.A.: The Automaticity of Everyday Live, Advances in Social Cognition, vol. X. Psychology Press, New York (2014)
18. Golemati, M., Katifori, A., Vassilakis, C., Lepouras, G., Halatsis, C.: Creating an ontology for the user profile: method and applications. In: Proceedings of the First International Conference on Research Challenges in Information Science, RCIS 2007, Ouarzazate, Morocco, 23–26 April 2007, pp. 407–412 (2007)
19. Casas, R., Blasco Marín, R., Robinet, A., Delgado, A.R., Yarza, A.R., McGinn, J., Picking, R., Grout, V.: User modelling in ambient intelligence for elderly and disabled people. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) Computers Helping People with Special Needs: Proceedings of the 11th International Conference, ICCHP 2008, Linz, Austria, 9–11 July 2008, pp. 114–122. Springer, Heidelberg (2008)
20. Rich, E.: Users are individuals: individualizing user models. Int. J. Man-Mach. Stud. **18**, 199–214 (1983)
21. Ur, B., Pak Yong Ho, M., Brawner, S., Lee, J., Mennicken, S., Picard, N., Schulze, D., Littman, M.L.: Trigger-action programming in the wild: an analysis of 200,000 IFTTT recipes. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 3227–3231. ACM, New York (2016)
22. Deza, M.M., Deza, E.: Encyclopedia of Distances, 3rd edn. Springer, Heidelberg (2014)
23. Binder, R.V.: Testing Object-Oriented Systems: Models, Patterns, and Tools, 1st edn. Addison-Wesley Professional, Boston (2000)
24. Aggarwal, C.C., Hinneburg, A., Keim, D.A.: On the surprising behavior of distance metrics in high dimensional space. In: Bussche, J., Vianu, V. (eds.) Database Theory—ICDT 2001: Proceedings of the 8th International Conference London, UK, 4–6 January 2001, pp. 420–434. Springer, Heidelberg (2001)

The Human Nervous System as a Model for Function Allocation in Human-Automation Interaction

Behice Durgun(✉)

Department of Anatomy, Faculty of Medicine,
Çukurova University, 01330 Adana, Turkey
bdurgun@cu.edu.tr

Abstract. The purpose of this presentation is to discuss structure and function of the human nervous system from the view point of the distribution of authority, sharing responsibility and controlling ways as a best model for designing of human-automation interaction. The cognitive view of function allocation was also discussed. The broad scope of ergonomics led to the most transparent, predictable and controllable automation system design. Function allocation is an ergonomic method to decide whether a particular function will be accomplished by a human, by an automation system or by human-automation interaction. Investigation of the human nervous system from integrative perspective provides the better understanding of the basis of automation and designing a better function allocation. High cognitive demands in human-automation interface can also influence the function allocation in man-machine system. It is expected that this presentation will offer the overview of this emerging area, as well as the new and future areas of its application.

Keywords: Human-automation interaction · Function allocation · Task analysis · Cognitive science · Human nervous system

1 Introduction

Automation and robotic systems have been smoothly integrated into countless industries, from plane to mobile phone. Automation depends on machines which not only process the information but also executes the dynamic functions many of which were previously performed manually. In this sense, sometimes dynamic functions cannot be controlled, and are affected by many factors. Therefore, the automation should be so designed that it allows the human and the machine to adapt to unexpected emergency situations [1]. Sheridan pointed out that the integration of technology and how work is designed, organized, and managed has not always been smooth, as there often has been time-lag between the application of workplace automation and workplace design and workforce needs [2].

In contrast to the extensive technical literature on automation, there is a small but growing research examining the function allocation involved in automation systems. Function allocation is the division of responsibility in human-automation interaction, and distinguishes what automation systems do and what humans do. By doing function

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_38

allocation, the designer takes into consideration the role of human, integration of human and automation system, fatigue, hazards, costs and human values. The Fitts list (1951) marked the beginning of function allocation research [3]. From the date of Fitts to now the approach to function allocation had been changed; human-centered automation studies lead the way for a better cooperation between human and automation [1]. With broad and developing the scope of ergonomics, human-centered automation has been the most studied issue for transparent, predictable and controllable system design [1, 4–8]. Parasuraman argues that better designs of systems and tasks need an understanding of the brain perceptual and cognitive processes [9]. However, investigation of the human nervous system from integrative perspective provides the better understanding of the basis of automation, designing a better function allocation and for the safer and more efficient working conditions [10].

Automation is a broad subject but this review outlines recent research and challenges about the area with an emphasis on allocation of function between human and automated systems, and on bringing human factors/ergonomics knowledge to the design process of human-automation interaction.

2 Automation

Automation is a technology of a system, a process, or a device for monitoring and controlling the production and function. This technology includes the mechanization and integration of the sensory process of environmental variables, data processing, decision-making and mechanical action [11]. Automation helps to reduce human error and to increase safety. Automation does not mean human replaces but rather changes it, sometimes with the ways that designers may not intend or not foreseen [11]. As a result, the human operator is forced to new cooperation demand [1]. This implies that in designing automation, from the lowest level of fully manual performance to the highest level of full automation should be taken into consideration. Parasuraman et al. proposed several levels of automation [7]. These levels are mainly based on the output functions performing decision and action selection; from the lowest level which all decision and action perform by a human to the highest level which the machine acts autonomously. Automation, however, may also perform the input functions, such as information acquisition and information analysis, that precede the decision-making and action. Wickens memory model provides a useful framework for analyzing the different neural processes used in interacting with systems and for carrying out a task analysis [12] (Fig. 1).

Investigation of the human nervous system from integrative perspective provides the better understanding of the basis of automation, designing a better function allocation and for the safer and more efficient working conditions. An analogy between automation and the human nervous system can be made. The nervous system also includes four steps to perform a motor action; sensory processing, perception, decision-making and response [10] (Fig. 2).

Automation systems use the several different mechanisms depend on how they perform the exact function. These are the artificial neural network; the distributed control system; the human machine interface; the supervisory control and data

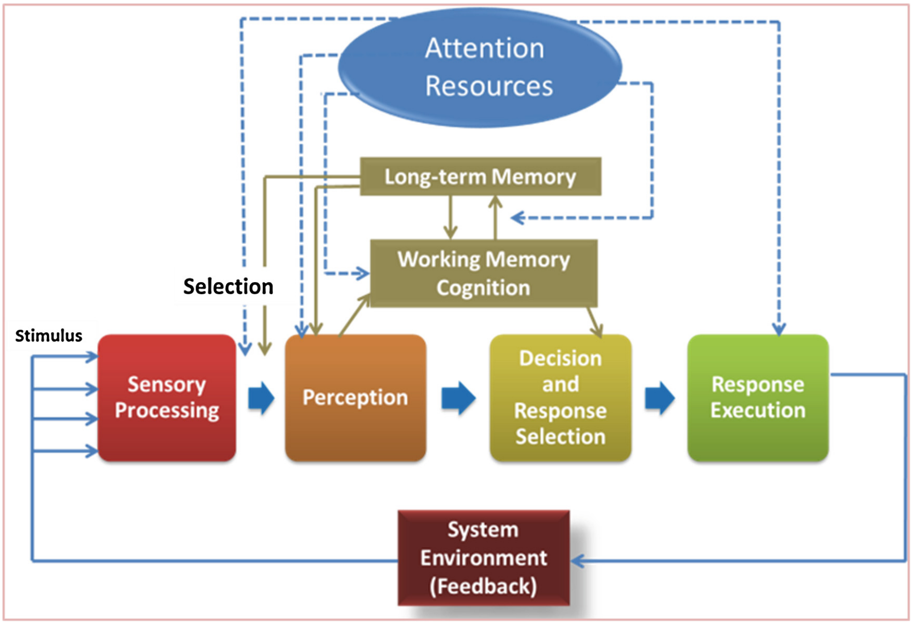


Fig. 1. The stages of information processing proposed by Wickens [12] (modified from Wickens, 1992).

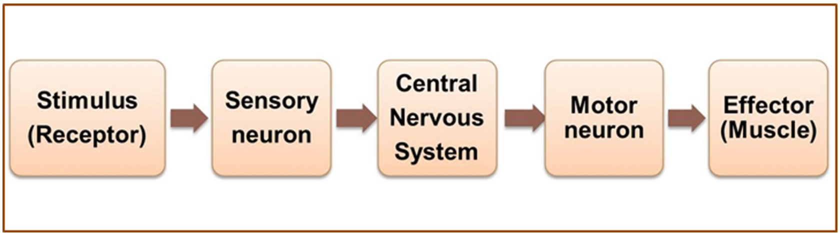


Fig. 2. The main steps of sensory-motor integration in the human nervous system.

acquisition; and the programmable logic controller. The artificial neural network is similar to the biological neuron connections that they can change the information on their course. In distributed control system, control spreading depends on the region of the system which needs operating. In a communication network, each control system connects to the others, which resembles the neural pathways of human nervous system. A human-machine interface system depends on human interaction with the system to act. A supervisory control and data acquisition system is a larger industrial control network that often has smaller sub-systems. These sub-systems include human-machine interface systems connected to remote terminal units which perform the translation of sensory signals into comprehensible data. Programmable logic controllers are real-time systems. This means that there are set a deadline and time-frame in which the desired result must be achieved.

The nervous system has been compared with an automation system, in that the brain and the spinal cord act as switching centers and the nerves act as cables for carrying messages to and from these centers. The nervous system serves as the chief coordinating agency for all systems. Conditions both within and outside the body are constantly changing. The nervous system must detect and respond to the changes (known as *stimuli*) so that the body can adapt itself to new conditions. Although all parts of the nervous system work in coordination, portions may be grouped together on the basis of either structure or function. There is both structural and functional hierarchy in the human nervous system. Therefore, the nervous system executes response from the basic reflex activity to the complex fine movements depending on at which level this four-step process is completed (Fig. 3). This is a magnificent example of hierarchical task analysis. In the most parts of the human nervous system, the steps also organized in a cyclic way (closed loop) or in a serial order (open loop); from stimulus to response [10].

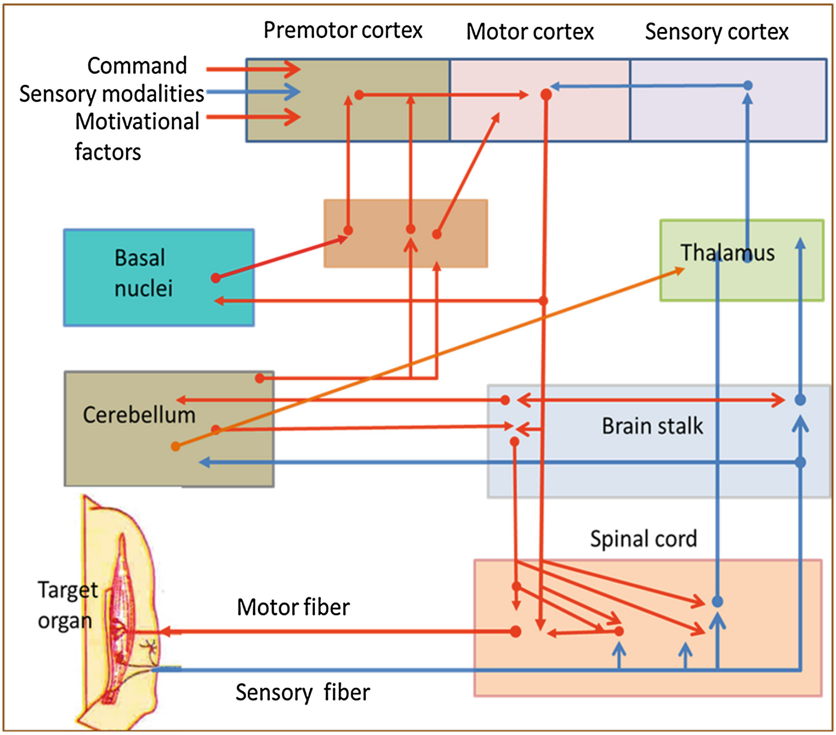


Fig. 3. The structural and functional hierarchy in the human nervous system. The nervous system executes response from the basic reflex activity to the complex fine movements depending on at which level the process is completed.

2.1 Adaptive Automation

Adaptive automation is a system in which both the user and the system starts changes in the level of automation. In adaptive automation, the level of automation or the number of systems operating on automation can be modified in real-time. So, the level or modes of automation engage more closely to operator's needs at any time [11, 13]. An example is an aircraft takeoff. An aircraft takeoff following an engine dysfunction may be automated at either a low or a high level depending upon the time [14]. Recent developments in neurological and physiological sensors to show users' mental state will increase interest in adaptive systems research and practice over the next few years. Adaptive systems are those that can appropriately change their behavior to fit the current context. The majority of adaptive systems are experimental and not practical because of the difficulty for accurate integration of user's mental state, environment, and task information [15]. The convergence of neuroscience and ergonomics (neuroergonomics, neuroengineering or cognitive engineering) is useful to understand the brain function underlying real-world human performance and to design the adaptive automation and, work environments for safer and more efficient operation [16, 17]. As a result, adaptive automation creates new challenges to both users and designers that go beyond traditional ideas of human-computer interaction and system design. The mechanism of the adaptive system similar to cerebellum serves as programmable logic controllers and plays an active role in motor coordination and the timing of movement. The cerebellum also detects errors in movement by comparing what is actually occurring to the intended action and it adjusts the correct movement [10].

2.2 Brain-Computer Interaction (BCI)

Brain-Computer Interaction systems use brain signals as an extra communication channel for human interaction with the environment. Different types of brain signals are used to control external devices without the need for motor output. This is helpful to people who either has only limited motor control or, patients with amyotrophic lateral sclerosis (ALS). In BCI, there is no motor control. Instead, the user is trained to engage in a specific type of mental activity associated with a unique brain electrical signals. The resulting brain potentials are processed and classified so as to provide a control signal in real time for an external device. There are four BCI components; signal acquisition, signal processing, the output device, and operating protocol [18]. BCI needs invasive techniques and therefore more attentive approach.

3 Human Physical and Mental Roles in Automation Systems

The human role in automation, especially in the high automation system is a challenge to designing better human-machine interface. Taking into consideration three levels, this challenge may be adequately addressed and analyzed [19]:

Micro Level. The main problem is how future interfaces are designed for preventing automation irony. The role of the operator, shared authority (human-machine cooperation) and adaptive automation seems to be more studied issues in the future [14]. New methodological approaches require for the assessment of basic physiological and mental parameters such as workload and situation awareness [20].

Meso Level. In the high-level automation, it is necessary to match the control over system and responsibility for system functions. Since the control is limited to the operational level, accountability may have to shift away from operator to operating organizations and system designers [21].

Macro Level. Recently, the multilayered networked system is used in aircraft. The important problem of the multilayered networked system is the governance or authority. Ilic reported that recent applications of telematics or concept of “smart grids” points to new modes of governance, which are emerging in practice, but still need to be understood in theory [22].

The human operator’s role in automation systems is to operate the automation and to intervene for taking manual control when necessary. One important determinant of operator’s choices of manual or automatic control is the degree of trust in the automation [23]. For the decision of the type and level of automation in any system design, a significant important point is to consider the consequences of human operator performance in the resulting system [22]. At high automation level, the arising question is the distribution of authority, responsibility and control between operator and automation system If two independently thinking entities start planning and acting together, the emergence of disagreement is inevitable. Baltzer et al., proposed the method of arbitration to solve this problem. Arbitration worked with different modalities as well as on different planning levels. An “interaction mediator” was designed to handle the complex interdependency in highly automated vehicles to incorporate a framework of modules [24]. In the human nervous system, cerebellum plays an interaction mediator role in motor coordination.

Automation systems often increase the workload and create unsafe working conditions [25]. Operational experience, field research, simulation studies, incidents and occasionally accidents have also shown that new and surprising problems arose. Disintegration that involves the interaction of operator and automation system are a remarkable and terrible way to failure in these complex work environments [4]. The importance of including human factors of the automation design process is to prevent “irony of automation” [26]. Human- in-the-loop simulation studies have shown that high automation systems increase human-automation performance, complacency and skill degradation, but decrease the situational awareness. This review also outlines situational awareness, complacency, and skill degradation by taking into consideration an operator’s physical and mental capabilities and limitations.

3.1 Physical Work

Traditional ergonomic evaluations focus solely on peripheral outcomes, such as force or muscle activity, and disregard the contributions to the brain during working, while cognitive ergonomic is an emerging field of study that focuses on the knowledge of human brain activities in relation to the control and design of physical tasks. Physical work can involve much static and dynamic work at different intensities, repetitions, and the duration, in turn, affects the autonomic responses. High automation systems leave fewer activities for people to perform. As a result, the operator becomes a more passive user instead of an active participant. This can actually inhibit operator's ability to detect critical signals or warning conditions. Further, an operator's manual skills worsen because of working automation for long periods.

3.2 Mental Work

Recently, the human mental workloads assessment is one of the most widely studied topics in ergonomics [27]. If operator's mental workload is either too high or too low, human-system performance may suffer from work environments. Therefore, the workload must be assessed at the design of new systems or the evaluation of existing ones [28]. Behavioral measures, such as accuracy and speed of response on primary and secondary tasks, ("reaction time" term has the same meaning) have been widely used to assess mental workloads [29]. However, measures of brain function offer some unique advantages for mental workloads assessment [27]. The typical finding in vigilance studies is that the detection rate of critical targets declines by time for the task. Vigilance-decreasing was attributed to a reduction in physiological arousal but more recent studies have attributed it to resource depletion. The parallel decline in vigilance performance and in blood flow velocity is found for both visual and auditory tasks.

Automation of decision-making functions may decrease the operator's awareness of the system and of certain dynamic features of the work environment because the operator is not actively engaged in evaluating the information sources leading to a decision. This might occur to systems where operators act as passive decision-makers monitoring a process when their intervention occurs only to prevent errors or incidents. The use of automation at the information analysis level may improve the operator's situation awareness [25].

Automation complacency occurs to conditions of the multiple-task load when a manual task to compete with the automated task for the operator's attention. If automation is highly, but not perfectly reliable in executing decision choices, then the operator may not check the automation and its information sources, and hence fails to detect the occasional times when the automation fails. Complacency can occur to both information automation and decision automation, but its effects on performance are greater with the decision automation [30]. The human operator will not skilled in performing decision-making function when this function is consistently performed by automation. Degradation of cognitive skills may be particularly important following automation failure. A recent simulation study of human control of a telerobotic arm used for movement towards hazardous materials found that following automation dysfunctions, performance was superior with an intermediate level of decision automation compared to higher levels. The reduced situation awareness, complacency and skill degradation collectively show that high-level automation can lead to "out-of-the-loop" unfamiliarity of the operator. All three sources of vulnerability may pose a threat to safety in the event of system failure. Automation that does not lead to unbalanced mental workloads, reduced situation awareness, complacency, or skill loss may nonetheless be associated with other human performance problems that impact on system performance, including mode confusion and low operator trust in automation [30]. The relative advantages of a specific level of automation can be determined by considering the consequences of human performance. There are few studies that have investigated the interaction between physical and mental work. The role of the human brain during physical and mental work should be within the scope of new studies. High cognitive demands can influence physical work, and physical activity can in turn influence mental processing.

4 Conclusion

The human nervous system as a model level of human-automation interaction and human roles in every level of automation were summarized with the expectation to be found useful by researchers in human factors and ergonomics, neuroscience, cognitive psychology, medicine, industrial engineering, and computer science. The highly automated machines are robots which include all stages of information processing. The most advanced robot has information processing rate of 500 million instructions per second, whereas the human brain alone has over 100 billion nerve cells, and each nerve make up to 10 000 connections to other nerve cells. Finally, there is no an automation system extended the border of human neural capacity. This means that high-level tasks such as strategic planning and decision-making in emergency conditions are still required human ability.

References

1. Hoc, J.M.: From human-machine-interaction to human-machine-cooperation. Ergonomics **43**, 833–843 (2000)
2. Sheridan, T.B.: Humans and Automation: System Design and Research Issues. Wiley, New York (2002)
3. de Winter, J.C.F., Dodou, D.: Why the Fitts list has persisted throughout the history of function allocation. Cogn. Technol. Work **16**, 1–11 (2014)
4. Christoffersen, K., Woods, D.D.: How to make automated systems team players. In: Salas, E. (ed.) Advances in Human Performance and Cognitive Engineering Research, vol. 2. JAI Press/Elsevier, Amsterdam (2002)
5. Billings, C.E.: Towards a human-centered aircraft automation philosophy. Int. J. Aviat. Psychol. **1**, 261–270 (1989)
6. Hollnagel, E., Woods, D.D.: Joint Cognitive Systems - Foundations of Cognitive Systems Engineering. CRC Press/Taylor & Francis, London (2005)
7. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. **30**, 286–297 (2000)
8. Young, M.S., Stanton, N.A.: Driving automation: learning from aviation about design philosophies. IJVD **45**, 323–338 (2007)
9. Parasuraman, R.: Neuroergonomics: research and practice. Theor. Issues Ergon. Sci. **4**, 5–20 (2003)

10. Durgun, B.: Sinir Sisteminin Fonksiyonel Anatomisi ve Ergonomi, 17. Ulusal Ergonomi Kongresi Bildiri Özetleri Kitabı, s. 48, Eskişehir (2011)
11. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse. Abuse. Hum. Factors **39**, 230–253 (1997)
12. Wickens, C.D.: Engineering Psychology and Human Performance. HarperCollins Publishers, New York (1992)
13. Hancock, P.A., Chignell, M.H.: Adaptive control in human-machine systems. In: Hancock, P.A. (ed.) Human Factors Psychology, pp. 305–345. Elsevier Science Publishers, North Holland (1987)
14. Inagaki, T.: Situation-adaptive autonomy: trading control of authority in human-machine systems. In: Automation Technology and Human Performance: Current Research and Trends, pp. 154–158. Erlbaum, Mahwah (1999)
15. Feigh, K.M., Dorneich, M.C., Hayes, C.C.: Toward a characterization of adaptive systems, a framework for researchers and system designers. Hum. Factors **54**(6), 1008–1024 (2012)
16. Parasuraman, R.: Putting the brain to work: neuroergonomics past, present, and future. Hum. Factors **50**(3), 468–474 (2008)
17. Karwowski, W., Siemionow, W., Gielo-Perczak, K.: Physical neuroergonomics: the human brain in control of physical work activities. Theor. Issues Ergon. Sci. **4**, 175–199 (2003)
18. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. **113**, 767 (2002)
19. Grote, G., Weyer, J., Stanton, N.A.: Beyond human-centred automation - concepts for human-machine interaction in multi-layered networks. Ergonomics **57**(3), 289–294 (2014)
20. Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., McMaster, R.: Representing situation awareness in collaborative systems: a case study in the energy distribution domain. Ergonomics **51**, 367–384 (2008)
21. Boy, G., Grote, G.: The authority issue in organizational automation. In: Boy, G. (ed.) The Handbook of Human-Machine-Interaction, pp. 131–151. Ashgate, London (2011)
22. Ilic, M.D.: Dynamic monitoring and decision systems for enabling sustainable energy services. Proc. IEEE **99**, 58–79 (2011)
23. Muir, B.M.: Trust in automation: part i. theoretical issues in the study of trust and human intervention in automated systems. Ergonomics **37**(11), 1905–1922 (1994)
24. Baltzer, M., Altendorf, E., Meier, S., Flemisch, F.: Mediating the interaction between human and automation during the arbitration processes in cooperative guidance and control of highly automated vehicles: base concept and first study. In: Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014, Kraków, Poland (2014)
25. Scerbo, M.W.: Adaptive automation. In: Karwowski, W. (ed.) International Encyclopedia of Ergonomics and Human Factors, pp. 1077–1079. Taylor and Francis, London (2001)
26. Goodrich, M.A., Boer, E.R.: Designing human-centered automation: tradeoffs in collision avoidance system design. IEEE Trans. Intell. Transp. Syst. **1**(1), 40–54 (2000)
27. Mehta, R.K., Parasuraman, R.: Neuroergonomics: a review of applications to physical and cognitive work. Front. Hum. Neurosci. **7**, 889 (2013)
28. Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A.: State of science: mental workload in ergonomics. Ergonomics **58**(1), 1–17 (2015)
29. Kaber, D.B., Endsley, M.R.: The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theor. Issues Ergon. Science **5**(2), 113–153 (2004)
30. Sauer, J., Chavaillaz, A., Wastell, D.: Experience of automation failures in training: effects on trust, automation bias, Complacency Performance. Ergonomics **14**, 1–14 (2015)

Citizen Science Involving Collections of Standardized Community Data

Werner Leyh[1(✉)], Maria Fava[2], Narumi Abe[2], Sandra Cavalcante[3], Leandro Giatti[3], Carolina Monteiro de Carvalho[3], Homero Fonseca Filho[4], and Clemens Jacobs[5]

[1] Department of Computer Science, University of São Paulo (USP), São Paulo, Brazil
WernerLeyh@yahoo.com

[2] São Carlos School of Engineering, University of São Paulo (USP), São Paulo, Brazil
mfava7@gmail.com, mail.narumi@gmail.com

[3] Department of Environmental Health, University of Sao Paulo (USP), São Paulo, Brazil
sandracavalcante@uol.com.br, lgiatti@usp.br, carvalhocm@gmail.com

[4] Environmental Management, School of Arts, Sciences and Humanities, University of São Paulo, USP, São Paulo, SP, Brazil
homeroff@usp.br

[5] GIScience, Institute of Geography, Heidelberg University, Heidelberg, Germany
c.jacobs@uni-heidelberg.de

Abstract. The interest of "Citizen Scientists" in their local environment is potentially of great value because they can assist in supplying essential "Environmental Knowledge" in an efficient and cost-effective way. This is particularly the case when "Volunteered Data" is registered in a standardized manner, interoperable with the data created by official institutions. The present work incorporates OpenStreetMap (OSM) and broadly accepted metadata-standards, that are controlled by scientific communities, to include the use of standardized interfaces for volunteered data contributions. An essential requirement for citizen science to operate, is the participation of the people. Spatial cognition is concerned with the acquisition, organization, employment, and examination of "knowledge about spatial environments". By this means "knowledge about spatial environments" is related to geographic proximity. Both OSM and metadata standards explore recent technologies for "Semantic Web" (SW) and "Linked Open Data" (LOD) enablement. The present study discusses the challenges and effects of standardized community contributions.

Keywords: Metadata · Dublin core standard (DwC) · Semantic Web · Semantic interoperability · Namespaces · WikiData · OpenStreetMap (OSM) · Spatial Data Infrastructure (SDI) · Citizen Science (CS) · Volunteered Geographic Information (VGI) · Ontology · Controlled vocabulary

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_39

1 Introduction

A recent study in [1] compared analyses of crowdcourced land cover data that were contributed by different groups, based on nationality and on domain experience:

The critical message arising from this research is, that it is important to consider and test for *potential variations in the way that landscape features are labelled and conceptualized* by different groups of contributors when analyzing crowdsourced data.

1.1 Related Work

1.1.1 Citizen Science (CS) – A Volunteered Geographic Information (VGI) Subset

The involvement of volunteers in environmental science and natural history has a long history, especially in Britain, long before it was termed "citizen science" [2]. The breadth of environmentally-based CS is immense. The guidelines for the UK Environmental Observation Framework (UK-EOF) define CS as "a volunteer collection of biodiversity and environmental data which contributes to expanding our knowledge of the natural environment, including biological monitoring and the collection or interpretation of environmental observations". CS and VGI both leverage volunteers to gather information and foster public interest. The ontology of VGI predicates that it is geographic in nature and dependent on volunteers; in contrast, CS refers to a practice, as opposed to the information itself. VGI encompasses a wide range of activities and practices, ranging from the 'fun' activities of locating summer holiday photographs to intensive surveying in the aftermath of an earthquake [3]. Within the VGI practices, there is a subset that falls into the category of CS - the involvement of non-professional scientists in data collection and, to some extent, its analysis [4]. Haklay [5] defines CS as the scientific activities in which non-professional scientists volunteer to participate in data collection, analysis and the dissemination of a scientific project and argues that not all CS is geographic; that is, although it involves a project in which a particular location on Earth plays a key role, within these VGI practices, there is a subset that falls into the category of CS. In [2], the authors recommend stretching the definition of CS beyond the contributory model so that it includes collaborative and co-created projects which require volunteers to carry out other tasks apart from the collection of environmental observations. Most studies that employ CS use volunteer-derived data as primary data, and adopt engagement strategies to the exclusion of Contractual and Co-created arrangements [6].

1.1.2 Citizen Science (CS) and the Growth in Electronic Communication

CS is likely to be the longest running of the VGI activities, with some projects requiring a continuous effort that may last for over a century [7]. Modern environmental management includes strong technocratic- and scientific-oriented management practices, while environmental decision- making is firmly based on scientific information about the environment. As a result, when an environmental conflict arises – such as a protest in the community over a noisy local factory – the evidence to back up the complaint, must be based on scientific data [7]. The growth in electronic communication has enabled communities to determine what methods to employ, find out about international standards and regulations and obtain information from scientific papers that can be combined with the local evidence.

1.1.3 The Wisdom of Crowds

Whereas official mapping agencies adopt procedures to control quality during the acquisition and compilation of geospatial data, VGI sources generally rely on crowd-sourcing - “the wisdom of crowds” is used as a technique to ensure quality and the convergence of truth [7]. One key advantage of the leading interactive web map VGI interface, OSM, is that its contributors are mainly concerned with their well-known local surroundings [8]. The data of (individual) mappers can be examined by local mappers (a crowd), which is the usual “quality assurance procedure” in OSM, as well as in many CS projects [4, 9]. CS Biodiversity projects very often rely on specialist reviews and the validation of observations as their principal quality control mechanism [4].

1.1.4 Localness Assumptions in Volunteered Geographic Information (VGI)

“After entering spatial data into GIS, the data usually require editing, to remove errors, or further processing. A common thread in studies of social media such as VGI, is the reliance on the so-called “Localness Assumption” [10, 11]. On the basis of this assumption, which is generally adopted implicitly, a unit of VGI in social media always represents the perspective or experience of a person who is local to the region of the corresponding geotag. In this case, it can be assumed that the closer volunteers are to the location of the data that they have uploaded, the more reliable the data will be [12] carried out a study of three separate social media communities (Twitter, Flickr, Swarm) to demonstrate that this localness assumption holds in about 75% of cases. Johnson et al. [10] found that in both Wikipedia and OSM, peer-review content about rural areas is: (i) of a significantly lower quality, (ii) less likely to have been produced by contributors concerned with the local area, and (iii) more likely to have been generated by automated software devices. In their review of VGI Quality Assessment Methods, Senaratne et al. [13] show that: (1) a majority of the reviewed papers focus on assessing map-based VGI. However, (2) implicit VGI (e.g. text-based Twitter or image-based Flickr) raises more concerns about quality than explicit VGI (e.g. map-based OSM); and, (3) this explicit VGI has been given significantly greater attention than implicit VGI, as a means of resolving problems of quality.

1.1.5 Controlled Vocabulary (CV) and a Lightweight Ontology (LA)

Controlled vocabularies (CV) such as terminologies and ontologies are regarded as a solution to the achievement of semantic interoperability across IT applications.

For example, the Dublin Core Schema is a small set of vocabulary terms that can be used to describe web resources (video, images, web pages, etc.), as well as physical resources such as books or CDs.

Darwin Core (DwC) is an extension of Dublin Core (DC) for biodiversity in-formatics. It is meant to provide a stable standard reference for sharing information on biological diversity [13].

The terms described in this standard are a part of a larger set of vocabularies and technical specifications under development and maintained by Biodiversity Information Standards (TDWG) (formerly known as the Taxonomic Databases Working Group (TDWG). The authors in [14, 15] argue that the unambiguous definition of DwC terms and their hierarchical structure, i.e., TAG "family = Fabaceae" is an implicit indication that TAG "order = Fabales", and reveals that the DwC is a light-weight ontology [14, 15]. This configuration lays stress on the automation of important tasks and interoperability, among the different knowledge organization systems (compare [16, 17] and Sect. 1.1.6).

1.1.6 Semantic Interoperability and Shared Controlled Vocabularies

Interoperability describes the extent to which systems and devices can ex-change data, and interpret that shared data. For two systems to be interoperable, they must be able to exchange data and subsequently display it in such a way that it can be understood by a user. "Semantic" interoperability provides interoperability at the highest level, and is the ability of two or more systems or entities to exchange information and use the information that has been exchanged.

Semantic interoperability is concerned not just with the packaging of data (syntax), but the simultaneous transmission of the meaning of the data (semantics). This is accomplished by adding data about the data (metadata), linking each data element to a shared, controlled vocabulary. The meaning of the data is transmitted with the data itself, in one self-describing "information package" that is independent of any information system. It is this shared vocabulary, and its associated links to an ontology, which provides the foundation and capability of machine interpretation, inferencing, and logic.

There is a close relationship between standards and semantic interoperability. A standard is a formal agreement of the meaning of a collection of concepts among communities that share a common interest. Conformance testing to a standard ensures, in principle, semantic interoperability within the scope of the standard. Standardized terminology is indeed one of the foundations of what is being developed in semantic networks. Controlled (standardized) vocabulary solves problems caused by the range of sources used to represent data values of specific metadata elements (compare e.g. [18]).

1.2 The Objective of the Present Study

Our study can be considered to be a direct response to the work of Hall et al. [16] who state that: "Our study revealed a fundamental tension between the need to produce structured data in a standardized way and OpenStreetMap's tradition of contributor freedom."

The aim of this study is to answer the research questions listed in Sect. 1.3. This will lead us directly to our objectives and challenges:

- (S1) On the one hand, in our daily communication, we use "standardized languages" with "controlled vocabularies": English, Chinese, French, etc.
- (S2) On the other hand, if we collect environmental data, it is often a challenging task to ensure that multiple, remote players agree on a common description of their data.

- (S3) In the light of (S1) and (S1), the study includes an analysis of the complex relationship between standardization and citizen participation.

1.3 An Attempt to Integrate "Crowd-Sourced" and "Official" Knowledge

The purpose of the present work is defined and determined by the following questions.

- (RQ0) Why is semantics important?
- (RQ1) What is a language? What are the language elements?
- (RQ2) What is vocabulary?
- (RQ3) What can be concluded from questions (RQ1) and (RQ2)?
- (RQ4) Why do we have difficulties in using well-defined languages for data collections?
- (RQ5) How does the description of standardized data affect the contributions made to Citizen Science?

1.4 Structure and Organization of the Paper

The remainder of the paper is structured as follows. Section 1 discusses relevant related work, while Sect. 2 describes in detail the overall approaches and methodology that are employed. In an attempt to clarify the benefits of the proposal, Sect. 3 analyzes the results and provides illustrative examples from a case study. Section 4 discusses these results together with their benefits, challenges and limitations. The paper provides a summary of its conclusions in Sect. 5.

2 Methodology

2.1 Applied Approaches

The following approaches are adopted:

- (MA1) A review of the literature summarized in the "Introduction" in Sect. 1;
- (MA2) A comparative study of daily communication between people on the one hand, and, the exchange of environmental data on the other (RQ1-3).
- (MA3) A response to the research questions (RQ0-5) outlined above within the constraints of the kind of contributions defined in Sect. 2.2.

2.2 Delimitation

The present work studies the impact made by the use of standardized community data. Factors such as preparedness, selection, preparation, motivation, benefits and other conditions that are broadly discussed in the literature, are not included in this research study (compare Sect. 1 and Comber et al. [1], Hall et al. [16] and Haklay et al. [20]).

3 Results

In this section, we represent the results of our study on the basis of the methods, approaches and protocols listed in Sect. 2. Our contributions ("the impact of standardized community data contributions in the context of "Semantic Web" (SW) and "Linked Open Data" (LOD) enablement") are in response to the research questions listed in Sect. 1.3.

First Stage. In the first stage an attempt is made to answer question (RQ1), (RQ2) and (RQ3) to show that a computer-mediated exchange of datasets is exposed to technical and non-technical filters similar to those found in direct human communication.

Second Stage. The second stage was based on the results about this data exchange and involved carrying out a literature review (Sect. 3 question RQ4), with regard to community contributed information and data. It was found, that on the one hand, Wikidata (the Wikipedia's structured data effort) and OSM support computing applications most effectively, provided that their structured data have a high degree of standardization.

On the other hand, in comparison with other attempts to design knowledge graphs on a large scale, "the Wikidata approach is on the chaotic side" [17].

Moreover, one major factor in the success of OSM-communities is their widespread adoption of a "be bold" principle that encourages contributors to "make changes as they see fit" and raises the freedom of contributors to the status of a core community.

Third Stage. The third stage is based on our findings from the first stage (where a "computer- mediated exchange of datasets is exposed to technical and non-technical filters") and our findings in the second stage ("Freedom versus Standardization"). It is argued in the third stage (RQ5) that "Controlled domain vocabularies may represent, similar to domain ontologies, domain knowledge".

3.1 (RQ0) Why is Semantics Important?

Generally speaking, a misunderstanding of the data can result in an invalid or unrepresentative analysis [18].

One new challenge arises from the huge increase in data production capacity by the varied experimental devices, which has an impact on the perennial issue of data sharing, the traditional practices for scientific research, and the modern, ontology-based methods for the definition of knowledge [19].

Semantic networking currently represents an advance in the long process of digital information analysis, from computer coding to semantic tagging and data retrieval. This process has now been consolidated by a wide range of standards that provide very high levels of technical, organizational and semantic interoperability [20].

3.2 (RQ1) What is a Language? What are the Language Elements?

According to [21], "language" can be defined as a system of conventional spoken, manual, or written symbols by means of which human beings, as members of a social group and participants in its culture, express themselves.

[22] defines language as a "system for communicating". Written languages use "symbols" (that is, characters) to build "words". The entire "set of words" forms the vocabulary of the language. The ways in which the words can be meaningfully combined are defined by the syntax and grammar. The actual "meaning of words and combinations of words" is defined by the "semantics".

3.3 (RQ2) What is Vocabulary?

According to the Oxford dictionary, vocabulary is "The body of words used in a particular language". According to [23] "vocabulary" can be defined as a set of "familiar" words within a person's language. When vocabulary, has grown with age, it serves as a useful and essential "tool for communication" and "acquiring knowledge".

Vocabulary may, however, be defined and controlled by a community for practical needs, without having to define the meaning or relationship between words. Once it has been arranged and structured in a hierarchy, the vocabulary is specialized and becomes a taxonomy with clear links between "words" [20].

3.4 (RQ3) What Can be Concluded from Questions (RQ1) and (RQ2)?

- Language is a system for communicating.
- We use familiar languages as tools for our daily communication.
- Written languages use "symbols" to build "words".
- The entire "set of words" forms "the vocabulary of our language".

Most people will automatically accept this conclusion, but this is not enough:

In our view people are not always willing or able to accept an opportunity to communicate.

It can also be concluded that as in the case of direct human communication, when establishing remote communication to exchange datasets through computers:

- The players will not always be available for these kind of data exchanges.
- These data-exchanges are exposed to many different technical and non-technical filters.

3.5 (RQ4) Why are There Difficulties in Using Well-Defined Languages for Data Collections?

The Challenge of Semantic Interoperability and Shared Controlled Shared Vocabularies. In summary, according to [17], language can be defined as a "system for communicating" whereas written languages use "symbols" to build "words" and the entire "set of words" forms the vocabulary of the language (RQ1).

"Semantic" interoperability can be accomplished by adding data about the data (metadata), and linking each data element to a controlled, shared vocabulary.

Exploring the Scientific Literature: The Case of WikiPedia/WikiData and OpenStreetMap. One major factor in the success of OSM-communities is their widespread adoption of a "be bold" principle. This principle encourages contributors to "make changes as they see fit" and raises the freedom of the contributor to the status of a core community value.

The structured data of Wikipedia are manifest in Wikidata, a major Wikimedia Foundation project, which has been described as a free linked database that can be read and edited by both humans and machines.

In comparison with other attempts to build knowledge graphs on a large scale, the Wikidata approach tends to be chaotic. There is no strict application of an upper ontology, no automated reasoning to support class inference or quality control, and no over-arching plan to govern the system's evolution [17].

OSM's structured data creation occurs in its tagging infrastructure, which lets editors specify the semantics of a geospatial entity using key-value pairs (e.g. its name, whether it is a restaurant or a hospital, etc.). For Wikidata and OSM to support computing applications most effectively, their structured data must have a high degree of standardization.

For Wikidata and OSM to support computing applications most effectively, their structured data must have a high degree of standardization. There is an inherent tension between the standardization needs of structured data and the ethos of ¨contributor freedom. Whereas Wikipedia tempers contributor freedom with a set of policies (such as "Neutral Point of View") that are strictly enforced by the community, OSM's "Good Practice" says "Nobody is forced to obey [the OSM guidelines], nor will OSM ever force any of its mappers to do anything.

Such challenges are addressed in [5, 24] and explored in detail in [16] (compare Sect. 1.1).

3.6 (RQ5) How Does the Description of the Standardized Data Affect the Contribution Made by CS?

It is certainly difficult to provide a detailed answer to this question that will apply to all the knowledge domains, but there is a significant general answer. In the light of Sect. 1.1.5, it can be argued (with regard to RQ5) that:

Controlled domain vocabularies" (CV) may represent a "domain of knowledge", similar to a "domain ontology". The reason for this is that even "flat vocabularies", in reality, may represent so-called "light-weight ontologies". They can be successfully used to represent knowledge in the domain of interest (compare Sect 1.1.5).

In view of this, it can be concluded that, for the application of CVs:

- CV can be applied in similar ways to domain ontologies;
- CV is regarded as a means of achieving semantic interoperability across IT applications.
- CV may thus also be regarded as a possible means of achieving semantic interoperability between Citizen Science (CS) and official institutions.
- CS-projects applying CVs may benefit from the effects of standardization in the same way as "normal" official institutions;

4 Discussion

One way to think about environmental data is as a network of connected entities, such as physical features, events, publications, people, species, sequences, images, and data collections that form the "environmental knowledge graph". Many questions in environmental informatics can be framed as paths in this graph [13, 25].

As already mentioned, CS-projects that employ CVs can benefit from the effects of standardization in the same way as "normal" official institutions; E.g. a serious challenge for many CS-projects is "establishing trust" in their products; The use of standardized vocabularies may assist in the development and sharing of validation tools (e.g. those based on the Darwin Core Standard the Global Biodiversity Facility which can ensure data validation, and are open for interested CS-communities).

Most of the world's knowledge remains locked up in unstructured texts. The use of CV facilitates also text mining and the automatic extraction of knowledge (e.g. OSM´s OverPass-API is a powerful example).

Major search engines are employing "schema.org" as a standard vocabulary for marking up web pages. It makes sense to use this for at least two reasons. The first is that it covers many entities that are often not included in domain-specific vocabularies. Secondly, there is a strong incentive to include a structured markup in web pages in order to improve the discoverability of search engines (compare [25]).

5 Conclusion

Controlled vocabularies can be applied in similar ways to domain ontologies.

Both may be "standardized", like the Dublin Core metadata standard and the Darwin Core metadata standard. This kind of standardization can lead to their (scientific, cultural, commercial) dissemination. Hence, this standardization can encourage the development of supporting products, applications and services.

In any case, applying controlled vocabularies means exploring an already developed, accumulated and maintained "domain knowledge", which, in many cases, is supported by highly sophisticated educational infrastructures (as in the case of our two examples, "Dublin Core" and "Darwin Core").

References

1. Comber, A., Mooney, P., Purves, R.S., Rocchini, D., Walz, A.: Crowdsourcing: it matters who the crowd are. the impacts of between group variations in recording land cover. PLoS ONE **11**, e0158329 (2016)
2. Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C., Robinson, L.D.: Understanding Citizen Science and Environmental Monitoring (2012)
3. Al-Bakri, M., Fairbairn, D.: User generated content and formal data sources for integrating geospatial data. In: Proceedings of the 25th International Cartographic Conference (2011)
4. Jacobs, C.: Data quality in crowdsourcing for biodiversity research: issues and examples. In: Capineri, C., Haklay, M., Huang, H., Antoniou, V., Kettunen, J., Ostermann, F., Purves, R. (eds.) European Handbook of Crowdsourced Geographic Information, pp. 75–86. Ubiquity Press, London (2016)
5. Haklay, M.: Crowdsourcing Geographic Knowledge. Springer Netherlands, Dordrecht (2013)
6. Wei, J.W., Lee, B.P.Y.-H., Wen, L.B.: Citizen science and the urban ecology of birds and butterflies—a systematic review. PLoS ONE **11**, e0156425 (2016)
7. Parr, D.A.: The Production of Volunteered Geographic Information: A Study of OpenStreetMap in the United States (2015). https://digital.library.txstate.edu/handle/10877/5776
8. Neis, P., Zipf, A.: Analyzing the contributor activity of a volunteered geographic information project — the case of OpenStreetMap. ISPRS Int. J. Geo-Inf. **1**, 146–165 (2012)
9. Eckle, M., de Albuquerque, J.P.: Quality assessment of remote mapping in OpenStreetMap for disaster management purposes. In: Proceedings of ISCRAM 2015 Conference - Kristiansand, 24–27 May, pp. 1–8 (2015)
10. Johnson, I.L., Lin, Y., Li, T.J.-J., Hall, A., Halfaker, A., Schöning, J., Hecht, B.: Not at home on the range. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI 2016, pp. 13–25. ACM Press, New York (2016)
11. Filho, H.F., Leite, B.P., Pompermayer, G.A., Werneck, M.G., Leyh, W.: Teaching VGI as a strategy to promote the production of urban digital cartographic databases. In: Joint Urban Remote Sensing Event 2013, pp. 222–225. IEEE, São Paulo (2013)
12. Fonte, C.C., Bastin, L., See, L., Foody, G., Estima, J.: Good practice guidelines for assessing VGI data quality. In: Proceedings of AGILE 2015, 9–12 June 2015, Lisbon, Portugal, pp. 2–5 (2015, in press)
13. Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M.M.: A review of volunteered geographic information quality assessment methods. Int. J. Geogr. Inf. Sci. **31**, 1–29 (2016). doi:10.1080/13658816.2016.1189556
14. Sachs, J., Finin, T.: Social and semantic computing in support of citizen science. In: CSWS 2011 Proceedings, pp. 46–51 (2011)
15. Sachs, J.: Using darwin core as a model: an ontologically minimalist approach to publishing occurrence data in RDF. In: Formal Models track of the Semantics for Biodiversity Symposium TDWG (2013). http://www.tdwg.org/fileadmin/2013conference/slides/Sachs_DwC_as_a_Model.pptx
16. Hall, A., Mcroberts, S., Thebault-Spieker, J., Lin, A.Y., Sen, S., Hecht, B., Terveen, L.: Freedom versus standardization: structured data generation in a peer production community. In: Proceedings of the 35th Annual ACM Conference on Human Factors in Computing Systems (CHI 2017) (2017). https://chi2017.acm.org/organizers.html

17. Good, B., Burgstaller-Muehlbacher, S., Mitraka, E., Putman, T., Su, A., Waagmeester, A.: Opportunities and challenges presented by Wikidata in the context of biocuration. In: Jaiswal, P., Hoehndorf, R., Arighi, C., Meier, A. (eds.) Proceedings of the Joint International Conference on Biological Ontology and BioCreative, Corvallis, Oregon, United States, 1–4 August 2016, vol. 1747, Corvallis (2016). CEUR-ws.org
18. Baclawski, K. (Ken Baclawski College of Computer and Information Science Northeastern University): Semantic interoperability for big data. In: http://www.ccs.neu.edu/home/kenb/class/data.semantics.html (ed.), Federal Meetup, 11 August 2014
19. Rafes, K., Rafes, K., Germain, C.: A platform for scientific data sharing to cite this version. In: Proceedings of the BDA 2015 Conference, 29 September–2 October 2015, Ile de Porquerolles, France (2015)
20. Henda, M.: Vocabulary standardization and semantic interoperability in education. http://www.benhenda.com/eng/node/64
21. Robins, R.H., Crystal, D.: Characteristics of language (2015). https://global.britannica.com/topic/language
22. Fonseca, Z.S.: The Nature of Human Language (2009). http://unad-act6humanlanguageblogspot.com.br/2013/10/the-nature-of-human-language.html
23. Wikipedia Community: Vocabulary (2017). https://en.wikipedia.org/wiki/Vocabulary
24. Mooney, P., Corcoran, P.: Has OpenStreetMap a role in digital earth applications? Int. J. Digit. Earth. **7**, 534–553 (2014)
25. Page, R.: Towards a biodiversity knowledge graph. Res. Idyeas Outcomes **2**, e8767 (2016)

The Signs of Semiotic Engineering in the IoT Interaction Design

Marcia Ikezaki Ferrari[1(✉)], Silvia Amelia Bim[2], and Plinio Thomaz Aquino Jr.[1,3]

[1] IPT – Instituto de Pesquisas Tecnológicas, Av. Prof. Almeida Prado, 532, Butantã, São Paulo, SP, Brazil
marcia@luzdigital.com.br, plinio.aquino@fei.edu.br

[2] UTFPR – Universidade Tecnológica Federal do Paraná, Av. Sete de Setembro, 3165, Rebouças, Curitiba, PR, Brazil
sabim@utfpr.edu.br

[3] Centro Universitário da FEI – Fundação Educacional Inaciana Pe. Saboia de Medeiros, Computer Science Departament, Av. Humberto de A. C. Branco, 3972, São Bernardo do Campo, SP, Brazil

Abstract. In an IoT (Internet of Things) environment new forms of interaction between human and computer can occur through devices that combine the object with sensors and actuators. In view of this scenario, one of the purposes of bringing together the areas of HCI (Human Computer Interaction) and IoT design is to understand how this new form of interaction with the participation of sensors and actuators happens, and thus contribute to the development of the design of interaction adequate to the needs of user. The objective of this paper is to establish the application of the concepts Semiotic Engineering (a theory of HCI), which is based on the study of signs and considers every form of interaction as an artifact to communicate a message between the system designer and the user, to contribute to effective and efficient communication, and consequently increase the quality of use of the system. The validation of this application occurred in a positive way, but with some restrictions that reflect the specific characteristics of this type of interaction, such as ubiquity and the distancing from the traditional forms of interaction (mouse, screen, and keyboard) between the person and the computer.

Keywords: Human computer interaction · Internet of Things · Semiotic engineering · Signs · Semiotic inspection method

1 Introduction

IoT, according to Gubbi [1], emerges as an important precursor of ubiquitous computing and its essence is the set of “things” interconnected by the internet, in such a manner as to provide varied information about the environment in which they are located at any time, besides allowing for the remote management of multiple devices near them. The “thing” in the context of IoT can be broad and denominate any object with processing capacity and connected to a network, with applications in several fields, for example, in environments that seek to support daily tasks and improve people’s quality of life [1–3].

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_40

Given this scenario, traditional forms of interaction between the human and the computer that happen through the monitor, mouse, and keyboard, can be replaced by other forms of interaction involving other senses of the human body, such as movement, gesture, and others [4].

Proposing theoretical models that help in the definition and design of such interactions involving these diverse forms of interaction between the human and the computer, as well as the implications in the evaluation of quality of use, are pointed out by Furtado [5] as one of the challenges in HCI (Human Computer Interaction) for the next decade.

Research by Koreshoff [4, 6] proposes an approximation between HCI and IoT as a way of minimizing problems (such as failures in understanding how the system works and how the interaction between the user and the IoT device can happen) that can arise in a system developed by the IoT designer, through the development of systems that respect the skills, knowledge, and dynamics of people's behavior. Other techniques such as profile mapping, context, interaction patterns, and interface patterns can also be approached as a means to reduce implementation-related failures in IoT designs [7, 8, 24].

One of the recent theories of HCI, Semiotic Engineering defines the interactive computational system as an artifact to communicate a message from the designer of the system to the user, called metacommunication. The study of signs is its fundamental basis, with communicability being the main property of the system to be evaluated for the quality of use of a computer system. Based on this theory, software should be developed considering the user's understanding of the application's functioning, to whom it is directed, its purpose, and the principles that define the possibilities of interaction with the system [9, 10].

Semiotic Engineering is not intended to determine solutions to the metacommunication failures detected in the interaction between the designer and the user through the interactive computational system, but must show what the failures were and how they happened [9]. There is also a need to investigate the application of Semiotic Engineering in new contexts, different forms of interaction, and different types of signs, seeking to establish improvements in the quality of the interactive computational system from the point of view of HCI [11].

The proposal of this paper is to establish a means to apply the concepts of signs in Semiotic Engineering to support the design of IoT interaction, based on efficient and effective communication in the design of the interaction which helps to reduce the failures and errors that may occur in the interaction between the human and the IoT environment, reducing the cost, time, and the reconstruction of the interactive system in an IoT environment, contributing positively to the quality of use of the interactive computational system.

2 IoT (Internet of Things)

Recommendation ITU-T Y.2060 [3] defines IoT as a global information infrastructure that enables interconnection of "things" (physical or virtual), exploring the capabilities of identification, capture of data, and processing and communication, making full use of "things" to offer services to all types of applications.

A concept widely used in literature review defines the IoT paradigm as a convergence of three visions that are: "Thing" Oriented Vision, "Internet" Oriented Vision, and Semantics Oriented Vision, representing the ubiquitous presence of objects, sensors, and actuators which interact and cooperate with each other to achieve a specific objective [2].

2.1 HCI in IoT

Koreshoff's research [4] carries out an approach on an IoT-centered model, starting with the mapping of works found in HCI that were classified based on each of the three visions and their intersections [2], as shown in Fig. 1. The result of the research compiled by each of the visions and intersections shows that there is a concentration in the "Thing Oriented Vision", explained by the proximity and constant interaction of the human with the object [4].

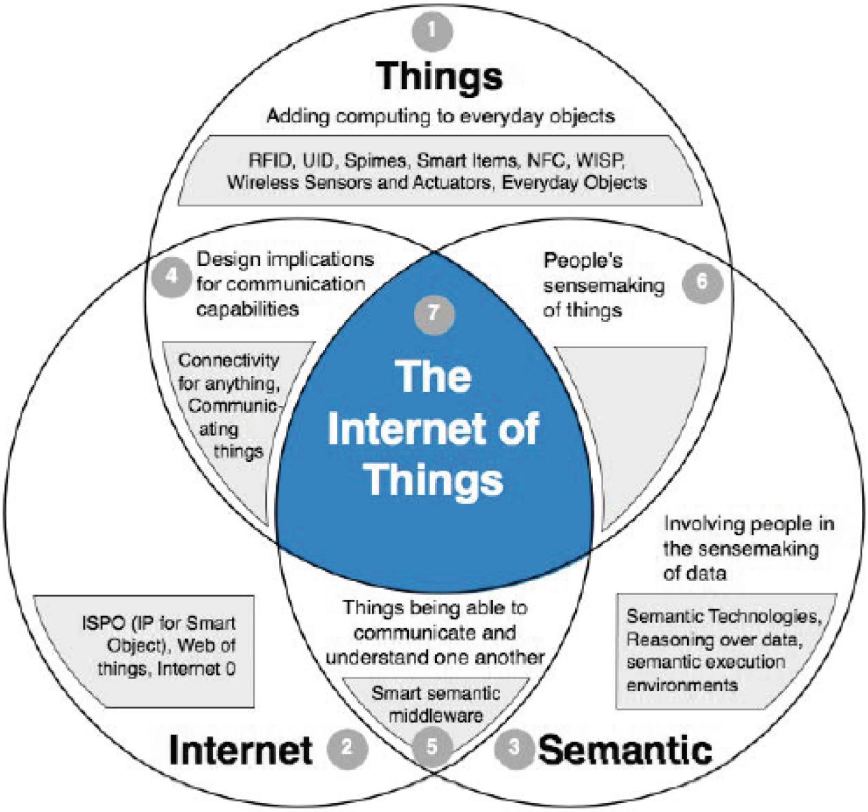


Fig. 1. Koreshoff et al. [4] vision Modified version of Atzori et al. [2] 'Internet of Things' paradigm

In a new study, Koreshoff et al. [6] seek to establish, with the help of classification tables, the direction of efforts in HCI for IoT in academia and what types and in what domains IoT products are being marketed. The authors believe that a correct understanding of research and commercial efforts can guide and help researchers to reflect and find a suitable way to incorporate the area of HCI into IoT.

3 Semiotic Engineering

Semiotic Engineering is a discipline that studies the phenomena of communication and signification, as well as the interlocutors involved in this process and in the context of design, which in HCI allows the understanding of the factors involved in the design, use, and evaluation of an interactive system [9].

In the vision of Semiotic Engineering, the HCI is characterized as a specific case of metacommunication, that is, the exchange of messages happens through a language established in the channel of interaction between the user and the designer to perform the tasks. It is a metacommunication because the message is indirect, because the designer is not present at the moment of interaction, and also unidirectional, because in the context of the interaction the user cannot continue communication with the designer [9].

The HCI artifacts are communicated through signs, in which the user must be able to interpret, learn, and adapt to various contexts of needs and opportunity. Signs and their process of signification are the basis of study of Semiotics. One of the main advantages of the semiotic view on HCI is to centralize the attention of researchers on the signs, which are computationally coded according to the intention and decision of the designer. Signs have two components, one of expression and the other of interpretation, used by the designer who is responsible for creating an interaction language to communicate with the user through his representative (the system conceived by the designer), called the designer's delegate [9, 11].

In the process of signification of the sign, there are three elements involved: the object (representation of the sign by the person), the representation (signification of a sign that generates another sign), and the interpretant (idea of the sign in the person's mind). The associations that form between the representation and the interpretant begin a process of interpretation of signs that happens through a set of prior knowledges determined by the systems of known signs, giving meaning to things. This process of interpretation can generate a new sign, in a theoretically indefinite and unlimited chain of ideas of associations, whose process is called unlimited semiosis, since the more a thought is meaningful, the more meaning is attributed to it [13].

From the point of view of Semiotic Engineering, the coding of signs represents the freezing of the designer's semiosis in each context. If the user does not recognize the sign or has difficulty in doing so, it means that the semiosis was unsuccessful, causing dissatisfaction and discomfort, and the failure in communication can be complete [9]. The role of the sign must overcome the language barrier and allow for a precise interpretation that reduces the time for understanding of the message and helps with quick decision making by the user [14].

The failure in quality of metacommunication is characterized by a communication breakdown, which is identified by the analysis of the signs present during the interaction between the human and the software and should be segmented into three classes: static, dynamic, and meta-linguistic [15, 16].

Communicability is the main attribute of quality of Semiotic Engineering that should enable the user, through his interaction with the system, to understand the premises, intentions, and decisions used by the designer during the development process. The greater the user's understanding of the designer's logic built in to the application, the greater is his chance of using the system effectively, efficiently, and productively [9, 10].

The SIM (Semiotic Inspection Method) is a method of inspection of Semiotic Engineering that does not require the participation of the user in the evaluation, being that the specialist incorporates the role of the user and evaluates the quality of emission of the metacommunication, trying to identify possible ruptures of communication. The metamessage is evaluated in a segmented way through each of the signs present in the interaction that are classified into static, dynamic, and meta-linguistic, generating a report with the problems encountered. These signs express a distinct meaning in the system and have a power of expression that serves different communicative purposes [15–17].

4 IoT Interaction Design and Semiotic Engineering

Analyzing interaction design from the viewpoint of communicability will allow the designer to identify if his message is sent to the user in a clear, objective, and unambiguous manner, indicating efficient and effective communication, allowing the user to perform his tasks and achieve his objectives satisfactorily.

As such, to try to establish a way to apply the concepts of Semiotic Engineering in the design of IoT environments, the main features in IoT described above are seen once again. IoT can be considered a system that combines sensors, actuators, and communication devices with the object, interconnected with each other and with the Internet, while being able to process information and exchange and send data [1–4, 6].

In the interaction between the human and the computer in IoT, the sensor is responsible for registering information captured from states external or internal to it, transforming the captured signals into information useful to the system. Considering an interaction in IoT, the devices responsible for user interaction are: the sensor that acts as the "eyes" of the system, being responsible for collecting information, and the actuator, that acts as the "hand" of the system, transforming the information into results [18]. Based on this information, the proposed structure for this interaction is presented in Fig. 2.

The sensors in IoT determine the form of interaction that can happen between the human and the object, which does not always happen via the traditional means (through mouse, keyboard, monitor, or touch), and it is necessary to consider other forms of interaction, sensation and perception, involving other senses of the human body, such as movement and gesture, among others. In the same way, the response of an IoT system can come in the form of light, color, temperature, among others, signifying a change in the current state of the object (thing) or the environment [6].

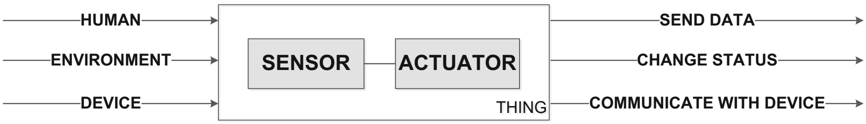


Fig. 2. Interaction structure by means of sensor and actuator

As shown in Fig. 2, in IoT the capture of information by the sensors does not depend only on the interaction with the person. Data collection can take place through other devices, objects or the environment in which the sensor is present, characterizing the ubiquity of the system. However, this is not the focus of this paper, since the point of interest is related to the interaction between the person and the sensor in IoT, establishing a form of interaction between the human and the "computer". Figure 3 highlights the point of interaction and the response that are the objects of study of this paper.

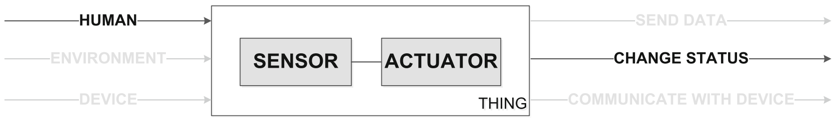


Fig. 3. Highlighted interaction between person and the sensor/actuator

Understanding the current context of users, devices, and the environment, considering the communication that happens between the computer and the user, between the user and the devices, and between the devices themselves, is one of the essential factors for the IoT system to meet its objectives [19].

Once established how and in what form the relationship between the person and the IoT system occurs, we look once again at the concept of interaction from the perspective of Semiotic Engineering, which considers interaction as a process of communication between people through a computational system, in which actions take place such as manipulation, communication, conversation, and exchange, among others [9, 12].

The interactive intellectual artifact is the result of intellectual exercise and human ingenuity, which is generated from the designer's interpretation of a problem, and the design of a solution must be presented in a linguistic coding in which the user must be able to understand the linguistic system and use the solutions proposed in the artifact. When an intellectual artifact is an interactive system, its design should consider the following aspects: artificial language should allow processing by a computer; the interface language that the user is going to interact with is always unique and new to the user and the artifact is considered a metacommunication [9].

The message must speak for the designer in the time of interaction, gradually transmitting to the user all the meanings coded by him, since the designer is not physically present at the time of interaction and is represented by the system [9, 11]. This message is transmitted through a set of computationally coded signs, which should inform the user how he can or should use the system to perform a task and achieve his goal [11].

In Semiotic Engineering signs are classified into meta-linguistic signs, static and dynamic. The meaning of static signs is interpreted independently of temporal or causal relationships, that is, their interpretation is limited by the elements visible at a given moment, in a single instant in time. Dynamic signs are linked to temporal and causal aspects, that is, to the interaction itself. Finally, meta-linguistic signs must inform, explain, clarify or illustrate explicitly the static and dynamic signs, as well as the functioning of the system. Thus, the success of the system is related to the efficiency of metacommunication, that is, it is up to the designer to engineer the signs, deciding which signs will compose the message to consistently and effectively inform his intentions to the user through a good communicability [11].

Communicability should indicate whether the set of computationally coded conversations between the delegate of the designer (system) and the user at the time of interaction happens effectively and efficiently [9, 11].

From the perspective of Semiotic Engineering, the IoT system can be defined as a computational artifact that can interact with a person, receiving information, processing it and assisting the person in their effort to achieve an objective in each context of use. As such, the IoT system be a linguistic intellectual artifact, whose system is considered as an artifact of metacommunication, since the system is the result of human effort with a language that allows its interpretation. Observing the IoT system as a metacommunication artifact, the designer is in the same communicative process as the user, as shown in Fig. 4.

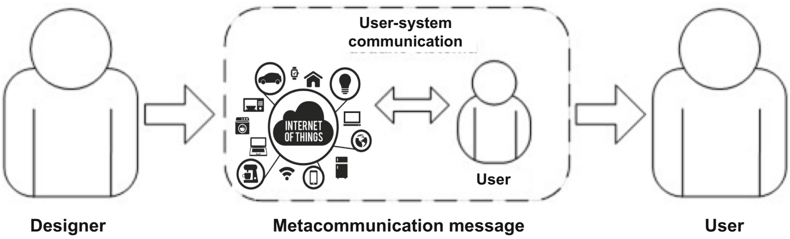


Fig. 4. Metacommunication between designer and user in an IoT system Adapted from Barbosa e Silva [12]

Semiotic Engineering is inherent to the construction of any system, as a common process on the part of system designers, precisely because it understands the interaction between the human and the computer as a communication between people, composed of designers and users [20].

The message that is part of the communication between the designer and the user is composed of a set of computationally coded signs [11]. Thus, human communication is closely linked to the generation and interpretation of signs, which is one of the human resources that enable communication [20].

In an IoT system, its interaction with the user happens through the sensors that can be considered as a sign, since in the real world it is part of the environment like any object, and on the interpreter's side, the translation of the quantity measured by it is the starting point for the process of semiosis that occurs within the interpreter [21].

The methods of Semiotic Engineering are interpretive, therefore the interpretation of what the static, dynamic, and metalinguistic signs are depends on the evaluator's view of each analyzed case, and how he can frame the concepts and materials with which he will work, while being part of the process of applying the method [16, 25].

5 Validation

Validation was accomplished through a workshop in which fourteen students enrolled in technology and engineering courses of a university participated, and who all had knowledge and skills in the development of software, applications and the like. Before the beginning of the activity, all participants received training on what is a sign and its importance in communication between people, with its influence on the understanding of the message.

The workshop activity consisted of developing designs in an IoT environment, based on a fictional scenario, with the layout of a house and a list of devices (sensors, microphones, loudspeakers, actuators, among others) that could be used in the design of a solution.

The participants were asked to prepare an IoT design for the environment of a home for an elderly person who lives alone, considering aspects such as well-being, comfort, and safety, among others.

The dynamics for the design of the interaction between the human and the IoT environment had no restriction on the use of devices (sensors, microphones, actuators, among others), since the main objectives in this workshop were: to obtain a set of data with the ideas and interaction design visions for an IoT environment, without considering technical or cost viability of this design. The idea was to present designers with a "blank canvas" on which they could create and design their interaction design without a predetermined template, without interferences or suggestions that could influence them during development and could affect the outcome of the design.

The data were collected through discussions, observations, and questionnaires, and the main results obtained from the interaction designs demonstrated the designer's concerns with the user's monitoring and safety, as shown in the graph of Fig. 5.

Based on the design features, the devices most used for monitoring were: the camera that allows visualization and tracking, and the presence/motion sensor. In the safety feature, the most used devices were: the temperature sensor that allows the measurement of ambient temperature and temperature of the human body. The display, microphone, and loudspeaker also indicate traditional forms of interactions that involve a screen where the user can search for or include information, microphones for voice commands, and loudspeakers for audible warnings, as shown in the graph of Fig. 6.

Based on the developed design, the designers were asked for solutions in which the IoT environment should send responses to the user. Overall, most participants opted for the use of smartphones and other devices that have video monitors, such as: tablets, monitors, TVs.

Based on the list of sensors and devices used by designers, and on the definitions of Souza and Leitão [11] for the three classes of signs present in the human-computer interaction, each design was classified according to the following characteristics: device

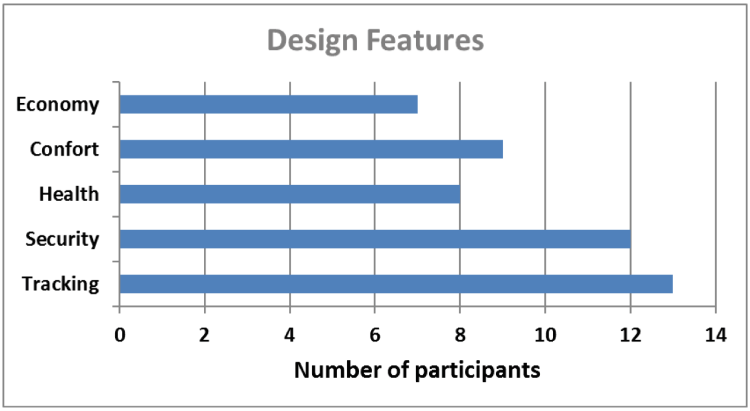


Fig. 5. Main aspects considered in the interaction design

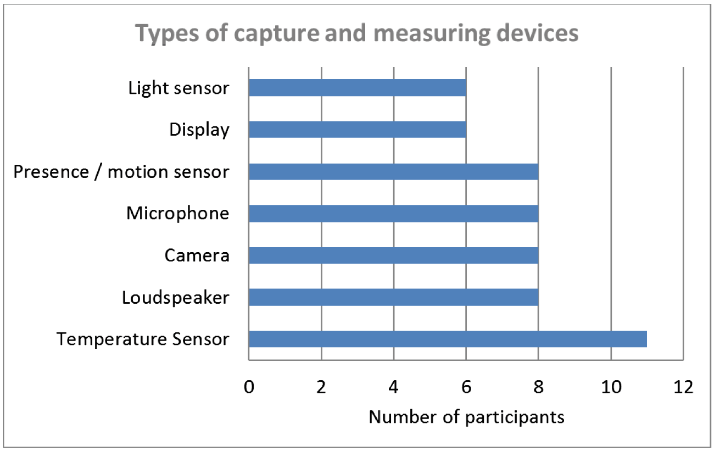


Fig. 6. Types of capture and measuring devices most used in the designs

used, ubiquitous (yes/no), input (yes/no), response/output (yes/no), and sign (metalinguistic, static, dynamic).

The compilation of the results showed some sensors used in different ways, that is, there were designs that considered an interaction between the person and the sensor, and others that considered the same sensor in ubiquitous situations.

Looking back at the definitions of each sign [11], static signs express the state of the system and its state is interpreted independently of the causal or temporal relationship of the interaction, that is, they are elements visible at a given moment interpreted in the spatial dimension. According to this definition, all sensors and devices used to capture image or sound, such as cameras and microphones, are responsible for the input of the system and have been classified as static signs, as they do not alter their internal and external state, and its interpretation is timeless, that is, there is no change in each moment of interpretation.

Dynamic signs express the behavior of the system and are subject to temporal and causal aspects, linked to the interaction itself. There is not necessarily a representation of a visible element, since the sign can only be identified with the interaction [11]. As such, all devices or objects that received quantities measured (collected) by the sensors or capture devices, and sent a response, altering its internal or external state, affecting the user, were considered as dynamic signs. The interpretation of the output is temporal and caused by an interaction of an input. The lack of the meta-linguistic sign in a design for an IoT environment may indicate the transparency of technology and ubiquity.

The compilation of the results obtained in the designs showed that there are sensors that can be used in ubiquitous or non-ubiquitous situations, such as the temperature sensor. This sensor can be used to capture the temperature of the human body and as such the input of the information depends on the interaction between the person and the sensor, which can happen in a wearable way or by contact. The temperature sensor can also be used to collect data from the environment or of a plant, regardless of human intervention, thus characterizing ubiquity. There are also sensors that are only used ubiquitously, as is the case, for example, of the sensors used to collect seismic information.

6 Discussion and Conclusion

This paper brought together the areas of HCI and IoT by establishing the IoT system as a metalinguistic artifact based on the concepts presented in Semiotic Engineering, identifying and classifying the sensors used in non-traditional interaction design in an IoT environment as static or dynamic signs. This approach determines the design of a user-friendly interaction design, whose correct understanding of the signs indicates effective and efficient communication between the user and the designer, contributing positively to the quality of use of the system.

Among the designs that were presented, 35% of the designers chose devices that allow an interface via monitor, tablet, or TV, among others. The option for this type of device may reflect a characteristic of the designer who is more accustomed to the development of software for the web, client servers, and mobile applications, as they still work out solutions for graphical interfaces using text and images for interaction with the user. Although the solutions presented in the exercise are functional, the IoT paradigm breaks the barrier, distancing itself from the traditional forms of interaction, allowing the designer to explore solutions that do not need a graphic interface.

Semiotic inspection was performed at design time considering the non-traditional forms of interaction, with it being possible to characterize and inspect the different signs in an IoT environment, reinforcing the concept that it is possible to inspect the signs independently of technology and domain. The absence of the metalinguistic sign in the interaction between the user and the IoT environment indicates an easy-to-use system without the need for explanations or help, transparency of technology, and ubiquity.

According to Koreshoff [6], scientific research has investigated one or two (at the most) sensors at the same time, while IoT products sold on the market combine several sensors and devices in a single product or environment. Likewise, in this study, all devices (sensor and actuator) used in the projects were analyzed and identified individually in static and dynamic signs, according to the definitions of Souza and Leitão [11]. However, in an IoT environment, the designer can use the devices in a grouped manner, generating a new type of device with numerous sensors, which could invalidate the individual classification of the signs, because in this group of devices it is possible to find sensors that do not depend on human interaction in data capture, characterizing ubiquity.

Ubiquity is one of the main characteristics observed in the IoT project. Under this aspect, the system cannot be characterized as a metalinguistic artifact according to the concepts of Semiotic Engineering because there is no interaction between the user and the system, that is, data collection is independent of user action and happens by means of other devices, objects or from the environment in which the sensor is present. However, even in a situation of ubiquity, the user may be affected by the response or reaction of the IoT environment.

The questions that arose during the application of the workshop stem from: the new technologies involved in IoT systems that involve many existing devices and can be used individually or grouped together and coupled with other objects, non-traditional forms of interaction between the human and the IoT environment, ubiquity, and the scarcity of material that brings together the areas of IHC and IoT.

As future work the signs are used to compose the interface interaction project in an IoT environment. The user in the IoT environment will have collaborative activities in a virtual environment [23]. EEG data is captured from user behavior using the virtual environment [22]. Real-world user behavior data (captured by IoT) and user behavior data in the virtual world (virtual reality and EEG) will be compared. Signs are elements studied in these environments as interface elements.

Acknowledgments. This work is supported by FEI (*Fundação Educacional Inaciana Pe. Sabóia de Medeiros*) and FAPESP (*Fundação de Amparo à Pesquisa do Estado de São Paulo*) for financial support of participation in the AHFE 2017.

References

1. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Future Gerer. Comput. Syst. **29**, 1645–1660 (2012)
2. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: a survey. Comput. Netw. **54**, 2787–2805 (2010)
3. ITU-T Y.2060: Telecommunication Standardization Sector of ITU. ITU-T Y.2060: Overview of the Internet of Things (2012)
4. Koreshoff, T.L, Leong, T.W., Robertson, T.: Approaching a human-centred internet of things. In: Proceedings of 25th Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration, Adelaide, pp. 363–366. ACM, New York (2013)

5. Futado, E.S., Militão, G., Andrade, R., Miranda, L.C., Oliveira, K.: Ubiquidade, Múltiplos Dispositivos e Tangibilidade. In: Baranauskas, M.C.C., de Souza, C.S., Pereira, R. (org.). I GranDIHC-BR - Grandes Desafios de Pesquisa em Interação Humano-Computador no Brasil. Relatório Técnico. Comissão Especial de Interação Humano-Computador (CEIHC) da Sociedade Brasileira de Computação (SBC), pp. 23–26 (2014)
6. Koreshoff, T.L., Robertson, T., Leong, T.W.: Internet of Things: a review of literature and products. In: Proceedings of 25th Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration, Adelaide, pp. 335–344. ACM, New York (2013)
7. Aquino Jr. P.T.: PICaP: padrões e personas para expressão da diversidade de usuários no projeto de interação. Escola Politécnica, Universidade de São Paulo, São Paulo. Doctoral dissertation (2008). http://www.teses.usp.br/teses/disponiveis/3/3141/tde-15092008-144412/
8. Masiero, A.A., Aquino, P.T.: Creating personas to reuse on diversified projects. In: Kurosu, M. (ed.) HCI 2015, pp. 238–247. Springer, Cham (2015)
9. De Souza, C.S.: The Semiotic Engineering of Human-Computer Interaction. The MIT Press, Massachusetts (2005). 307 p.
10. Prates, R.O., Barbosa, S.D.J.: Introdução à Teoria e Prática da Interação Humano Computador fundamentada na Engenharia Semiótica. In: Kowaltowski, T., Breitman, K. K. (eds.) (org.) XXVII Congresso da Sociedade Brasileira de Computação. Jornadas de Atualização em Informática (JAI), pp. 263–326 (2007)
11. de Souza, C.S., Leitão, C.F.: Semiotic engineering methods for scientific research in HCI. Synth. Lect. Hum.-Centered Inform. **2**(1), 1–122 (2009)
12. Barbosa, S.D.J., da Silva, B.S.: Interação Humano-Computador. Elsevier, Rio de Janeiro (2010). 408 p.
13. Peirce, C.S.: Semiótica. Editora Perspectiva, São Paulo (2012). 337 p.
14. Bhutkar, G., Poovaih, R., Katre, D., Karmarkar, S.: Semiotic analysis combined with usability and ergonomic testing for evaluation of icons in medical user interface. In: Proceedings 3rd International Conference on Human Computer Interaction, Bangalore, pp. 57–67. ACM, New York (2011)
15. de Souza, C.S., Leitão, C.F., Prates, R.O., da Silva, E.J.: The semiotic inspection methods. In: Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems, Natal, pp. 148–157. ACM, New York (2006)
16. de Reis, S., Prates, R.O.: Applicability of the semiotic inspection method: a systematic literature review. In: Proceedings of 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction (IHC + CLIHC 2011), Porto de Galinhas, pp. 177–186. Brazilian Computer Society, Porto Alegre (2011)
17. Bim, S.A.: Obstáculos ao ensino dos métodos de avaliação da Engenharia Semiótica. 2009. 181f. Tese (Doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro (2009)
18. ITU Internet Reports. International Telecommunication Union. ITU Internet Reports 2005: Ubiquitous Network Societies: Their Impact on the Telecommunication Industry, Geneva (2005)
19. International Telecommunication Union. ITU Internet Reports 2005: The internet of things, Geneva (2005)
20. Monteiro, I.T.: Autoexpressão e engenharia semiótica do usuário-designer. 2015. 312. Tese (Doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro (2015)
21. Suárez, L.L.: Conhecimento Sensorial – Uma análise segundo a perspectiva da Semiótica Computacional. 2000. 156f. Dissertação (Mestrado) – Universidade Estadual de Campinas, Campinas (2000)

22. do Amaral, V., Ferreira, L.A., Aquino Junior, P.T., de Castro, M.C.F.: EEG signal classification in usability experiments. In: 2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), Rio de Janerio, pp. 1–5 (2013). doi:10.1109/BRC.2013.6487469
23. Goebbels, G., Aquino Jr., P.T., Lalioti, V., Goebel, M.: Supporting team work in collaborative virtual environments. In: Proceedings of ICAT 2000 - The Tenth International Conference on Artificial Reality and Tele-existence (2000)
24. Aquino Jr., P.T., Filgueiras, L.V.L.: A expressão da diversidade de usuários no projeto de interação com padrões e personas. In: VIII Brazilian Symposium on Human Factors in Computing Systems (IHC 2008), pp. 1–10. S. B. Computação, Porto Alegre (2008)
25. de Reis, S.: Avaliação do Método de Inspeção Semiótica. 2012. 224f. Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Belo Horizonte (2012)

Thing to Service: Perspectives from a Network of Things

Sharon S. Chinoy[1,2](✉)

[1] Harris Corporation, Melbourne, FL, USA
schinoy1986@my.fit.edu
[2] Florida Institute of Technology, Melbourne, FL, USA

Abstract. As technology evolves to the Internet of Things (IoT), understanding context of the information about and from those things can be challenging. Numerous distributed heterogeneous devices reporting large quantities of data, and that may be interconnected to provide evolving services to multiple stakeholders will be characteristic of IoT ecosystems. Although the IoT is in the incubation phase, we can gain perspectives from current Networks of Things. In this article the author defines characteristics of complex domains that affect network management, including a case study of the Federal Aviation Administration (FAA) National Airspace System (NAS). Today's tools, standards, and models are not sufficient to capture operating domain knowledge consistent to these characteristics. The author offers a new model to serve as a framework to obtain domain knowledge context for the individual resources and their applicable services, while simplifying complexity needed for effective maintenance management.

Keywords: Complex networks · Internet of Things · Network management · Services

1 Introduction

The IoT is a growing field as the Internet, low cost devices, and cloud storage become ubiquitous. An objective of the IoT is to enable resources to collect and share sensor data through Internet communication. A variety of resources may be connected, including appliances, smart meters, smart phones, and autonomous vehicles. As data becomes more readily available, new services will be created for various stakeholders.

Organizations who maintain and manage current large domain infrastructures are challenged by these same characteristics (i.e., increases in resource number, resource variety, data quantity, and data consumers). Today's tools, standards and models for network management systems fall short in accommodating these conditions. It is proposed that context must be integrated into a maintenance management system design in order for it to provide useful information and enable efficient maintenance operations within a complex domain. A new model is needed to serve as a framework for capturing context into network management tools.

Section 2 of this paper discusses the need for context and defines complex domain characteristics that affect network management. Section 3 describes current tools,

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_41

standards, architectures, and models and why they are insufficient to accommodate complex domains characteristics. Section 4 proposes a new model as a framework for capturing contextual domain information. Section 5 discusses future research to use the model to develop an ontology for network management.

2 Context for Network Management

Figure 1 illustrates general maintenance operations functions. When system events and trends are reported, the operators must be able to detect and isolate conditions, assess their impact to domain mission services, coordinate maintenance actions with stakeholders, and restore the condition back to normal operations. These same activities are performed for planned maintenance activities. Data from system events and processes are then used for analysis to measure service level agreements or improve the architecture or processes.

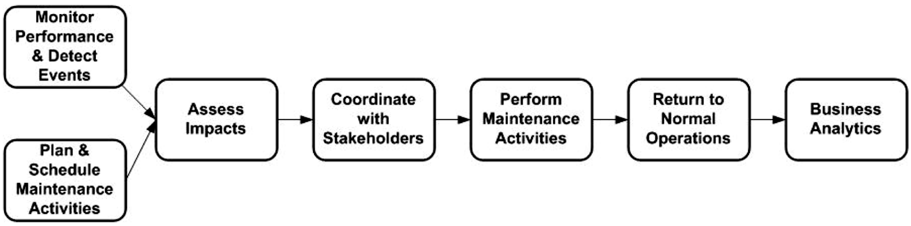


Fig. 1. General network management operations.

Although this may appear as a simple process, network maintenance is increasingly complex in many domains: cyber security analysis; incident response; power transmission and distribution; Supervisory Control and Data Acquisition (SCADA) systems; oil and gas, petroleum, and mineral processing plants; and telecommunications and Internet Service Provider networks. For critical systems, domain stakeholders are responsible for ensuring overall mission operations during anomalous conditions. The maintenance organization must work in conjunction with operational stakeholders to assess impacts, determine if it is better to continue using the failed resources and when corrective actions can begin. The impacts may result in operational stakeholders subsequently initiating mitigation actions. Furthermore, external stakeholders may be impacted by an outage or may be responsible for conducting the recovery.

Central to effective decision-making is context. Even when the maintenance organization detects a failure, context guides the course of action for recovery. For example, if redundant entities exist, they can mitigate the event to continue operations. In addition, if a single outage occurs, it may necessitate a different response than multiple resource outages. Even for the same resource, situational context at event occurrence may warrant that repair be delayed, the failure to be considered insignificant, or urgent restoral needed as soon as possible. Domain operating rules establish event criticality and response, and the response can vary within context [1, 2].

Cognitive processing towards decision-making balances action with information known at the time, weighing potential outcomes and risk [3, 4]. In order to make effective decisions, operators must have current information, be able to integrate information into patterns, and compare it to goals. Operators want automation that acts as a partner to integrate, interpret and correlate data within context, indicates trends, helps to understand and respond to situations, and shares data across stakeholders [5]. Context is essential for providing needed information.

2.1 Characteristics of Complex Domains

A literature review was performed to establish a contextual foundation for network management in complex domains. Table 1 summarizes complex domains' characteristics that influence maintenance [5–23].

2.2 Case Study of the FAA National Airspace System (NAS)

This section describes the FAA NAS in reference to the categories in Table 1.

Pace. The NAS is operational 24 hours a day, seven days a week. At the macro level, the NAS supports an aviation industry that creates $1.5 trillion to the U.S. economy and serves approximately 70,000 daily flights and 756.3 million passengers per year [24]. At the meso level, the NAS enables air traffic controllers to control traffic and manage traffic flow, stakeholders to obtain weather and flight information, and pilots to navigate and land aircraft. At the micro level, numerous individual systems provide capabilities and are maintained by the Technical Operations (Tech Ops) organization.

Architecture. The FAA's NAS is a complex System of Systems. NAS systems include at least 168,000 managed devices [25, 26]. Many contractors developed the systems over many decades. This resulted in numerous technologies, protocols, interfaces, and maintenance tools. Currently, many NAS systems are not natively Internet Protocol (IP). Configurations of the same system can vary across locations.

Location. Systems are geographically distributed across the U.S. Some are in remote locations, difficult to access.

Resiliency. Since safety is of utmost importance, the FAA institutes multiple levels of redundancy to continue service in degraded operations. Depending on significance, this can range from no redundancy, a redundant device (e.g., card, power supply) or circuit, a redundant subsystem, a duplicate system, or a unique backup system.

Data Flows. Emerging architectures are resulting in multiple systems interfacing to provide a single end-to-end service. Capturing and maintaining these service flows can be challenging, especially when they cross FAA programs, and changes occur over time. As the System Wide Information Management system becomes operational, NAS data will be more widely available to consumers who may be internal or external to the FAA.

Services. The systems provide many NAS services. Services can vary in importance. As a result of increased data availability, new services are being offered and new roles for stakeholders may emerge (such as the Airline Operations Centers).

Table 1. Characteristics of complex domains that affect maintenance management

Category	Characteristics
Pace of domain	• Highly dynamic operations with a fast time factor • May provide critical capabilities (e.g., health care, air traffic control, nuclear), such that maintenance must be coordinated with stakeholders
Architecture	• "System of Systems" as a result of decades of system procurement • Many various technologies • Technologies may change often with obsolescence or software updates, or may remain constant for years
Location	• Distributed systems in multiple geographical locations • Some locations may be difficult to reach (e.g., on a mountaintop)
Resiliency	• Redundant elements • Backup systems
Data flows	• Complex data flow paths use multiple elements for end-to-end connectivity • May traverse external domains for end-to-end connectivity • May use fixed system elements or may vary based on dynamic routing

Services	• Resources support services that accomplish the domain's mission
	• Evolution towards "service ecosystems"
	  • Capabilities obtained or offered as a service (e.g. XaaS) in lieu of owning products (e.g., infrastructure (IaaS), platform (PaaS), software (SaaS)
	  • Services as knowledge products offered by sharing and transforming data through multiple interfacing networked devices, perhaps across domains (e.g., weather data)

Insufficient tools	• Independent software and workstations • Complex data not fused • Physical and functional information not correlated • Alarms not correlated to root cause • Too many alarms, especially transient, false, or low priority alarms • Do not determine recovery alternatives • Do not determine event significance to mission context • Do not project future events
Range in operator domain knowledge and capabilities	• Experts with implicit knowledge • Those who lack site specific knowledge • Operators lacking detailed knowledge of the technology • Distrust of the technologies • Having to maintain multiple mental models
Various organizational structures	• Centralized control with increased responsibilities and coordination • Distributed work where stakeholders have different roles, but their success depends on collaboration • Multiple organizations responding as one, requiring sharing of people and technological resources (e.g., in incident response)

Stakeholders	• Deficient communications among stakeholders
	• Increasing number of stakeholders
	• External stakeholders
	  • Third party provides service to domain
	  • Domain products offered as services to other parties

Tools. Maintenance tools lack integration for end-to-end service management, and do not offer features for prediction, decision support, or analytics. When assessing planned or unscheduled outages, this requires personnel to mentally correlate systems' components forming a service thread, with current and projected status [27].

Operator Knowledge and Capabilities. In the case of the FAA, Tech Ops must be able to isolate failure occurrence, and assess its impact in order to coordinate with Air Traffic (AT), who must then make operational decisions (e.g., position reconfigurations or traffic flow changes). The impact is based on NAS service significance, whether backups are available to continue functionality, and the number of resource outages. Operational decisions are also influenced by traffic volume and weather. As the number of equipment items increases, Tech Ops perceives their workload as increasing. When problem solving, they have difficulty understanding equipment functional relationships and system-to-system relationships [25].

Organization and Stakeholders. The NAS is used and supported by multiple organizational groups within AT and Tech Ops, each with dedicated responsibilities. Tech Ops and AT organizations are geographically distributed. Tech Ops may work in the same facility as AT personnel, or may work from a central control facility that supports multiple AT locations. Although the FAA is responsible for managing the NAS, they may need to coordinate with third parties. In some cases, service providers offer and maintain capabilities (e.g., for telecommunications or for Flight Services). In other cases, FAA equipment is located at places that are not FAA owned (e.g., airports, contract towers, or commercial locations). In addition, data is provided to external users (e.g., airlines) and may be integrated within their tools. When there is a service outage, external users may contact the FAA for troubleshooting.

There is often a lack of a common frame of reference between stakeholder groups due to terminology and operating differences [28]. Various stakeholders need shared situation awareness, yet information tailored to their perspectives and goals.

3 Designing Maintenance Systems

The maintenance design of a complex life-critical system is an important design aspect in order for personnel to effectively maintain availability. The question is how to provide information to enhance situation awareness and be effective to accommodate stakeholder goals and perspectives within context. Design considers system requirements, commercial software products, standards, and models.

3.1 System Requirements

Many complex domains have become System of Systems architectures, evolved from years of procurements. Often each new procurement defined maintenance requirements specific to its system and the latest technological advances, in isolation from other systems or as an enterprise. This is true for the FAA's NAS. Because there have been no overarching enterprise requirements for maintenance design, each NAS system has a unique approach, and the maintenance tools are independent from each other. This causes increased workload.

One challenge for defining enterprise level maintenance requirements is that the system itself may impose unique information and control requirements. For example, a maintenance design for a satellite system necessitates different status and controls than a computer processing system or a radar system or a simple sensor. Yet each system may be used concurrently within one domain. In some cases a system may be a new innovation, and maintenance tool requirements may emerge over time.

Another challenge in defining overarching requirements is understanding and incorporating various stakeholder perspectives, especially as domains become service ecosystems [29].

3.2 Tools

Current commercial tools tend to provide silos of network management information (e.g., element and server managers, or bandwidth and traffic usage). Some can only monitor IP devices, which limits capabilities for large complex domains comprised of both analog and digital devices. Even if the infrastructure is purely IP, presenting devices solely by their IP addresses provides limited information to support the network operators. In addition, these tools do not enable information fusion of end-to-end components that form data flows and service threads, redundancy relationships, and other domain knowledge to support maintenance personnel's cognitive functions for tracking patterns in fault propagation [3, 30]. These patterns aid fault isolation and impact assessment [31]. Domains either resort to managing their enterprises through multiple tools (increasing operator workload), or providing custom solutions designed with a unified user interface accessing data from a suite of commercial tools.

At the core of a network management system is its ability to receive an event message or parameter value, compare it to a threshold, and flag nominal or outage conditions. Historically it was much easier to assess whether system resources were usable (i.e., up) or not usable (i.e., down). As systems become IP enabled, the definition of an outage is more obscure. For example, Quality of Service (QoS) factors (e.g., latency and jitter) influence performance decisions, but there is no distinct threshold for declaring when it becomes unacceptable [32]. Another example is at the application layer where a software product may rely on multiple data sources. If one data source is not available, is the software product still usable? The answer may be domain and situation specific.

Even today, tools present too many event messages, in spite of the Three Mile Island accident raising awareness that excessive and confusing alarms can contribute to catastrophic consequences [33, 34]. Events should be filtered and categorized for importance to the domain.

In a study specific to an FAA maintenance tool, the researchers identified display information important to the maintenance operators. These included site (location), equipment, status, event timestamp, description, and criticality [35]. In addition, users wanted data filtered to facilities only within their responsibility.

To promote situation awareness, tools must provide useful, reliable information, in a format that is easily perceived and interpreted [23, 36].

3.3 Standards and Guidelines

Network management standards tend to focus on protocols for (1) obtaining device status (e.g., Simple Network Management Protocol (SNMP), Syslog protocol, Common Management Information Protocol (CMIP)), for (2) monitoring IP data flows (e.g., IP Flow Information eXport (IPFIX) protocol), or for (3) IP performance metrics [37]. While it is necessary to have mechanisms to detect device status information, it is not sufficient for today's complex domains that require more contextual information for equipment configurations and impact to services and stakeholders.

Human engineering standards for designing maintenance user interfaces tend to focus on aspects related to hardware configurations (e.g., weight, reach and access) and general user interface characteristics (visual and audible alarms, color-coding and labels) [38, 39]. These standards, as well, focus on a limited set of factors for a modern maintenance organization. The FAA recognizes the need for better guidance. In one study they assessed status coding of various deployed tools to better define alarm, alert, normal, and unmonitored conditions, and high-level events that may result in those conditions. However, the researchers note that even if the recommendations are implemented, it does not ensure the maintenance tool's design is effective [40].

3.4 Models

Network management related models offer frameworks for network communication (e.g., Transmission Control Protocol (TCP)/IP protocol suite, or the Open System Interconnection (OSI) model), architectures, and telecommunication function models. Architecture and function models offer some level of context, but their perspective is towards the telecommunications service provider and its interaction in delivering and managing ordered telecom, not the perspective of the customer enterprise in which telecom is part of the end-to-end service supporting a domain's mission. Common architecture and function models include: (1) the Telecommunications Management Network (TMN) framework; (2) the Fault, Configuration, Accounting, Performance, Security (FCAPS) model; (3) the Fulfillment, Assurance, Billing (FAB) model; and (4) the Information Technology Infrastructure Library (ITIL).

Classical models from human performance research in complex socio-technical systems are Rasmussen's Skills, Rules, Knowledge (SRK) model and the Abstraction Hierarchy (AH) model. As systems become more complex, workers are transforming from hands-on skilled workers of a mechanical era to knowledge workers in an information-based era. Human performance relies on knowledge work. When faced with an unfamiliar situation or one that involves uncertainty and risk, the operator uses his mental models to determine an action plan, while fusing conditions, rules, and know how. By capturing information that supports domain knowledge, less experienced workers can more readily gain expertise, and all can make more effective decisions.

In the AH model, cognitive processes for decision-making in response to anomalous system events rely on transformation strategies of system physical to functional properties [41]. This model is foundational for recognizing that maintenance operators make decisions with various types of system information, but it fall shorts in addressing distributed heterogeneous devices, expanding service ecosystems and multiple stakeholders' perspectives.

4 The Thing to Service Model

Based on insights from complex domains, the Thing to Service Model is proposed as a holistic framework to elicit and structure information for an enterprise network within a distributed, inter-twined service ecosystem (Fig. 2).

By considering the resources physical attributes in relation to service usage, mission operating rules, and key stakeholders (i.e., "social objects") it can bridge the gap for collecting domain contextual information needed for decision-making [15, 36]. The model proposes that multiple perspectives are needed to support stakeholders. Information from any perspective (or multiple perspectives) may be retrieved at any time, in support of current tasks and goals. An event, time, and situation context affect which perspective is requested for decision-making. An event may include performance data, a failure indication, or the necessity to schedule a maintenance activity. Time may reflect current or future time or a time range in conjunction with other simultaneous

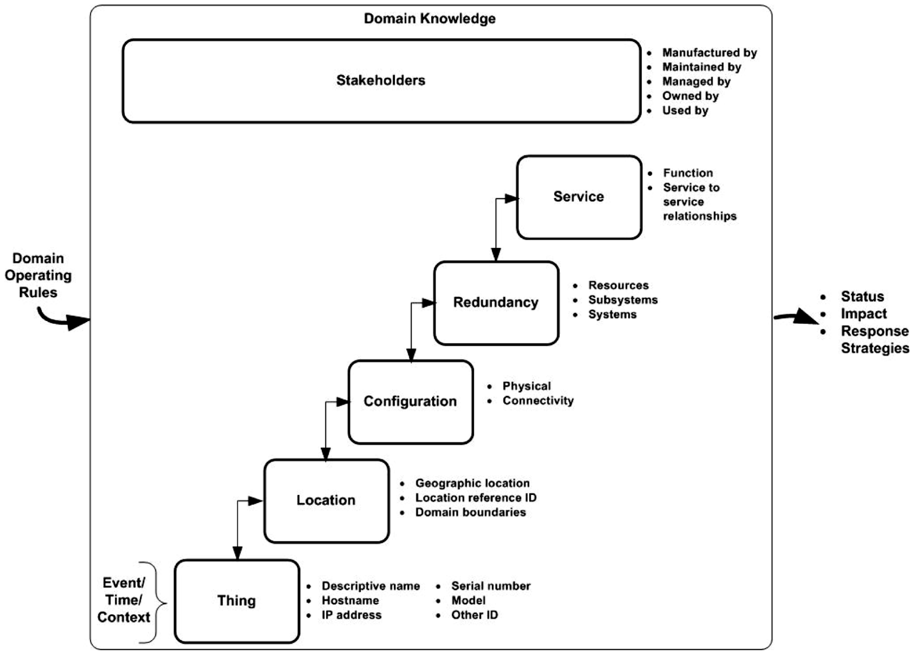


Fig. 2. Thing to service model

contextual situations (e.g., other outages or weather conditions). Domain operating rules define how data is used, displayed, and interpreted. The rules in conjunction with the perspectives determine status at the various levels, impact to services, and response strategies.

Technical staff must be able to assess resource performance in order to recognize failures and analyze trends. At the Thing Perspective, resources may be identified by multiple naming conventions (e.g., a descriptive name, hostname, IP address, or MAC address). Each relevant name may be useful at different times when accomplishing decision-making tasks or corrective actions. Historically, network management tools focused on managed devices with alarm mechanisms or performance trend data collection. However, this alone provides limited information in complex domains.

When an outage occurs, maintainers must be able to locate the device, which is provided through the *Location Perspective*. Location information is typically thought of as a latitude/longitude or postal address. Geographic location information is particularly important when the enterprise is geographically distributed. Contextual location information may include domain specific operating boundaries such as airspaces, regions, or landmarks (e.g., airports). Location information is useful for grouping resource responsibilities in large distributed networks, assessing if multiple outages are site or area specific, and informing technicians where to go to fix the issue.

The next level is the *Configuration Perspective*, which captures resource configuration information. This perspective can represent a physical configuration and connectivity to provide an end-to-end capability. The Configuration Perspective can be the hardest to represent in complex domains, especially if elements are dynamically added, changed, or removed over time and vary across installations. A resource may be as simple as a sensor or router. In more complex configurations, it may represent a card within a module within a shelf within a rack within a subsystem within a system. It may also be atypical configurations such as an antenna assembly. The resource may have connectivity interfaces (e.g., interfaces to other system elements, cabling, Local Area Networks, and Wide Area Networks). Connectivity information can be more difficult to capture and assess when it is combined with the next level, the Redundancy Perspective.

The *Redundancy Perspective* provides relationships of redundant configurations, whether at the resource, subsystem, or system level. Where configurations have a dedicated redundant component (e.g., primary/backup), it is more easily managed. However where there are multiple non-dedicated redundant capabilities (e.g., redundant routers within a network with data dynamically flowing in one of multiple paths), it can be harder to model and assess in real-time.

The highest level is the *Service Perspective*. This perspective indicates usage and encapsulates the other perspectives. The Service Perspective relays the function of inter-connected resources, and can provide operations impact information when assessing outage conditions. It may also have inter-connected service relationships, such that one service is used as input to create a subsequent service, and so forth in the service ecosystem. If it is not possible to capture all inter-relationship data flows from one service to another, when problems occur it can be difficult to identify root cause or spanning of affected parties. An end user may report a problem, but the problem may actually occur at lower levels of the data flow.

Overarching all perspectives is the *Stakeholder Perspective*. This perspective represents the service ecosystem's multitude of stakeholders. While the contents of this article have focused on the maintenance organization (e.g., Tech Ops) and the primary users (e.g., Air Traffic Control), other stakeholder information may need to be captured. In some cases, resources may be managed by the domain's maintenance organization, but maintained or owned by a service provider (e.g., a telecommunications company). Or resources may be maintained by the domain's maintenance organization, but owned by another party (e.g., an airport authority). Furthermore, there may be several service consumers that extend beyond the primary users (e.g., Airline Operations Centers or industry partners with third party tools that provide information to other customers). When an outage occurs, it may be necessary to identify, notify, and coordinate with many stakeholders.

The Stakeholder Perspective can also support tailoring display information. For example, some stakeholders may be restricted or may desire to only view resources for which they use or are responsible. Other stakeholders may not want to see any resources, but rather only services.

Although this model has been described for usage in network management, it could also serve as a framework for collecting foundational information for IoT 'big data'.

5 Future Research

Future research will investigate using the Thing to Service model as a framework for defining operational context in a network management ontology. Ontologies offer a means to represent domain knowledge through objects, properties, and relationships, using common semantics. Benefits include knowledge-sharing, flexibility for multiple perspectives, inferencing of concepts, reasoning, and decision support [42–44].

6 Conclusion

As discussed, there are several challenges for network management design in large mission critical domains. Current tools, standards, and models are not sufficient for defining operational context to support real-time decision-making and response of sensor data. Service ecosystems are evolving with distributed heterogeneous interfacing resources, self-adapting services, and emergent stakeholder relationships. This will be characteristic of the IoT. A new holistic paradigm is needed to provide tangibility to these ecosystems. The Thing to Service Model offers a framework to characterize domain knowledge related to resources, configurations, services, stakeholders, and relationships between them. The framework is scalable for a multitude of devices across a distributed network and service ecosystem. The model serves as a foundation for capturing relevant information that can subsequently be used for user interface presentation and business analytics. As the desire for "smart" products and environments continues, capturing domain-specific knowledge will remain the challenge for obtaining relevant and useful information.

References

1. Mumaw, R.J., Roth, E.M., Vicente, K.J., Burns, C.M.: There is more to monitoring a nuclear power plant than meets the eye. Hum. Factors **42**(1), 36–55 (2000)
2. Adkisson, L., Karna, K., Katz, D., Karna, A., Dontas, K.: Identification of artificial intelligence (AI) applications for maintenance, monitoring, and control of airway facilities. Technical Note DOT/FAA/CT-TN92/41, Federal Aviation Administration (1994)
3. Woods, D.D.: Cognitive demands and activities in dynamic fault management: abductive reasoning and disturbance management. In: Stanton, N. (ed.) Human Factors in Alarm Design, pp. 63–92. Taylor and Francis Ltd., London (1994)
4. Grote, G.: Management of Uncertainty. Springer, London (2009)
5. Bradshaw, J.M., Carvalho, M., Bunch, L., Eskridge, T., Feltovich, P.J., Forsythe, C., Hoffman, R.R., Johnson, M., Kidwell, D., Woods, D.D.: Coactive Emergence as a Sensemaking Strategy for Cyber Operations (2012). http://cmapsinternal.ihmc.us/rid=1MJNFT79F-1R8PZR9-1D4R/Coactive%20Emergence-2012.pdf
6. Vargo, S.L., Lusch, R.F., Wieland, H., Archpru Akaka, M.: Knowledge service engineering: a service-dominant logic perspective. In: Kantola, J., Karwowski, W. (eds.) Knowledge Service Engineering Handbook, pp. 139–153. CRC Press, Boca Raton (2012)
7. Ahlstrom, V., Hartman, D.G.: Human error in airway facilities. Report DOT/FAA/CT-TN01/02, Federal Aviation Administration (2001)
8. Ahlstrom, V., MacMillan, J., Tenney, Y.J., Pew, R.W., Cranston, R.L., Mogford, R.: Human error mitigation in future operations control centers. Technical Note DOT/FAA/CT-TN99/14, Federal Aviation Administration (1999)
9. Truitt, T.R., Ahlstrom, V.: Situation awareness in airway facilities: replacement of maintenance control centers with operations control centers. Technical Note DOT/FAA/CT-TN00/09, Federal Aviation Administration (2000)
10. Antonovsky, A., Pollock, C., Straker, L.: Identification of the human hactors contributing to maintenance failures in a petroleum operation. Hum. Factors **56**(2), 306–321 (2014)
11. Burns, C.M., Barsalou, E., Handler, C., Kuo, J., Harrigan, K.: A work domain analysis for network management. In: Proceedings of the IEA 2000/HFES 2000 Congress, I, vol. 200, pp. 469–472 (2000)
12. Chow, R., Vicente, K.J.: A field study of collaborative work in network management: implications for interface design. In: Proceedings of the HFES 45th Meeting, pp. 356–360 (2001)
13. Close, D., Babski-Reeves, K., Younan, N., Schulz, N.: Cognitive task analysis of power grid monitors. In: Proceedings of the HFES 53rd Annual Meeting, vol. 53, pp. 304–308 (2009)
14. D'Amico, A., Whitley, K., Tesone, D., O'Brien, B., Roth, E.: Achieving cyber defense situational awareness: a cognitive task analysis of information assurance analysts. In: Proceedings of the HFES Annual Meeting 2005, pp. 229–233 (2005)
15. Handley, H.: A network model for human interoperability. Hum. Factors **56**(2), 349–360 (2014)
16. Johnson, C.W.: Understanding failures in international safety-critical infrastructures: a comparison of European and North American power failures. In: Mohan, D.J., Mullane, M. (eds.) 26th International Conference on System Safety, Vancouver, pp. 25–29 (2008)
17. Kuo, J., Burns, C.M.: A work domain analysis for virtual private networks. In: 2000 IEEE International Conference System, Man and Cybernetics, Nashville, TN, vol. 3, pp. 1972–1977 (2000)

18. Lenox, M.M., Connors, E.S., Endsley, M.R.: A baseline evaluation of situation awareness for electric power system operation supervisors. In: Proceedings of the HFES Annual Meeting 2011, pp. 367–378 (2011)
19. Li, X., Horberry, T., Powell, M.: Human control in mineral processing plants: an operator centered investigation. In: Proceedings of the HFES Annual Meeting, vol. 54, pp. 284–288 (2010)
20. Li, X., Powell, M.S., Horberry, T.: Human factors in control room operations in mineral processing: elevating control from reactive to proactive. J. Cogn. Eng. Decis. Making **6**(1), 88–111 (2012)
21. Pacley, A.L., Reichert, M.E.: Advanced technology master caution panel (ATMCP): tracking task status by monitoring information technology (IT) resources. In: Proceedings of the 48th HFES Annual Meeting, pp. 846–850 (2004)
22. Romero, D.D., Graven, T.G.: Visual representation of connectivity information for efficient system understanding. In: Proceedings of the 57th HFES Annual Meeting, pp. 2102–2016 (2013)
23. Stanton, N.A., Salmon, P., Jenkins, D., Walker, G.: Human Factors in the Design and Evaluation of Central Control Room Operations. CRC Press, Boca Raton (2010)
24. National Air Traffic Controllers Association: Aviation Facts (2015). http://www.natca.org/index.php/acronyms-breakdown/facts-sheet
25. Pierce, L.G., Byrne, C.L., McCauley, D.: Planning for organization development in operations control centers. Report DOT/FAA/AM-12/6, Federal Aviation Adminstration Civil Aerospace Medical Inst (2012)
26. Harris Corporation: FAA Telecommunications Infrastructure Data Sheet (2013). http://download.harris.com/app/public_download.asp?fid=521
27. Chinoy, S.S., Fischer, D.J.: Advancing situational awareness for technical operations. In: Proceedings of 56th Annual ATCA Conference (2011)
28. Ahlstrom, V., Muldoon, R., Dorsey, H.: Factors affecting the communication of status information between technical operations and air traffic personnel. Technical report DOT/FAA/TC-09/01, Federal Aviation Administration (2009)
29. Nightingale, D.J., Rhodes D.H.: Enterprise systems architecting: emerging art and science within engineering systems. In: MIT Engineering Systems Symposium, Cambridge, MA (2004)
30. Stanton, N.: Alarm initiated activities. In: Stanton, N. (ed.) Human Factors in Alarm Design, pp. 93–118. Taylor and Francis Ltd., London (1994)
31. Olsen, M., Parcel, A.: Conceptual model for monitoring nextgen faa services. In: ATCA Proceedings from the 59th Annual Conference and Exhibition, pp. 70–77 (2014)
32. Caldwell, B.S., Wang, E.: Delays and user performance in human-computer-network interaction tasks. Hum. Factors **51**(6), 813–830 (2009)
33. Wallin, S., Ahlund, C., Nordlander, J.: Rethinking network management: models, data-mining and self-learning. In: 2012 IEEE Network Operations and Management Symposium (NOMS): Dissertation Digest (2012)
34. The President's Commission: Report of the President's Commission on the Accident at Three Mile Island (1979). http://www.threemileisland.org/downloads/188.pdf
35. Hilburn, B., Ahlstrom, V.: Information requirements analysis for remote maintenance monitoring interfaces. Technical report DOT/FAA/TC-12/38, Federal Aviation Administration (2012)
36. Mann, D., Brooks, J., DeRosa, J.: The relationship between human and machine-oriented standards and the impact to enterprise systems engineering. MITRE Technical report MTR100156, MITRE Corporation (2011)

37. Ersue, M., Claise, B.: An Overview of the IETF Network Management Standards. Internet Engineering Task Force (IETF) (2012). ISSN: 2070-1721
38. Department of Defense: Design Criteria Standard: Human Engineering, MIL-STD-1472G (2012)
39. Federal Aviation Admin.: Human Factors Design Standard (HFDS) for Acquisition of Commercial-Off-the-Shelf (COTS) Subsystems, Non-Developmental Items (NDI), and Developmental Systems (2003)
40. Ahlstrom, V., Muldoon, R.: Visual coding of status for technical operations systems. Technical report DOT/FAA/CT-TN04/20, Federal Aviation Administration (2004)
41. Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans. Syst. Man Cybern. **13**(3), 257–266 (1983)
42. Gruber, T.: Ontology. In: Liu, L., Ozsu, M.T. (eds.) Encyclopedia of Database Systems. Springer, Heidelberg (2009)
43. Kolas, D., Self, T.: Towards an effective methodology for rapidly developing component-based domain ontologies. In: Proceedings of Ontology for the Intelligence Community (2009)
44. Noy, N.F., McGuinnes, D.L.: Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory, Stanford University, Stanford (2001)

Designing a Cognitive Concierge Service for Hospitals

Hari Madduri[1(✉)] and Neil Gomes[2]

[1] IBM, Burnet Road, Austin, TX 11501, USA
madduri@us.ibm.com
[2] Thomas Jefferson University and Jefferson Health, Philadelphia, USA
neil.gomes@jefferson.edu

Abstract. This paper and associated talk is based on our experience at Jefferson Health implementing a cognitive patient concierge system. In this paper and presentation we describe: how the environmental constraints in a typical hospital (e.g., network security and privacy considerations) affect the design, how to use current speech recognition technologies to deliver acceptable levels of cognitive accuracy, how to connect various sensors and get their data correlated with speech input, how to interface with the hospitals' information systems and building automation systems (e.g., to control HVAC systems), and how to design for solution enhancement and maintenance. We also illustrate other human factors for placement of microphones, speakers and controls in patient rooms for both ergonomic and hygienic reasons.

Keywords: Cognitive concierge · In-patient concierge · Cognitive patient care

1 Introduction

Artificial Intelligence (AI) has been around since the 1950s. However, consumer products with cognitive capabilities have become mainstream only recently. Amazon's Echo, Google's Google Home, and other such devices are now becoming increasingly popular. These are speech-based and users simply talk to the device to either get information or accomplish tasks. The device answers questions with appropriate audible information. The device fulfills simple tasks by asking compatible devices and systems to do their part (e.g., like setting the thermostat to a given temperature). While most of these speaker-based products primarily target consumers at home, they have also begun to attract considerable interest from businesses in healthcare, hospitality, retail, etc. Many leading hotel chains, hospitals and office building managers are experimenting with natural language interactions to get work done without involving a lot of people. Natural language understanding by systems requires some cognitive capabilities which is a part of AI.

While we hear terms like AI, Cognitive and Machine Learning frequently these days, there aren't any widely-accepted common definitions for these. AI in the past was broadly used to indicate anything that showed any kind of intelligence in machines. Today, there are different classifications of AI. The classical AI is called general-purpose AI where the objective is to mimic the general intelligence of humans. Another

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2_42

kind of AI, which is domain-specific or task-specific is now called pragmatic AI [1] and it uses cognitive capabilities and machine learning algorithms to recognize patterns. For our purposes, we will define Cognitive *as something of or relating to intellectual activity (thinking, reasoning, remembering, imagining or learning).*

Pragmatic AI needs to understand and interact with things in the real-world, and to enable this, we can associate it with IoT (Internet of Things), which we can define as *the internetworking* of *physical objects (things) embedded with electronics, software, sensors, actuators, and network connectivity that enables these objects to collect and exchange data.*

The combination of Cognitive capabilities and IoT allows us to infuse intelligence into hitherto "dumb" devices and construct systems/solutions that enable higher end-user satisfaction while reducing the support costs. IBM's Watson is a collection of cognitive capabilities delivered as referenceable/callable cloud services. By appropriately connecting physical objects to the Watson Cognitive IoT platform and a select set of Watson cognitive services, we can build smarter IoT systems (Fig. 1).

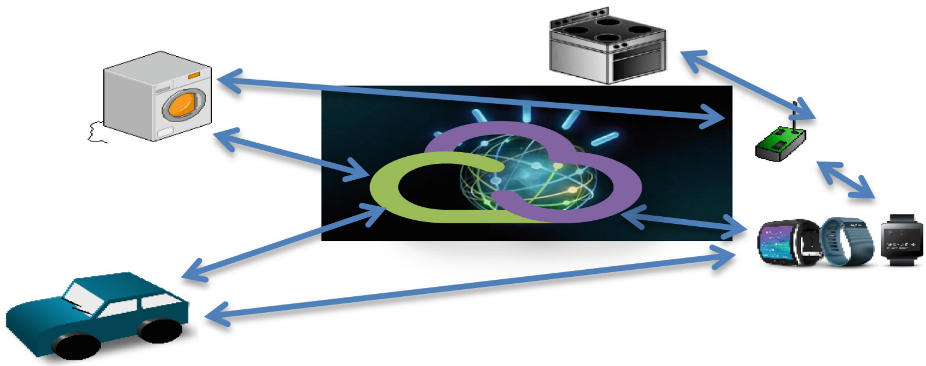


Fig. 1. Objects connected to Watson IOT platform & Watson Cognitive Services

Devices connected to Watson cognitive services can leverage Watson's ability to understand natural language, reason and learn from input data, external events, and user interactions. As a result, the devices become smarter.

In the following sections, we illustrate this by describing in detail a solution we built at Jefferson Health to provide automated in-patient concierge services. We will also talk about design considerations in a typical hospital environment that include network security, patient privacy, services provided by different vendors (e.g., TV network, Building Management, etc.), and ergonomic and hygienic factors relating to IoT equipment. In the last section we talk about extensions to our work and future potential.

2 Application to In-Patient Concierge Service

As illustrated earlier, IoT devices can send data and receive actions from cloud-based cognitive platforms/services and thus act smart or cognitive. A question that we asked ourselves was "Can a patient's room become smart by instrumenting the room with IoT devices and connecting them to cloud-based Watson cognitive services?" The answer is 'yes'. Not only individual devices but also connected spaces like patient rooms can be made cognitive by suitably engineering the connections, software, and operations. The Technology Innovation and Consumer Experience team at Jefferson Health, IBM, and Harman International worked jointly last year to produce the in-patient concierge service illustrated below (Fig. 2).

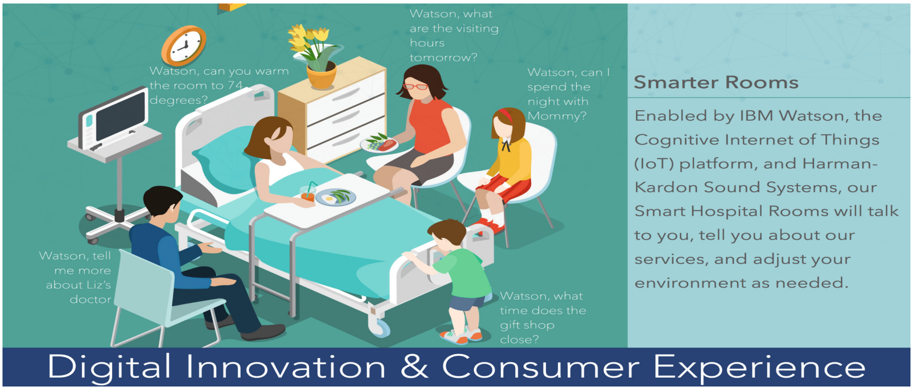


Fig. 2. Concept of in-patient concierge service at Jefferson University Hospital

The idea of an in-patient concierge service aided by IBM Watson cognitive services is fairly intuitive. We place speakers with built-in microphones (custom-developed by Harman International) in patient rooms. Whenever the patient wants to ask a question or request something he/she simply talks to the speaker in a natural manner. The voice is carried to IBM services, the intent (not just the words) is understood and is fulfilled either via a voice response or by a physical action (e.g., turn on a light).

The initial objective of the project was kept fairly modest and the use cases were limited to simple requests for information or requests for simple actions like playing music or turning lights on/off. The primary purpose was to establish feasibility and measure the accuracy and response times. The primary business benefits being targeted were:

- To provide a better patient experience
- Save staff time on tasks that could be automated
- Gather information for further analysis

A summary table of use cases and their expected business benefits is presented below (Fig. 3):

Use Case	Business Benefits
Visiting hours on a given day	Common question. Saves staff time.
Food service times	Common question. Saves staff time.
Information about clinicians	Saves staff time. Reduces patient anxiety.
Turn on/off/dim the lights	Saves staff time. Fewer staff interruptions.
Play ambient sounds…Raise/lower volume…	Saves staff time. Fewer staff interruptions.
Adjust the room temperature* (invoking Building Automation system API)	Saves staff time. Fewer staff interruptions.
Notifications/Reminders (e.g. “remind me to walk every 4 hrs”)	Saves staff time. Fewer staff interruptions. Enhances patient well-being.
Conduct a survey via the speaker and record responses in a database	Quick and easy. Saves time and money to conduct surveys. Data collected can be mined for correlations with IOT data.

Fig. 3. Use cases and business benefits

3 Solution Architecture

Figure 4 shows the solution architecture and the sequence for a simple request use case:

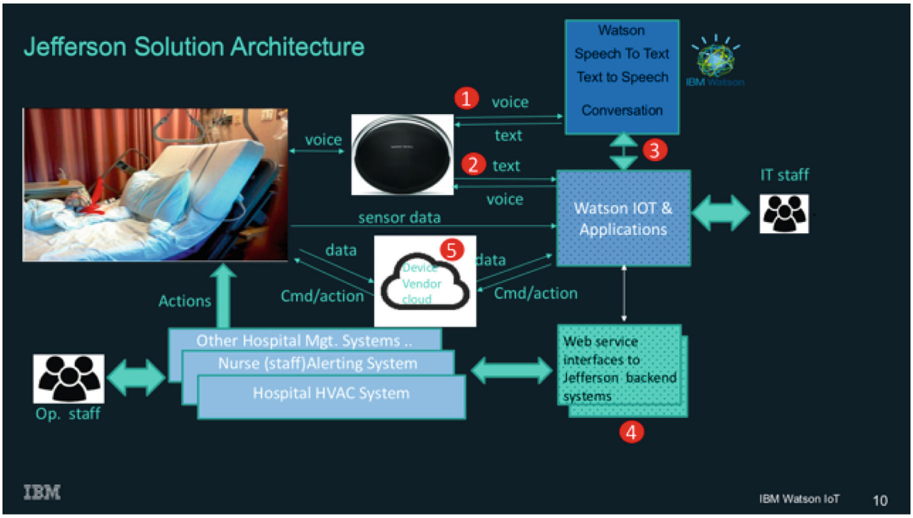


Fig. 4. Solution architecture and a use case sequence

1. The patient asks for information, e.g. “When can my mom visit me?” The audio is sent to the Watson Speech to Text service and is transcribed into text.
2. The text is sent to the Watson Cognitive IoT platform and a custom application built on the Watson IoT platform.
3. The custom application on the Watson IoT platform orchestrates calling one or more Watson services to understand the intent behind the request (the intent is ‘visiting hours’ in this example), and calling fulfillment services provided by Jefferson and returning the answer in speech form back to the speaker.

4. Jefferson provided fulfillment services are implemented with callable REST (Representational State Transfer) interfaces and these serve as a thin veneer over Jefferson information systems (that contain information about visiting hours, doctor biographies, etc.) and other hospital management systems (e.g., building control systems to manipulate temperature in a room).
5. Finally, some use cases may need sensor data to ascertain the current conditions in the room such as current temperature via a sensor. Some sensor devices may be directly connected to the Watson Cognitive IoT platform while some others may be connected to the device vendor's cloud. In the latter case, adapters need to be built to get the sensor data from the vendors' clouds and also to transmit commands to the actuators in the sensors (e.g., to turn off a light/fan/etc.).

The above sequence gives an idea of how a simple use case is implemented. In general, we observed the following design pattern for conversational systems, which applies not only to patients in a hospital but also to other automated concierge services in hotels, offices, etc.

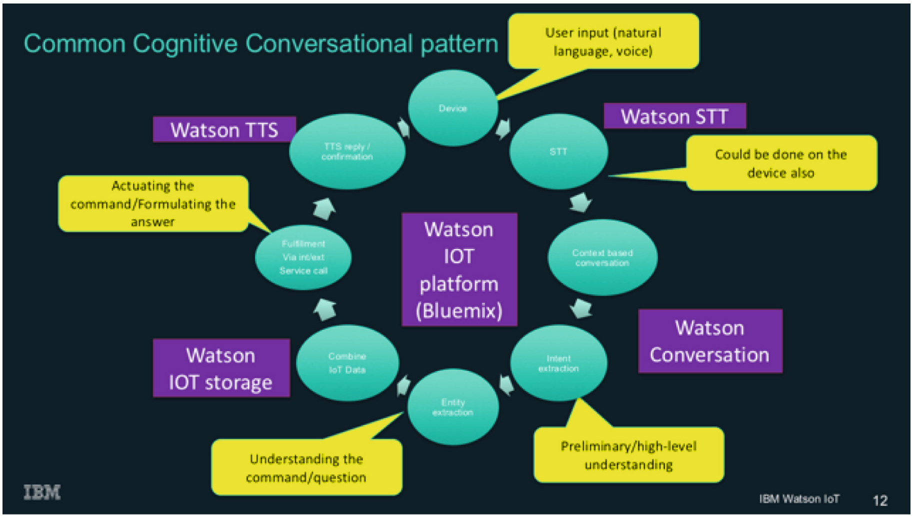


Fig. 5. Common design pattern for cognitive concierge systems

The cycle depicted in Fig. 5 is fairly self-explanatory. The purple boxes indicate the specific Watson services used to fulfill those functions. The Watson IoT platform in the center represents both the platform and the custom applications built on it. This core orchestrates the overall flow and talks to other fulfillment services. Though the above diagram doesn't depict it, the same flow can be used for gathering feedback from a human user and improve the system over time. For example, every time a question is answered, the verbal feedback from the user can be transcribed and understood just like his question/request and it can be stored in a database to train the conversation and speech recognition models to improve their accuracy.

4 Design Considerations in a Hospital Environment

A hospital environment is unique and unlike most other environments in which such a concierge solution could be deployed. As such, we had to consider several factors while designing the software architecture and hardware to deliver this solution. Key considerations are described below:

4.1 Sensitive Personal Information

Most information in a clinical environment such as a hospital could be classified as PHI (Protected Health Information) and is defined as: "any information in a medical record that can be used to identify an individual, and that was created, used, or disclosed in the course of providing a health care service, such as a diagnosis or treatment." [2].

In the early stages of development of the Proof of Concept (PoC), we only handled non-sensitive information to test our hypothesis that the solution would work and be of value to the patient and the organization. However, throughout the design of the solution we planned for accommodating and handling protected health information and thereby meeting regulatory requirements.

We dealt with questions about where we would store the audio files generated via speech before they were transmitted to the Natural Language Processing (NLP) engine to be parsed into text, where that text would be stored so that it could be analyzed by Watson, etc. We also had to consider how and where we would store data so that we could use shared Watson services.

Other considerations in handling sensitive personal information included authenticating users on the device in the room, ensuring that the system is not always-on and listening (which could be viewed as invasion of privacy) but triggered by the user, that the audio output is contained within the boundaries of the room and that patient confidentiality is ensured in replaying recorded information.

4.2 Controlling Equipment in a Hospital Environment

The next challenges we faced were around controlling devices/equipment in the environment of care using the IBM Cognitive IoT (Internet of Things) platform. The first challenge is to ensure that there is a centralized system for building automation controls and it has application programming interfaces (APIs) to control heating and air-conditioning in patient rooms. In some older hospitals and with some vendor equipment (HVAC, TV, etc.) this may turn out to be a big challenge and thus it needs to be addressed early on.

The next challenge area is with sensors and actuators in the room. In phase 1 however, we did implement use cases that connected smart devices (e.g., lights) to a smart hub and controlled lighting in the room. Sensors and actuators, although inexpensive, are not easily connected to platforms they are not natively built for and usually store data in their own cloud platforms, which meant developing software adapters to extract this data and get it to our cognitive IoT solution platform of choice, especially since some service providers don't have open APIs for such data extraction. Thus, it is important to consider the feasibility of integrating sensors/actuators with your chosen cognitive solution platform.

Additionally, and most importantly, hospital environments are highly guarded, secured and controlled. They already have complex equipment that sometimes can interfere with IoT sensors. Thus one has to consider both hospital-network-security compliance and interference with other equipment.

4.3 Importance of Training Cognitive Services

The cognitive services used, like speech-to-text or intent classification, need to be trained with domain-specific data. For example, when a patient asks to know more about a particular doctor, the system should be able to recognize the name of the doctor. Given that speech recognition systems don't know proper names they fail to recognize such utterances. IBM's Watson speech-to-text service has a customization feature which allows it to be trained with domain-specific vocabulary. Likewise, careful design of intent classes and training are necessary to attain high levels of accuracy and thus customer satisfaction.

5 Ergonomic and Hygiene Considerations

The unique hospital environment also presented us with challenges and considerations surrounding hygiene and ergonomics for the device in the room. To this end, Harman International worked closely with us to determine the standards for such devices in the hospital setting and addressed some of the challenges in early prototypes and is continuously improving them. Some of these challenges are highlighted below. Note that in dealing with these challenges in early development, we are producing a solution that is relevant to real hospital settings and is a viable commercial offering.

5.1 Speaker Design Challenges

A key technical challenge with the speaker systems was the ability to manage multiple such systems remotely using a platform and this is where Harman International played a key role in the project. Harman designed a commercial product especially for this initiative that could be connected via Ethernet or WiFi to secure networks in clinical environments and that could be managed remotely using a software platform. For example, we may have to refresh the firmware on a number of speakers all at once, or we may want to selectively enable/disable them, or we might want to wipe the data clean on these devices. This kind of centralized manageability feature vastly differentiates our solution from any bootstrapped systems that use consumer or off-the-shelf products, most of which do not have such features.

Other speaker design challenges included designing a speaker/multi-purpose device (e.g. a clock radio) that would meet hospital hygiene standards and that could be cleaned using harsh (but effective) disinfecting solutions that are common in clinical environments and designing it in a way that it would be best positioned to collect audio and return audio.

6 Future Directions

6.1 Future Vision

We have a lot of exciting future plans for this solution, some of which are described in Figs. 6 and 7 below and include infusing the cognitive platform into other apps such as our "At Your Service" (a custom mobile app developed by and used at Jefferson Health) administrative patient rounding app that enables our leadership to regularly collect feedback from employees and patients, enabling them to effect changes to the environment of care while they are with these employees or patients and do so via the Cognitive IoT platform using voice or application commands. We also plan to remember patient room preferences and use them when patients visit us, providing a more personalized environment.

Fig. 6. Using the cognitive patient concierge with hospital rounding systems

The future for this system is very bright and we are only scratching the surface with these potential use cases. Our leadership and design teams have already defined several other use cases that are now being validated and architected for development in future phases.

6.2 Building a Cognitive Platform

Our current implementation is only a modest beginning. A lot more can be done by exploiting IoT device data in combination with other sources of data. For example, we can provide personalized services by knowing the heating/cooling preferences or

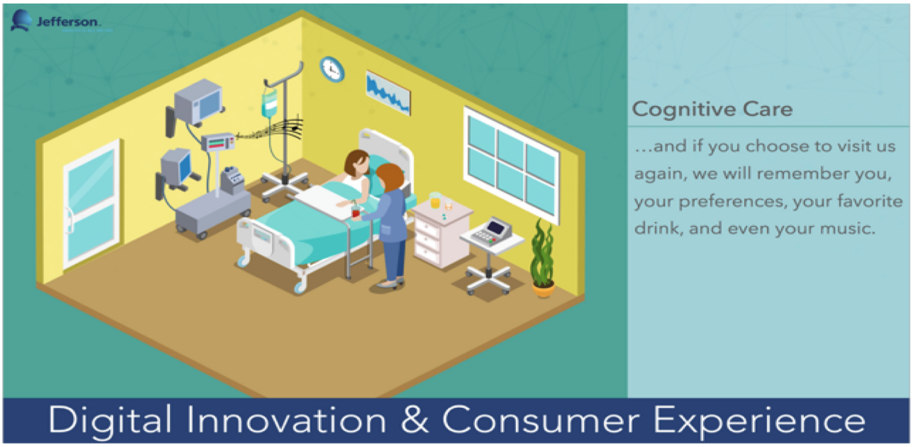


Fig. 7. Using the cognitive patient concierge to deliver personalized care

TV-watching preferences of a patient. We can also make the user's speech interpreted in context instead of running each request through a text classifier mechanically. This makes the verbal interactions more natural. We can also incorporate learning into the system and update the user profile based on past actions requested by a patient and his/her reactions to the responses provided. We can also imagine accumulating data and running analytics on data to anticipate the needs of the patients and catering to them with the intention to engage, delight, and build a reputation for customer service.

These kinds of uses will only grow with time and to be able to support them, we need a flexible platform that allows multiple tools to be used in different combinations. For example, we will need the ability to ingest different kinds of data (structured and unstructured data), store, cleanse, connect with other sources of data, analyze (i.e., run analytics) and generate actionable insights. Furthermore, once these insights are generated they should be easily incorporated into the hospitals' operational systems and business processes.

Key Attributes of the Platform

- Ability to enable natural language conversations (including interrupting the conversation, starting another conversation and then resuming the original conversation). Humans can switch context and so should the systems we build.
- Ability to train for various domains of knowledge. For example, doctor biographies, facility information, curated medical information, etc.
- Ability to extend with varied new capabilities (a.k.a. skills) to accomplish different tasks (e.g., operate TV, lights, curtains, AC, make restaurant reservations, order food, etc.)
- Personalized – ability to know the patient and adapt responses to him/her
- Contextual – human interactions are always understood in context
- Learning – ability to learn based on past experiences and user guidance
- Anticipate and Act – Predicting needs and fulfilling them to delight the user

7 Summary and Conclusion

In this short paper, we have described a cognitive in-patient concierge system that we built as a joint project between Thomas Jefferson University Hospitals, IBM, and Harman International [3]. We have highlighted the design considerations and unique challenges of a hospital environment. We have also indicated solutions and new directions in which this work can be extended. A number of enterprises in various domains are already exploiting (or planning to capitalize on) conversational chat bots for increased differentiation, customer satisfaction, and staff efficiency. We believe we have created a unique, practical, commercial, voice-enabled, and managed solution that is aware of healthcare constraints and requirements. We believe it could also be used in other environments such as hospitality and transportation to deliver a more engaging, interactive, and consumer-centric experience, which also saves time and reduces cost in most such environments.

Acknowledgements. IBM, Jefferson Health, Harman, and Persistent Systems Limited teams.

References

1. Artificial Intelligence: What's Possible For Enterprises In (2017). https://www.forrester.com/report/Artificial+Intelligence+Whats+Possible+For+Enterprises+In+2017/-/E-RES136288?objectid=RES136288
2. Truevault Blog: What is Protected Health Information? https://www.truevault.com/protected-health-information.html
3. IBM Blog: How Watson IoT may transform your next hospital stay, https://www.ibm.com/blogs/internet-of-things/cognitive-announcement-jefferson-hospital/

Author Index

C. Baldwin (ed.), *Advances in Neuroergonomics and Cognitive Engineering*,
Advances in Intelligent Systems and Computing 586, DOI 10.1007/978-3-319-60642-2