Advances in Intelligent Systems and Computing 488

Kelly S. Hale
Kay M. Stanney *Editors*

Advances in Neuroergonomics and Cognitive Engineering

Proceedings of the AHFE 2016 International Conference on Neuroergonomics and Cognitive Engineering, July 27–31, 2016, Walt Disney World®, Florida, USA

Springer

Advances in Intelligent Systems and Computing

Volume 488

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

Kelly S. Hale · Kay M. Stanney
Editors

Advances in Neuroergonomics and Cognitive Engineering

Proceedings of the AHFE 2016 International Conference on Neuroergonomics and Cognitive Engineering, July 27–31, 2016, Walt Disney World®, Florida, USA

Springer

Editors
Kelly S. Hale
Design Interactive, Inc.
Orlando, FL
USA

Kay M. Stanney
Design Interactive, Inc.
Orlando, FL
USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-41690-8 ISBN 978-3-319-41691-5 (eBook)
DOI 10.1007/978-3-319-41691-5

Library of Congress Control Number: 2016944328

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Advances in Human Factors and Ergonomics 2016

AHFE 2016 Series Editors

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

7th International Conference on Applied Human Factors and Ergonomics

Proceedings of the AHFE 2016 International Conference on Neuroergonomics and Cognitive Engineering, July 27–31, 2016, Walt Disney World®, Florida, USA

Advances in Cross-Cultural Decision Making	*Sae Schatz and Mark Hoffman*
Advances in Applied Digital Human Modeling and Simulation	*Vincent G. Duffy*
Advances in Human Factors and Ergonomics in Healthcare	*Vincent G. Duffy and Nancy Lightner*
Advances in Affective and Pleasurable Design	*WonJoon Chung and Cliff (Sungsoo) Shin*
Advances in Human Aspects of Transportation	*Neville A. Stanton, Steven Landry, Giuseppe Di Bucchianico and Andrea Vallicelli*
Advances in Ergonomics In Design	*Francisco Rebelo and Marcelo Soares*
Advances in Ergonomics Modeling, Usability & Special Populations	*Marcelo Soares, Christianne Falcão and Tareq Z. Ahram*
Advances in Social & Occupational Ergonomics	*Richard Goossens*
Advances in Neuroergonomics and Cognitive Engineering	*Kelly S. Hale and Kay M. Stanney*
Advances in Physical Ergonomics and Human Factors	*Ravindra Goonetilleke and Waldemar Karwowski*
Advances in The Ergonomics in Manufacturing: Managing the Enterprise of the Future	*Christopher Schlick and Stefan Trzcielinski*
Advances in Safety Management and Human Factors	*Pedro Arezes*
Advances in Human Factors, Software, and Systems Engineering	*Ben Amaba*
Advances in Human Factors and Sustainable Infrastructure	*Jerzy Charytonowicz*
Advances in The Human Side of Service Engineering	*Tareq Z. Ahram and Waldemar Karwowski*

(continued)

(continued)

Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries	*Sacit Cetiner, Paul Fechtelkotter and Michael Legatt*
Advances in Human Factors in Sports and Outdoor Recreation	*Paul Salmon and Anne-Claire Macquet*
Advances in Human Factors and System Interactions	*Isabel L. Nunes*
Advances in Human Factors, Business Management, Training and Education	*Jussi Kantola, Tibor Barath, Salman Nazir and Terence Andre*
Advances in Human Factors in Robots and Unmanned Systems	*Pamela Savage-Knepshield and Jessie Chen*
Advances in Design for Inclusion	*Giuseppe Di Bucchianico and Pete Kercher*
Advances in Human Factors in Cybersecurity	*Denise Nicholson, Janae Lockett-Reynolds and Katherine Muse*

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 through 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.

The book is organized into five main Parts:

Part I: Influencing Design with Cognitive Engineering and Neuroergonomics
Part II: Cognitive Computing
Part III: Physiological Monitoring and Interaction
Part IV: Theoretical Advances in Cognitive Engineering and Neuroergonomics
Part V: Assessing Cognition and Performance

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 goes to the board members listed below:

M. Fafrowicz, Poland
C. Fidopiastis, USA
C. Forsythe, USA
X. Fang, USA
Q. Gao, China
Y. Guo, USA
P. Hancock, USA
D. Kaber, USA
K. Kotani, Japan
B. Lawson, USA
S. Lee, Korea
H. Liao, USA
Y. Liu, USA
T. Marek, Poland
J. Murray, USA
C. Nemeth, USA
D. Nicholson, USA
A. Ozok, USA
O. Parlangeli, Italy
R. Proctor, USA
D. Rodrick, USA
A. Savoy, USA
D. Schmorrow, USA
N. Stanton, UK
K. Vu, USA
T. Waldmann, Ireland
B. Winslow, USA
G. Zacharias, USA
L. Zeng, USA

It is our hope that professionals, researchers, and students alike find the 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 do not 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.

Orlando, FL, USA
July 2016

Kelly S. Hale
Kay M. Stanney

Contents

Part I Influencing Design with Cognitive Engineering and Neuroergonomics

Cued Recall with Gaze Guiding—Reduction of Human Errors with a Gaze-Guiding Tool 3
Barbara Frank and Annette Kluge

Investigation of the Recognition of Different Font Sizes on Human-Machine Interface and Physiological Characteristics for Aged People 17
Shengwen Luo, Li Ding, Linghua Ran and Yan Li

Cognitive Ergonomics Applied to the eQRH: Developing an Electronic Quick Reaction Handbook for Use During Aviation Emergencies 27
Marcia Crosland, Wesley Wang, Jerry Ray, Stuart Michelson and C.J. Hutto

The Impact of Chinese Cultural Elements on the User Experience in Government Website Design 39
Tian Lei, Xu Liu, Lei Wu, Ziliang Jin and Yuhui Wang

Research on the Influence of Abstract Knowledge to the Individual Cognitive Behavior and Innovative Design Thinking 51
Yongjia Zhou, Qing Xue, Jia Hao and Min-xia Liu

How the Color Matching and Arrow Position Affect the Design of the Bus Stop Board: An Eye Movement Experimental Research 65
Chuanyu Zou, Na Lin and Yunhong Zhang

Measuring the Amplitude of the N100 Component to Predict the Occurrence of the Inattentional Deafness Phenomenon 77
Eve Florianne Fabre, Vsevolod Peysakhovich and Mickaël Causse

Part II Cognitive Computing

Application of a Simulation-Based Software Tool for the Prospective Design of IT Work Places 87
Nico Feller, Andreas Amann, Ulf Müller, Michael Schiffmann, Oliver Kurscheid and Markus Gorzellik

Evaluation of User's Affective Engagement While Interacting with Educational Technologies: A Pilot Study 97
Marvin Andujar, Patricia Morreale, Yerika Jimenez, Luis Jimenez and Juan E. Gilbert

On the Modelling an Artificial Cognitive System Based on the Human-Brain Architecture 107
Olga Chernavskaya, Dmitry Chernavskii and Yaroslav Rozhylo

Identity Verification Using a Kinematic Memory Detection Technique 123
Merylin Monaro, Luciano Gamberini and Giuseppe Sartori

Emotiv-Based Low-Cost Brain Computer Interfaces: A Survey 133
Naveen Masood and Humera Farooq

Part III Physiological Monitoring and Interaction

Test-Retest Stability of EEG and Eye Tracking Metrics as Indicators of Variations in User State—An Analysis at a Group and an Individual Level. 145
Jessica Schwarz and Sven Fuchs

Transcranial Direct Current Stimulation (tDCS) Versus Caffeine to Sustain Wakefulness at Night When Dosing at Start-of-Shift. 157
Lindsey McIntire, R. Andy McKinley, Justin Nelson and Chuck Goodyear

The Effects of Transcranial Direct Current Stimulation (tDCS) on Training During a Complex Procedural Task 173
R. Andy McKinley, Lindsey McIntire, Jeremy Nelson, Justin Nelson and Charles Goodyear

Eliciting Sustained Mental Effort Using the Toulouse N-Back Task: Prefrontal Cortex and Pupillary Responses 185
Mickaël Causse, Vsevolod Peysakhovich and Kevin Mandrick

An Eye Movement Research on 2D Interactive Game Design 195
Na Lin, Lei Feng, Tian Kang, Shiyu Fan and Yunhong Zhang

Part IV Theoretical Advances in Cognitive Engineering and Neuroergonomics

Cognitive and Emotional-Motivational Implications in the Job Design of Digitized Production Drilling in Deep Mines 211
Mohammed-Aminu Sanda

Measuring Operator Emotion Objectively at a Complex Final Assembly Station 223
Sandra Mattsson, Dan Li, Åsa Fast-Berglund and Liang Gong

A Comparative Study on 3D/2D Visual Search Performance on Different Visual Display Terminal 233
Yunhong Zhang, Ruifeng Yu, Lei Feng and Xin Wu

Aging Effects on Brain Efficiency in General Aviation Pilots 243
Zarrin Chua and Mickaël Causse

Simulating Team Cognition in Complex Systems: Practical Considerations for Researchers 255
Michael McNeese, Nathan J. McNeese, Tristan Endsley, James Reep and Peter Forster

Fatigue and Theories of Resources: The Energetic Metaphor of the Mind Functioning Might Be Reconsidered 269
José J. Cañas

A New Method for the Objective Registration of Mental Workload 279
Thea Radüntz

Part V Assessing Cognition and Performance

Effect of Time Pressure on Work Efficiency and Cognitive Judgment 293
Junpei Yabuki and Hiroshi Hagiwara

Cognitive Architecture Based Platform on Human Performance Evaluation for Space Manual Control Task 303
Yanfei Liu, Zhiqiang Tian, Yuzhou Liu, Junsong Li and Feng Fu

Multi-level Cognitive Cybernetics in Human Factors 315
Daniel N. Cassenti, Katherine R. Gamble and Jonathan Z. Bakdash

Accessibility Evaluation: Manual Development and Tool Selection for Evaluating Accessibility of E-Textbooks 327
Yu Ting Sun, Aaron K.M. Manabat, Mei Ling Chan, Isis Chong and Kim-Phuong L. Vu

Development of a Scoring System for Evaluating the Accessibility of eTextbooks 339
Mei Ling Chan, Yu Ting Sun, Andriana M. Tesoro and Kim-Phuong L. Vu

Measuring the After-Effects of Disruption on Task Performance 351
Robert G. Abbott, Eric Moyer and Chris Forsythe

Complexity and Reliability as Basic Quantitative Characteristics of Computer Based Tasks 361
Inna Bedny, Gregory Bedny and Waldemar Karwowski

Windshield Frame Shape and Awareness of the External World While Driving an Automobile 375
Yusaku Takeda, Koji Iwase, Toshihiro Hara, Atsuhide Kishi, Kazuo Nishikawa, Richard Veale, Masatoshi Yoshida, Tadashi Isa and Takahide Nouzawa

Managing Mediated Interruptions in Manufacturing: Selected Strategies Used for Coping with Cognitive Load 389
Jessica Lindblom and Jonna Gündert

The Design Features of Pictorial Face Design for Facilitating Exertion Perception 405
Ding Hau Huang, Wen Ko Chiou, Bi Hui Chen and Yi-Lang Chen

Design an Interactive Game App of Horticultural Therapy for Older Adults 413
Pin-Yi Lai and Chien-Hsu Chen

Visual Psychophysical Thresholds for Perceiving Objects and Letters on Monocular Head-Up Displays in Indoor and Outdoor Lighting Conditions 421
Breanne K. Hawes, Tad T. Brunyé and Brian P. Westgate

Effects of Biasing Information on the Conceptual Structure of Team Communications 433
Michael T. Tolston, Victor Finomore, Gregory J. Funke, Vincent Mancuso, Rebecca Brown, Lauren Menke and Michael A. Riley

Part I
Influencing Design with Cognitive Engineering and Neuroergonomics

Cued Recall with Gaze Guiding—Reduction of Human Errors with a Gaze-Guiding Tool

Barbara Frank and Annette Kluge

Abstract Gaze guiding has been found to be effective for skill acquisition. In three studies, 120 participants learned how to operate a simulated process control task and acquired complex cognitive skills. The studies differed only in the simulated process control tasks: Study 1 consisted of a fixed-sequence task, Study 2 of a contingent-sequence task and Study 3 of a parallel-sequence task. After two weeks, the acquired skill had to be recalled. The Gaze-Guiding group received the help of a gaze-guiding tool in week 3, while the Control group received no help. The results of all studies imply that the gaze-guiding tool supported the correct execution of the tasks. In Study 3, the gaze-guiding tool also supported a higher production outcome compared to the Control group. Gaze guiding can be used as a cued recall tool for skills which require the exact execution of a procedure for different task types.

Keywords Gaze guiding · Dual task · Fixed task · Decision making · Complex cognitive skill · Standard operating procedures

1 Introduction

Printing a document from a computer is easy: Press print and the document is printed. But when a system failure arises, printing gets annoying, e.g. when the document is stuck in a "queue". Often, the computer user does not know what to do next, even if the specific problem has occurred before. First, the user has to know why the problem occurred. Second, the user has to decide whether to (a) fix the problem by him/herself—"pause", "cancel", end queue service in system settings, restart or try all of these together or (b) search for instructions on the internet and

B. Frank (✉) · A. Kluge
Department of Work, Organizational and Business Psychology, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany
e-mail: barbara.frank@rub.de

A. Kluge
e-mail: annette.kluge@rub.de

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_1

follow them. In such a case, the system failure would be easier to fix if the system detected the reason for the fault and provided support adapted to the user's skills. In other words, when the user has already pressed "cancel" (which is part of the fixing procedure) but does not know what to do next and stops interacting, the system would tell the user what to do next.

Such a scenario and skill loss can also occur in any situation in which a skill is rarely used. In industries like process control and air transportation it is difficult and yet important to maintain rarely used skills over a period of non-use. These rare situations in which initially learned skills are necessary are called non-routine situations [1]. Examples are the start-up of a plant, solving issues, taking corrective actions after incidents, or taking over when automation breaks down. In such cases, operators follow 'standard operating procedures' (SOP): They have to know which procedure is necessary, what the content of the procedure is, and how the procedure is executed [1]. A single handling error can provoke hugely harmful consequences, which can be prevented by extensive initial skill training and recurrent or refresher training after a defined time period [2, 3]. As such tasks are performed by human-computer interaction, the recall process of once-learned skills can be supported by a gaze-guiding tool that is implemented on the computer interface. A gaze-guiding interface fades in information into the control interface based on the current situation of the system and the intended procedure [4]. The present paper analyzes the potential of gaze guiding for cued skill recall in three complex task types: a fixed-sequence task, a contingent-sequence task and a parallel-sequence task.

1.1 Complex Cognitive Skills, Complex Tasks and Retention

An operator learns complex cognitive skills which consist of a combination of sub-skills. Generally speaking, a skill is a well-organized performance that requires coordinated processes of perception, cognition and action, is acquired through practice, and is performed with economy of effort [5, 6]. Complex cognitive skills are skills that are necessary e.g. in computer programming and dynamic process control tasks [7]. They can be divided into motor sub-skills and cognitive sub-skills [7]. Skills that have once been acquired can decay after periods of non-use [2] due to a decrease in retrieval strength (New Theory of Disuse [8]). Following this approach, skill retention requires the maintenance of high retrieval strength of learned skills.

Complex tasks consist of multiple elements that need to be executed [9]. According to Kluge [1], a complex task "can be decomposed into part-tasks that include sequences of steps, which need to be integrated and coordinated based on attentional processes and need to be orchestrated based on the simultaneous processing of knowledge elements into a interdependent team to meet the organizational goals" (p. 35). For the execution of complex tasks, complex cognitive skills are required, which are learned based on standard operating procedures (SOP) [10].

Such complex tasks in non-routine situations can consist of fixed-sequence tasks, contingent-sequence tasks and parallel-sequence tasks.

- In fixed-sequence tasks, the operator first needs to ascertain what kind of task has to be executed (e.g. start-up of a plant or error management) and then needs to execute the initially learned standard operating procedures sequentially (SOPs) [1].
- Dynamic decision making can be defined by multiple, interdependent and real-time decisions, occurring in an environment that changes independently and as a function of a sequence of actions [11]. In such an environment, decisions under certainty take place: The operator is aware of possible alternatives, consequences and the order of preferences [12]. A contingent-sequence task under certainty can consist of a fixed-sequence task in which at a special point or under a special condition, the operator has to perform the next steps based on a correct gathering of information and interpretation of the situation.
- Parallel-sequence tasks basically consist of two sequences which have to be synchronized in time [6, 13, 14]. In these tasks e.g. the operator has to control a second task while executing a first task, and both tasks are executed based on SOPs. A conscious, directed attention allocation and time-sharing is necessary to perform the task [13, 15]. An example of such a task is when a pilot is controlling different instruments during take-off, and consequently has to divide his/her attention according to change frequency and how valuable and costly the attention is [16].

1.2 *Gaze Guiding*

Cued recall is the recall of elements from memory triggered by cues [17]. An example of cued recall is remembering words with the support of word categories [18]. Possible methods for applying cued recall to complex cognitive skills are job aids. Job aids are used to process, store and extend information [19, 20]. Skills that are used infrequently, consist of sequences, and contain a large amount of information can be supported by job aids [21]. A procedural job aid is designed to guide users through the procedure step by step [20, 22]. A dynamic computer-based job aid such as the present gaze-guiding tool supports the operator if s/he fails to remember the next step, and interrupts the interaction with the system. Dynamic computer-based job aids have been used for the learning of skills and can be applied through various methods, such as the “attention guidance technique” and “visual cueing”. These methods assist learners in complex learning environments in which it may be challenging to detect relevant information and cues [23]. The attention guidance technique has been particularly applied in skill acquisition [24]. Attention guidance facilitates the search for relevant information and problem solving with salient, critical information [25]. A comparable method is “visual cueing”, which has also been used for learning [26]. Visual cueing highlights e.g. tasks, graphics or animations with arrows, circles or distinctive colors [26]. Some studies have shown that visual cueing guides attention and supports fast skill acquisition of information, retention and problem solving [26–28]. However, various studies found that simply highlighting relevant areas was not sufficiently effective for improving performance and reducing errors [24, 25, 29]. As a conclusion, it seems reasonable to provide further textual information, in addition to highlighting relevant areas, and to ensure a resource-friendly processing and understanding [24]. This can be offered by a gaze-guiding tool, which is able to guide the operator's gaze and consequently attention to the relevant areas and provide further information [30]. The optimal gaze-guiding tool balances the operator's mental workload and does not overextend her/him with too much information [31]. Recent studies have shown that gaze-guiding methods have been used successfully as learning support [32]. For instance, the performance of surgery students was found to be positively supported by a form of gaze guiding showing the next relevant steps as compared to exploration [32].

Based on the assumption that cued recall embodied by gaze guiding harbors the potential to support the operator's skill recall after a period of non-use, the following hypothesis is developed (and analyzed for each study):

Hypothesis: After a period of non-use and in a retention assessment, a group which is supported by gaze guiding shows a superior performance compared to a group without gaze guiding and free recall.

2 Method

2.1 Sample

Studies 1–3 were conducted from October 2014 to December 2015 with a sample of 100 participants for each study. In the following, only a subsample is analyzed to investigate gaze guiding in particular. The three studies differed only with respect to the learned task type (fixed-sequence task, contingent-sequence task, parallel-sequence task).

Study 1. From October 2014 to December 2015, 40 engineering students (15 female) participated in the study. Four participants were excluded based on the selection criteria (see below). The participants were recruited by postings on social networking sites and flyers handed out to engineering students. To ensure technical understanding which was required for the technical task, only students from faculties of engineering were recruited. Participants received 25€ for taking part. The study was approved by the local ethics committee. Participants were informed about the purpose of the study and told that they could discontinue participation at any time (in terms of informed consent). All participants were novices in learning the process control task used in the study. The recruitment was similar for all three studies.

Study 2. From April to July 2015, 40 students took part in the study (12 female). Five participants were excluded based on the selection criteria (see below).

Study 3. From October to December 2015, 40 students took part in the study (13 female).

2.2 Process Control Tasks

The complex cognitive skill in the present studies was performed in a simulated process control task: The participants had to know the content of the particular start-up procedure (SOP) and how to interact with the interface. The participants learned how to operate the microworld Waste Water Treatment Simulation (WaTrSim). The operation of WaTrSim includes the start-up of the plant, which is assumed to be a non-routine task that requires skill retention [10]. In WaTrSim, the operator's task is to separate waste water into fresh water and gas by starting up, controlling and monitoring the plant. The operation goal is to maximize the amount of purified water and gas and to minimize the amount of waste water (only separated purified gas outcome was used for the calculations because this shows that the participants performed the last steps of the start-up procedure; water had already been produced after the operation of column K1). This goal is achieved by considering the timing of actions and following the start-up procedure. The start-up procedure differed in all three studies.

Study 1. The operation included the start-up procedure of the plant as a *fixed-sequence task* comprising 13 steps (Table 1). Performing the WaTrSim start-up procedure correctly and in a timely manner led to a production outcome of a minimum of 200 l of purified gas. The minimum amount of purified gas in the initial training was used as selection criterion (>=200 l). Start-up time was max. 180 s.

Study 2. The operation included the start-up procedure of the plant as a *contingent-sequence task* comprising 13 steps and following five steps for each condition. The following five steps had to be executed depending on the conditions: heating W1 > 15 °C or heating W2 < 70 °C. After one of the conditions had occurred, the correct four steps had to be executed (Table 1). Performing the WaTrSim start-up procedure correctly and in a timely manner led to a production outcome of a minimum of 100 l of purified gas. The minimum amount of purified gas in initial training was used as selection criterion (>=100 l). Start-up time was max. 240 s.

Study 3. The operation included the start-up procedure of the plant as a parallel-sequence task. Two sequences had to be operated in parallel: 13 steps for sequence 1 and three steps for sequence 2. Sequence 2 had to be executed when the level of tank Bf had reached >75 % or <25 %. After one of the conditions had occurred, the correct two steps had to be executed (Table 1). Performing the WaTrSim start-up procedure (both sequences in parallel) correctly and in a timely manner led to a production outcome of a minimum of 200 l of purified gas.

Table 1 Description of fixed-sequence task (Study 1), contingent-sequence task (Study 2) and parallel-sequence task (Study 3)

Step	Study 1	Study 2		Study 3		
	Fixed-sequence task Start-up procedure: 13 steps	Contingent-sequence task Start-up procedure: 13 steps and 2 × 5 steps		Parallel-sequence task Start-up procedure: Sequence 1 (13 steps) and Sequence 2 (2 × 3 steps); both sequences had to be executed in parallel		
1	LIC V9: flow rate 500 l/h	LIC V9: flow rate 500 l/h		LIC V9: flow rate 500 l/h	A	Monitor tank level of tank Bf constantly
2	V2 deactivate follower control	V2 deactivate follower control		V2 deactivate follower control		
3	Valve V1: flow rate 500 l/h	Valve V1: flow rate 500 l/h		Valve V1: flow rate 500 l/h		**Condition 1**:
4	Wait until R1 > 200 l	Wait until R1 > 200 l		Wait until R1 > 200 l	B	Tank level Bf > 75 %
5	Valve V2: flow rate 500 l/h	Valve V2: flow rate 500 l/h		Valve V2: flow rate 500 l/h	C	FIC V8: flow rate 90 %
6	Wait until R1 > 400 l	Wait until R1 > 400 l		Wait until R1 > 400 l	D	Heating W2: 70 °C
7	Valve V3: flow rate 1000 l/h	Valve V3: flow rate 1000 l/h		Valve V3: flow rate 1000 l/h		
8	Wait until HB1 > 100 l	Wait until HB1 > 100 l		Wait until HB1 > 100 l		**Condition 2**:
9	Activate heating HB1	Activate heating HB1		Activate heating HB1	B	Tank level Bf > 25 %
10	Wait until HB1 > 60 °C	Wait until HB1 > 60 °C		Wait until HB1 > 60 °C	C	FIC V8: Flow rate 10 %
11	Activate column K1	Activate column K1		Activate column K1	D	Heating W2: 20 °C
12	Valve V4: flow rate 1000 l/h	Valve V4: flow rate 1000 l/h		Valve V4: flow rate 1000 l/h		
13	Valve V6: flow rate 400 l/h	Valve V6: flow rate 400 l/h		Valve V6: flow rate 400 l/h		
14		W1 > 15°C **OR**	W2 > 70°C			
15		LIC V8 deactivate	LIC V8 deactivate			
16		LIC V9 700 l/h	LIC V9 600 l/h			
17		LIC V8 500 l/h	LIC V8 400 l/h			

(continued)

Table 1 (continued)

Step	Study 1	Study 2		Study 3		
	Fixed-sequence task Start-up procedure: 13 steps	Contingent-sequence task Start-up procedure: 13 steps and 2 × 5 steps		Parallel-sequence task Start-up procedure: Sequence 1 (13 steps) and Sequence 2 (2 × 3 steps); both sequences had to be executed in parallel		
18		Heating W1 15 °C	Heating W2 70 °C			
	Min. prod. outcome IT: 200 l	Min. prod. outcome IT: 100 l		Min. prod. outcome IT: 200 l		
	Max. start-up time: 180 s	Max. start-up time: 240 s		Max. start-up time: 240 s		

IT initial training

The minimum amount of purified gas in initial training was used as selection criterion (>=200 l). Start-up time was max. 240 s.

2.3 WaTrSim Gaze Guiding

The gaze-guiding tool used in the present study (Fig. 1) has been described, pre-tested [33] and evaluated as effective in a process control task [4]. The gaze-guiding tool is technically implemented with a transparent, darkened interface that leaves space for the relevant cue element, a red-orange flashing frame of the cue element, and additional textual information of the following step with a pictogram (Fig. 1). The gaze-guiding tool is faded in at a predefined time point and after the system has reached the relevant conditions for the next steps in order to ensure that the operator has time to recall and react independently without the gaze-guiding tool. The gaze-guiding tool is only faded in if the operator has not executed the relevant next step in a timely manner. After the gaze-guiding tool has been faded in, it disappears again when the operator interacts with WaTrSim. It is faded in repeatedly until the operator executes the correct adjustments of the relevant step [4, 33].

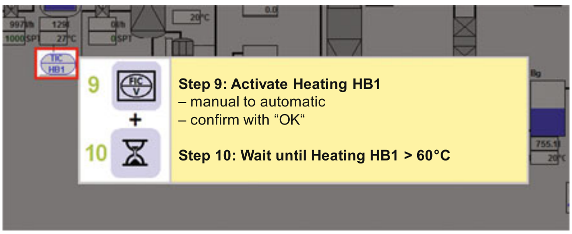


Fig. 1 Example of the gaze-guiding tool. The interface is *darkened*, the valve is highlighted with a flashing *red-orange* frame and a *yellow* text box is displayed

2.4 Research Design

Studies 1–3. The studies consisted of a 2 × 2 mixed experimental design (between- and within-subjects design). One experimental group (Gaze-Guiding group) and one Control group were analyzed at two measurement time points (week 1: initial training and week 3: retention assessment).

2.5 Variables

Independent Variables. The *gaze guiding* is described in 2.3 (adapted to the sequence that needs to be executed in Studies 1–3).

Dependent Variables. The *production outcome* was measured by the produced amount of purified gas. For the calculations, the best trial of the initial training (week 1) and the first trial of the retention assessment (week 3) were used. The minimum production outcome at the initial training depended on the task. Study 1: 200 l, Study 2: 100 l, Study 3: 200 l.

The *start-up mistakes* are the sum of incorrect valve adjustments and procedure mistakes, such as adjustment of the incorrect valve flow rate, and depended on the task. The start-up mistakes of the best production outcome trial of the initial training and of the first trial of the retention assessment were used for the calculations. Study 1: 0–15 mistakes, Study 2: 0–19 mistakes, Study 3: 0–21 mistakes (recalculated in percentages, 0–1).

Control Variables. *Start-up time:* The start-up time was measured for the best production outcome trial of the initial training (week 1) and the first trial of the retention assessment (week 3). The start-up time was limited depending on the task. Study 1: 0–180 s. Study 2: 0–240 s. Study 3: 0–240 s.

Retentivity was measured with the Wilde Intelligence Test-2, which consists of verbal, numerical and figural information [34]. First, the participants had to memorize the verbal, numerical and figural information for four minutes. After a disruption phase of 17 min, they then answered reproduction questions related to the memorized information, choosing one of six response options (scores from 0 to 21; identical for Studies 1–3).

2.6 Procedure

The procedure was similar for Studies 1–3. The participants attended twice: initial training and (after two weeks) retention assessment.

The initial training (week 1) lasted for 120 min. After completing a test concerning retentivity, participants explored the simulation twice. They were then given information and instructions about the start-up procedure. After this, they trained the start-up procedure (fixed-sequence task, contingent-sequence task or parallel-sequence task) with a manual. Following the training, the participants had to perform the start-up procedure four times without help. They were required to produce a minimum of purified gas (Study 1: 200 l, Study 2: 100 l, Study 3: 200 l).

Two weeks after the skill acquisition, the retention assessment (week 3) took place, which lasted for approximately 30 min. The participants were asked to start up the plant up to five times without help (the first trial was used to assess skill retention). The Gaze-Guiding group received the support of the gaze-guiding tool if necessary.

3 Results

The descriptive statistics are given in Table 2. In Study 1, four participants and in Study 2, five participants were excluded due to a too low production outcome at initial training. The groups did not differ in terms of control variables ($p > 0.05$).

3.1 Hypothesis-Testing

To test the hypothesis, repeated measures ANOVAs with the two measurement time points initial training and retention assessment were conducted to measure skill retention.

Study 1: Fixed-Sequence Task (N = 36; Fig. 2**)**. The analysis showed no significant difference between the Gaze-Guiding and Control group in terms of production outcome (interaction of time*group: $F(1, 34) = 1.97$, $p = 0.169$, $n_p^2 = 0.06$). The analysis of start-up mistakes showed that the Gaze-Guiding group made significantly fewer start-up mistakes than the Control group (interaction of time*group: $F(1, 34) = 27.28$, $p < 0.001$, $n_p^2 = 0.05$).

Study 2: Contingent-Sequence Task (N = 35; Fig. 2**)**. The analysis showed no significant difference between the Gaze-Guiding and Control group in terms of production outcome (interaction of time*group: $F(1, 33) = 0.82$, $p = 0.372$, $n_p^2 = 0.02$). The analysis of start-up mistakes showed that the Gaze-Guiding group made significantly fewer start-up mistakes than the Control group (interaction of time*group: $F(1, 32) = 17.60$, $p < 0.001$, $n_p^2 = 0.04$).

Study 3: Parallel-Sequence Task (N = 40; Fig. 2**)**. The analysis showed that the Gaze-Guiding group produced significantly more purified gas than the Control group (interaction of time*group: $F(1, 38) = 7.98$, $p = 0.008$, $n_p^2 = 0.17$). The analysis of the start-up mistakes showed that the Gaze-Guiding group made significantly fewer mistakes compared to the Control group (interaction of time*group: $F(1, 38) = 4.96$, $p = 0.032$, $n_p^2 = 0.12$).

Table 2 Descriptive statistics for Study 1, Study 2 and Study 3

	Study 1		Study 2		Study 3	
	GG (n = 18)	CG (n = 18)	GG (n = 16)	CG (n = 19)	GG (n = 20)	CG (n = 20)
Independent and control variables						
Sex	6 female 12 male	9 female 9 female	6 female 10 male	6 female 13 male	7 female 13 male	6 female 14 male
Age	21.61 (2.03, 18–25)	22.33 (2.79, 18–28)	22.88 (2.31, 20–29)	22.63 (3.24, 19–30)	23.10 (3.89, 18–31)	22.20 (3.24, 18–29)
Retentivity	14.44 (3.20, 8–19)	15.33 (1.97, 12–20)	12.50 (3.58, 7–19)	12.74 (3.12, 6–18)	14.72 (3.72, 7–19)	16.36 (2.56, 12–20)
Start-up time (IT)	89.67 (7.97, 70–103)	96.06 (8.48, 80–113)	156.38 (37.17, 108–231)	130.32 (35.90, 87–218)	121.90 (28.06, 82–191)	118.85 (25.12, 93–172)
Start-up time (RA)	143.78 (29.88, 62–180)	151.72 (31.54, 93–180)	234.88 (9.24, 216–240)	225.79 (33.96, 139–240)	210.45 (41.17, 114–240)	223.95 (28.48, 136–240)
Dependent variables—IT (week 1)						
Prod. outcome	348.49 (47.60, 259.80–439.93)	337.46 (33.23, 260.00–399.30)	336.73 (132.23, 151.20–512.00)	396.45 (151.38, 112.90–607.94)	463.74 (108.42, 202.00–627.01)	489.84 (77.72, 320.36–581.01)
Start-up mistakes	0.09 (0.13, 0–0.4)	0.11 (0.15, 0–0.47)	0.19 (0.15, 0–0.48)	0.25 (0.29, 0–0.85)	0.26 (0.21, 0–0.69)	0.27 (0.19, 0–0.74)
Dependent variables—RA (week 3)						
Prod. outcome	146.20 (119.31, 0–471.93)	82.80 (112.08, 0–299.93)	5.25 (11.10, 0–40.00)	18.73 (63.68, 0–273.71)	124.80 (169.62, 0–499.01)	43.93 (81.61, 0–253.36)
Start-up mistakes	0.14 (0.16, 0–0.47)	0.53 (0.21, 0–0.87)	0.23 (0.19, 0–0.53)	0.67 (0.25, 0.16–1)	0.42 (0.28, 0–1)	0.64 (0.23, 0.06–0.86)

Note M(*SD*, Range); *GG* Gaze-Guiding group, *CG* Control group; Start-up time Study 1: 180 s, Study 2: 240 s, Study 3: 240 s; *IT* initial training, *RA* retention assessment; Minimum production outcome at initial training of Study 1: 200 l, Study 2: 100 l, Study 3: 200 l; Start-up mistakes are given in percentages: 0–1

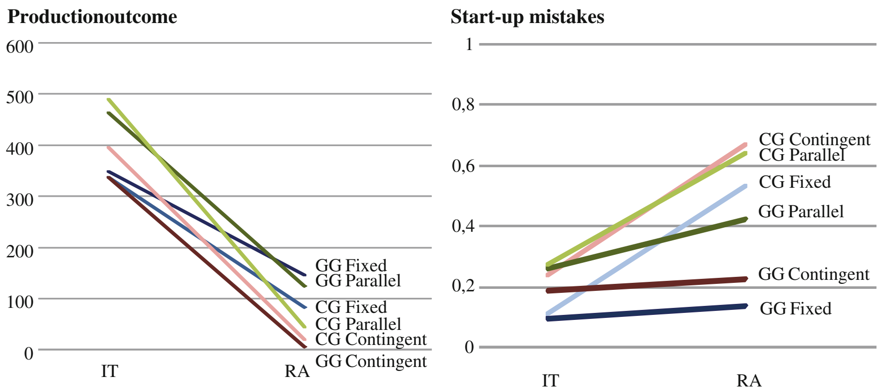


Fig. 2 Production outcome and start-up mistakes for gaze-guiding group (*GG*) and control group (*CG*) of study 1 (*fixed*), study 2 (*parallel*) and study 3 (*contingent*)

Studies 1–3 (N = 110). Only start-up mistakes were analyzed for Studies 1–3 as a whole due to different production outcome criteria (and start-up times) in the three studies. The analysis showed that all three Gaze-Guiding groups (treated as one group) made significantly fewer mistakes than the Control groups (treated as one group): $F(1, 108) = 29.77$, $p < 0.001$, $n_p^2 = 0.22$ (interaction of time*group).

4 Discussion

The results show that the gaze guiding tool supported the cued skill recall after a period of non-use for the correct execution of all three tasks. Using the gaze guiding tool results in comparable start-up mistakes to the initial training and in fewer start-up mistakes compared to the Control group in all tasks. This indicates that the gaze-guiding tool has the potential to support the execution of the tested task types.

Nevertheless, the effectiveness differs between the tasks that are supported. The results indicate that gaze guiding supported a higher production of purified gas for the parallel-sequence task compared to the Control group and when comparing two measurement time points, but this was not shown for the fixed-sequence and the contingent-sequence task. Especially in the contingent-sequence task, the gaze-guiding tool was not found to be effective (production outcome $M = 5.25$). It can be assumed that gaze guiding decreases in efficacy with increasing number of skill elements in a task. In this respect, the affordances of the contingent sequence were higher than in the other two sequences: The contingent sequence consisted of 23 steps (both conditions; fixed sequence: 13 steps) and the participants had to decide which procedure to follow depending on predefined cues that needed to be acknowledged. Accordingly, an incorrect situation assessment could result in the execution of the wrong follow-up sequence, thus leading to zero production outcome. In sum, it can be concluded that the gaze-guiding format chosen in the presented studies is not a "one-size-fits-all"-sequences solution. Instead, further research is needed to determine the benefits and limitations of different gaze-guiding design principles for different sequences.

The results also suggest that the gaze-guiding tool cannot support the performance with regard to production outcome, which depends on the speed of sequence execution: In all groups, the skill level decreased from initial training to retention assessment, and for the contingent-sequence task in particular, no benefit of the gaze guiding could be shown. The descriptive statistics of the control variable start-up time suggests that the groups using gaze guiding started up WaTrSim more slowly at the retention assessment than at the initial training. This may be a reason why the production outcome was lower at the retention assessment even if the start-up procedure was executed correctly. It can therefore be suggested that gaze guiding is a helpful tool for the correct execution but that there is also some potential for development in terms of optimizing the timing of the gaze guiding. For example, the timing could be adjusted so that gaze guiding is faded in earlier, enabling the participants to act earlier.

In cases of emergencies, incidents or breakdowns, the correct execution is important for controlling the plant safely. The present results represent a promising start point for developing a gaze-guiding tool that is able to support operators in executing SOPs after longer periods of non-use.

Acknowledgments The research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) with grant number KL2207/3-3. Part of the results are also published in Frank and Kluge [35, 36].

References

1. Kluge, A.: The Acquisition of Knowledge and Skills for Taskwork and Teamwork to Control Complex Technical Systems. Springer, Heidelberg (2014)
2. Arthur, W., Bennett, W., Stanush, P.L., McNelly, T.L.: Factors that influence skill decay and retention: a quantitative review and analysis. Hum. Perform. **11**, 57–101 (1998)
3. Kluge, A., Frank, B.: Counteracting skill decay: four refresher interventions and their effect on skill and knowledge retention in a simulated process control task. Ergonomics **57**, 175–190 (2014)
4. Weyers, B., Frank, B., Bischof, K., Kluge, A.: Gaze guiding as support for the control of technical systems. Int. J. Inf. Syst. Crisis Response Manag. **7**, 59–80 (2015)
5. Patrick, J.: Training: Research and Practice. Academic Press, San Diego (1992)
6. Proctor, R.W., Dutta, A.: Skill Acquisition and Human Performance. Sage Publications, Thousand Oaks (1995)
7. Van Merriënboer, J.: Training Complex Cognitive Skills: A Four-Component Instructional Design Model for Technical Training. Educational Technology, Englewood Cliffs (1997)
8. Bjork, R.A., Bjork, E.L.: Optimizing treatment and instruction: implications of a new theory of disuse. In: Nilsson, L.G., Ohta, N. (eds.) Memory and Society Psychological Perspectives, pp. 109–134. Psychology Press, Hove (2006)

9. Proctor, R.W., Vu, K.-P.: Laboratory studies of training, skill acquisition, and retention of performance. In: Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (eds.) The Cambridge Handbook of Expertise and Expert Performance, pp. 265–286. Cambridge University Press, Cambridge (2006)
10. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance. Prentice Hall, Upper Saddle River (2000)
11. Brehmer, B.: Dynamic decision making: human control of complex systems. Acta Psychol. **81**, 211–241 (1992)
12. Dörner, D., Bick, T.: Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität. Huber (1994)
13. Wickens, C.D., McCarley, J.S.: Applied Attention Theory. CRC Press, Boca Raton (2008)
14. Wickens, C.D.: Multiple resources and mental workload. Hum. Factors **50**, 449–455 (2008)
15. Schumacher, E.H., Seymour, T.L., Glass, J.M., Fencsik, D.E., Lauber, E.J., Kieras, D.E., Meyer, D.E.: Virtually perfect time sharing in dual-task performance: uncorking the central cognitive bottleneck. Psychol. Sci. **12**, 101–108 (2001)
16. Moray, N.: Monitoring behavior and supervisory control. In: Boff, J., Kaufmann, K., Thomas, J.P. (eds.) Handbook of Perception and Human Performance, pp. 40–51. Wiley, New York (1986)
17. Moult, M.: Cued Recall, Encyclopedia of Clinical Neuropsychology, pp. 751–752. Springer, Heidelberg (2011)
18. Carrier, M., Pashler, H.: The influence of retrieval on retention. Mem. Cognit. **20**, 633–642 (1992)
19. Kluge, A., Grauel, B., Burkolter, D.: Combining principles of cognitive load theory and diagnostic error analysis for designing job aids: effects on motivation and diagnostic performance in a process control task. Appl. Ergon. **44**, 285–296 (2013)
20. Swezey, R.W.: Design of job aids and procedure writing. In: Salvendy, G. (ed.) Handbook of Human Factors, pp. 1039–1057. Wiley, Hoboken (1987)
21. Rossett, A., Gautier-Downes, J.: A Handbook of Job Aids. Pfeiffer & Co, San Francisco (1991)
22. Salas, E., Wilson, K.A., Priest, H.A., Guthrie, J.W.: Design, Delivery, and Evaluation of Training Systems. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, pp. 472–512. Wiley, Hoboken (2006)
23. Mayer, R.E., Moreno, R.: A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. J. Educ. Psychol. **90**, 312–320 (1998)
24. de Koning, B.B., Tabbers, H.K., Rikers, R.M.J.P., Paas, F.: Attention guidance in learning from a complex animation: seeing is understanding? Learn. Instr. **20**, 111–122 (2010)
25. de Koning, B.B., Tabbers, H.K., Rikers, R.M., Paas, F.: Attention cueing as a means to enhance learning from an animation. Appl. Cogn. Psychol. **21**, 731–746 (2007)
26. Lin, L., Atkinson, R.K.: Using animations and visual cueing to support learning of scientific concepts and processes. Comput. Educ. **56**, 650–658 (2011)
27. Amadieu, F., Mariné, C., Laimay, C.: The attention-guiding effect and cognitive load in the comprehension of animations. Comput. Hum. Behav. **27**, 36–40 (2011)
28. Jamet, E.: An eye-tracking study of cueing effects in multimedia learning. Comput. Hum. Behav. **32**, 47–53 (2014)
29. Grant, E.R., Spivey, M.J.: Eye movements and problem solving guiding attention guides thought. Psychol. Sci. **14**, 462–466 (2013)
30. Kluge, A., Greve, J., Borisov, N., Weyers, B.: Exploring the usefulness of two b variants of gaze-guiding-based dynamic job aids for performing a fixed-sequence start-up procedure after longer periods of non-use. Int. J. Hum. Factors Ergon. **3**, 148–169 (2014)
31. Vidulich, M.A., Tsang, P.S.: The confluence of situation awareness and mental workload for adaptable human-machine systems. J. Cogn. Eng. Decis. Mak. **9**, 95–97 (2015)
32. Vine, S.J., Masters, R.S., McGrath, J.S., Bright, E., Wilson, M.R.: Cheating experience: guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. Surgery **152**, 32–40 (2012)

33. Bischof, K., Weyers, B., Frank, B., Kluge, A.: Gaze Guiding zur Unterstützung der Bedienung Technischer Systeme. In: Weisbecker, A., Burmester, M., Schmidt, A. (eds.) Mensch und Computer 2015—Workshopband. De Gruyter Oldenbourg, Berlin, pp. 61–68 (2015)
34. Kersting, M., Althoff, K., Jäger, A.O.: Wilde Intelligence Test-2 (WIT-2). Hogrefe, Göttingen (2008)
35. Frank, B., Kluge, A.: Cued skill recall: the effect of gaze guiding to support complex cognitive skill recall to reduce errors and retain performance, under review
36. Frank, B., Kluge, A.: Recall enhancement with gaze guiding: performance support and error reduction in dual tasks. under review

Investigation of the Recognition of Different Font Sizes on Human-Machine Interface and Physiological Characteristics for Aged People

Shengwen Luo, Li Ding, Linghua Ran and Yan Li

Abstract With the growing population of the elderly people, designers of the human-machine interface must consider their psychophysiological characteristics. Three groups of subjects with a mean age of 21.0, 38.5 and 61.5 years, participated in the experiments. First they were examined by the perimeter. Then their near vision were measured by the Landolt ring test. Finally a Chinese character choosing task was done. The senile group had a marked decrease of light sensitivity and also a decline of near vision. For the 9-point font, there was no significant difference among three groups. However, for the 6-point font, the senile group reacted slowly. In the future, for the sake of the aged people, it would be of great importance to work out the critical and optimal font sizes under different light conditions. To obtain this goal, more experiments should be done and more subjects should be involved.

Keywords Font sizes recognition · Physiological characteristics · Human-Machine interface design

S. Luo · L. Ding · Y. Li (✉)
School of Biological Science and Medical Engineering, Beihang University,
No. 37 Xueyuan Road, Haidian District, Beijing 100191, People's Republic of China
e-mail: yanli_buaa@buaa.edu.cn

S. Luo
e-mail: 13101045@buaa.edu.cn

L. Ding
e-mail: ding1971316@buaa.edu.cn

L. Ran
China National Institute of Standardization, No.4 Zhi Chun Road,
Haidian District, Beijing 100088, People's Republic of China
e-mail: ranlh@cnis.gov.cn

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_2

1 Introduction

The human-machine interface design is important in modern machine. The study of it is widely used in areas of cell phone, computer, tablet PC and other intelligent electronic products. Many factors in the human-machine interface design seriously affect the quality of user experiences, and one of the factors is font size.

Aging is the time-dependent accumulation of cellular insults or damage accompanied by subsequent functional decline that increases organisms' vulnerability to death [1]. Advanced age is widely recognized as one of the biggest risk factors for many of the leading causes of vision loss, such as Presbyopia, cataract, glaucoma, age-related macular degeneration (AMD) and diabetic retinopathy [1, 2]. One of reasons why the advanced age influence the vision is that the level of some endocrine substances produce changes in the eyes. According to recent studies, melanopsin, melatonin rhythms, and many other endocrine substances are responsible for the human vision [3, 4], and some of these substances would decrease or change when human is getting old. Not only the endocrine substances, but also some functional cells change in the development of human's life. And the human anatomy theory explains the sensitivities of human's eyes depend on the rod cells and cone cells in the eye. Rod cells produce rhodopsin to make human's retina sensitive to the weak light and therefore it matters to eyes function such as dark adaption. Moreover, many other factors can also influence the synthesis of rhodopsin, for example, the lack of vitamin A assimilated from outside. As people are getting old, the synthesis function of their body's cells, including rod cells and cone cells, will decay. This change is one of the reasons causing weakness of the older man's sensitivity to light.

Some researches prove that the aged people have different performances in the function of eyes. Raymond P. Najjar discusses age-related decrease in non-visual sensitivity to light. The age has a great relationship with human's non-visual sensitivity [5], which relates to one of the intrinsically-photosensitive retinal ganglion cells [6]. And the human's visual sensitivity also relates to another intrinsically-photosensitive retinal ganglion cells, namely roll cells and cone cells. Joanne M. Wood and his team [7] discover that different aged drivers also had different visual abilities of seeing pedestrians at night. It proves that some skills of human's eyes become weaker when human is getting old.

The physiological characteristics of human's eyes are meaningful in the human-machine interface design. Different font sizes influence the information transmission efficiency [8]. Andrew R. Whatham and his partners carry out a series of experiments in which subjects with normal vision are asked to read isolated lowercase single letters and lowercase words of 4, 7 and 10 letters, in separate tests. And the result reveals that critical character sizes have a functional relation with the subjects' acuity, which proves the font size is able to be the measurement of human's vision [9]. The surrounding luminance also makes the difference in the recognition of the word which prove that the light intensity is one of the important factors in the font recognition [10]. The human characteristics which relates to the light intensity is their sensitivity to light. Thus, we make a hypothesis that the sensitivity to light influence the human's performance on the human-machine interface; ultimately, vision and sensitivity to light may influence the design of font sizes on the human-machine interface.

Based on the abovementioned studies, the hypothesis of this study is that human's age, vision acuity and sensitivity to light affect the ability of recognition of different font sizes on human-machine interface. And its aim is to discover the relations among them, hoping to provide a reference for the human-machine interface design. In this study, three experiments are designed to measure the physiological characteristics of subjects' eyes. The first experiment measures the subjects' sensitivity to light using IVS-201B automatic perimeter. The second experiment applies an E-prime program to record the reaction time of recognizing the different font sized Chinese characters. And the third is the Landolt ring experiment to measure the subjects' near vision. The statistical analysis will be used to find out the relations.

2 Methods

2.1 Subjects

Eighteen subjects participated in the project (N = 18). They were divided into three groups by age, and each group consisted of three males and three females. The average age of each group was 21.0 ± 2.0, 38.5 ± 5.5 and 61.5 ± 5.5, respectively.

2.2 Experimental Procedures

Apparatus.

(1) IVS-201B Automatic perimeter: the subject was required to cover one eye with an eye patch and receive the check of the perimeter to measure Central 30° field of vision. The measured results were represented by numbers (refer to Fig. 1), and the larger the number is, the more sensitive to light.
(2) Lenovo X201 computer: for running E-prime programs.

E-prime Programs. Two E-prime programs were used in the experiment. The subjects finished the tests under the environment with indoors illumination rated at 300 lx.

One of the programs was aimed to identify the ability of recognizing the different font sizes on the human-computer interaction interface. The program asked the subjects to choose two Chinese characters on the same page. The displayed page

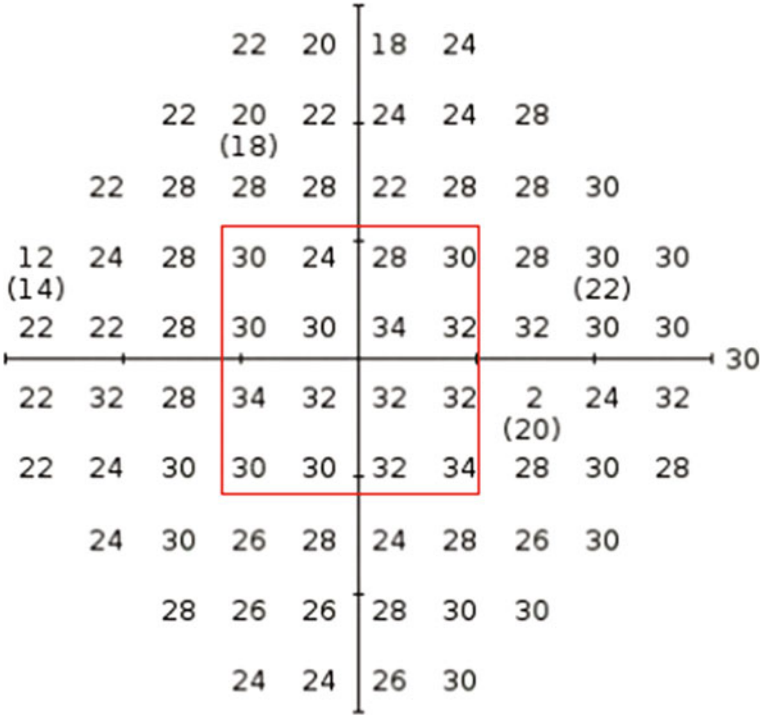


Fig. 1 Digitized sensitivity to light

had nine optional rectangles and each contained two characters inside except the middle one. The subjects were supposed to choose the middle character first and then identify the same character from the surrounding rectangles. Both the reaction time (RT) and accuracy were recorded by E-Prime (Fig. 2).

Another program was meant to check the subjects' near vision. It used Landolt ring to serve the purpose. Subjects were asked to sit about forty centimeters away from the laptop's screen and look at the screen horizontally. When the Landolt ring appear in the center, they were required to choose the right direction of the gap on the Landolt ring by pressing four buttons 'w, a, s, d', which represented four different directions. There were 29 levels of Landolt ring sizes with visual angles ranging from 1 to 15 arc minutes. E-prime recorded reaction time (RT) and accuracy of each trial correctly (Fig. 3).

2.3 Statistical Methods

(1) First, this study listed the averaged value and standard deviation of each group to give a clear description;

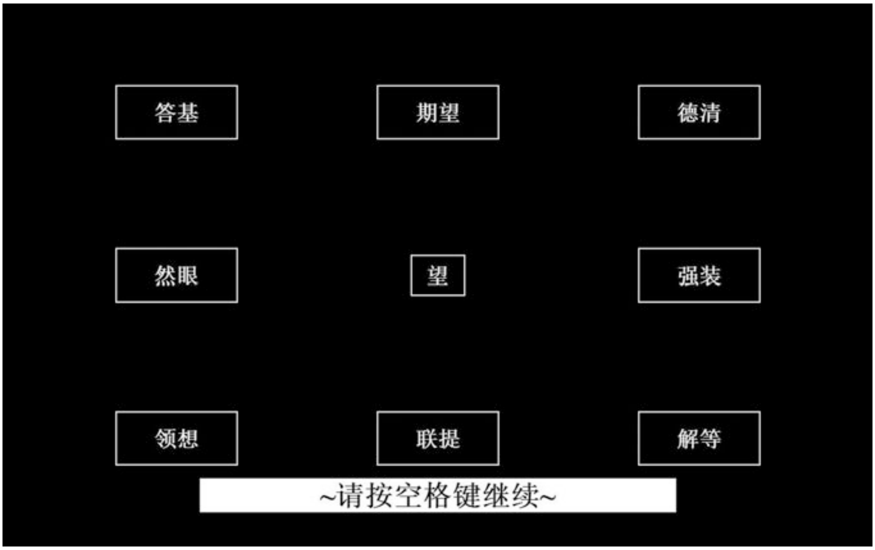


Fig. 2 The interactive interface that measured the reaction time

Fig. 3 The interface of the program that tested the vision

(2) Second, two-tailed paired t-test was used to get the *H*-value and *P*-value. T test is the most frequently used test to determine if two sets of data are significantly different from each other.

(3) Lastly, linear correlation analysis was used to judge the linear correlation among factors and get the correlation coefficient (R) and *P*-value.

3 Results

The 4 × 4 centering matrix (shown in Fig. 1) was chosen to represent each subject's sensitivity to light. The reason why these data has been chosen was that the subjects mostly had the best feeling to light in this area and normal people used this part of their eyes to watch things for most of the time. This area of human's eyes has a great influence on human's life and its value is meaningful in statistics.

The vision of each subject was represented by the lowest level of Landolt ring size that he was able to recognize. The ability of subjects to recognize the font size on computer was measured by the reaction time on recognizing 6 font size and 9 font size. In analyzing this outcome, the recognition accuracy was ignored because the two font sizes were large enough for subjects to do almost every trial correctly.

3.1 Averaged Value and Standard Deviation Analysis

According to the averaged analysis (see Table 1), the sensitivity to light and vision decreased along with the increase of age especially between the senile group and the middle aged group. It might be due to the degeneration of subject's eyes as they are getting old. Reaction time of 6-point character choosing task also increased along with the increase of age.

3.2 T-Test Result Analysis

T test result revealed that the sensitivity to light had significant difference between middle aged and senile group, the same can be said of the difference between youngest and senile group, especially in the left eyes' data. The other factors did not reveal the obvious regulation in the comparison (Table 2).

Table 1 Averaged values and standard deviation of each group

Group	Youngest	Middle aged	Senile
Age	21.67 ± 1.47(yrs)	39.00 ± 4.24(yrs)	60.83 ± 4.07(yrs)
Sensitivity to light (OS)	31.73 ± 0.72	31.75 ± 1.17	25.13 ± 5.06
Sensitivity to light (OD)	31.83 ± 1.55	31.92 ± 1.13	25.10 ± 4.67
vision	2.17 ± 2.40 (level)	3.17 ± 2.56 (level)	5.17 ± 2.71 (level)
RT(6)	3127.84 ± 961.97 (ms)	3486.27 ± 570.19 (ms)	3907.95 ± 1232.53 (ms)
RT(9)	2517.25 ± 564.19 (ms)	2477.64 ± 308.21 (ms)	2787.25 ± 532.72 (ms)

(*OS* means the left eyes, *OD* means the right eyes)

Table 2 *P*-values from t-test among groups

Group	Youngest and middle aged	Middle aged and senile	Youngest and senile
Sensitivity to light (OS)	0.323090	0.032024	0.026538
Sensitivity to light (OD)	0.919726	0.241816	0.189437
vision	0.580456	0.269875	0.095162
RT(6)	0.345665	0.541139	0.160116
RT(9)	0.818502	0.354997	0.568550

3.3 Correlation Analysis of Factors

Age and Sensitivity to Light. The correlation analysis revealed the regulation between age and averaged sensitivity to light. The *P*-value is 0.054 and the correlation coefficient (R) that is −0.5395 which prove that the sensitivity to light has negative correlation with age. This result stated that the age may influences the human's sensitivities to light.

Age and Reaction Time. The results did not reveal that age is one of the influencing factors of the reaction time in recognizing the characters. The correlation *P*-values were 0.165 (6-point font), 0.283 (9-point font) and the correlation coefficients (R) were 0.341 (6-point font), 0.267 (9-point font). The explanation of this phenomenon could be due to the fact that the font size was not small enough.

Vision and Reaction Time. The results did not indicate that vision was the influencing factor of the reaction time, either. The correlation return *P*-values were 0.398 (6-point font), 0.201 (9-point font) and the correlation coefficients (R) were 0.212 (6-point font), 0.316 (9-point font). The reason may be that the font sizes were large enough for the subjects to see them clearly despite the differences in their vision.

Sensitivities to Light and Vision. The linear correlation was not obvious between averaged sensitivity to light and vision. *P*-value was 0.1025 and the correlation coefficient (R) was −0.432. The relation may exist between them but need further researches to testify.

Sensitivities to Light and Reaction Time. Averaged sensitivity to light had significant relation to their reaction time for 9-point characters. The correlation return *P*-values were 0.545 (6-point font), 0.016 (9-point font) and the correlation coefficients (R) were −0.153 (6-point font), −0.560 (9-point font). Special attention should be paid to this relation in future studies.

4 Conclusion

The results suggest although senile group seem to have disadvantages in sensitivity to light, vision and reaction time, t-test and linear correlation prove that age only has significant negative correlations with sensitivity to light. When designing the luminance of human-machine interface, the relation between the sensitivity to light and age should be taken into consideration. However, another relation between the sensitivity to light and reaction time of the recognizing experiments shows significance for the larger characters (9-point font size). The reason of this phenomenon may be that the experiments have controlled the indoors illumination rated at 300 lx. In this condition, the difference among the individuals' sensitivities to light are not revealed by the recognizing experiments. In the further research, the indoor illumination condition should be changed to find out the relation between these two factors and number of subjects should be increased.

The vision test reveals that the senile group have a certain gap with the youngest and the middle aged one. However, the difference is not big enough to justify the relation between the age and the vision. The relations of vision, sensitivity to light, and the reaction time are not revealed by the experiment. The probable reasons could be that the font sizes in the experiment is not small enough to reveal the difference in the subjects' vision.

This study discusses the sensitivity to light, reaction time on recognizing experiments, age and vision, but there exist some limitations:

(1) The quantity of the subjects is too small. In the future research, the quantity of the samples should be increased;
(2) Except the factors that these experiments included, there are also some other factors having a great impact on the font recognition which need considering in the future research;

To conclude, In the future, for the sake of the aged people, it would be of great importance to work out the critical and optimal font sizes under different light conditions. To obtain this goal, more experiments should be done and more subjects should be involved.

Acknowledgments This study was supported by science and technology support program 2014BAK01B01.

References

1. Alavi, M.V.: Aging and vision. In: Retinal Degenerative Diseases, pp. 393–399. Springer International Publishing (2016)
2. Ko, P., Mohapatra, A., Bailey, I., Sheedy, J., Rempel, D.: Effects of font size and reflective glare on text-based task performance and postural change behavior of presbyopic and nonpresbyopic computer users. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 56, no. 1, pp. 2378–2382. SAGE Publications (2012)

3. Vartanian, G.V., Li, B.Y., Chervenak, A.P., Walch, O.J., Pack, W., Ala-Laurila, P., Wong, K. Y.: Melatonin suppression by light in humans is more sensitive than previously reported. J. Biol. Rhythm. **30**(4), 351–354 (2015)
4. Danilenko, K.V., Plisov, I.L., Wirz-Justice, A., Hebert, M.: Human retinal light sensitivity and melatonin rhythms following four days in near darkness. Chronobiol. Int. **26**(1), 93–107 (2009)
5. Najjar, R.P., Chiquet, C., Teikari, P., Cornut, P.L., Claustrat, B., Denis, P., Gronfier, C.: Aging of non-visual spectral sensitivity to light in humans: compensatory mechanisms? PLoS ONE **9** (1), e85837 (2014)
6. McDougal, D.H., Gamlin, P.D.: The influence of intrinsically-photosensitive retinal ganglion cells on the spectral sensitivity and response dynamics of the human pupillary light reflex. Vision. Res. **50**(1), 72–87 (2010)
7. Wood, J.M., Lacherez, P., Tyrrell, R.A.: Seeing pedestrians at night: effect of driver age and visual abilities. Ophthal. Physl. Opt. **34**(4), 452–458 (2014)
8. Grobelny, J., Michalski, R.: The role of background color, interletter spacing, and font size on preferences in the digital presentation of a product. Comput. Hum. Behav. **43**, 85–100 (2015)
9. Whatham, A.R., Mermoud, C., Déruaz, A., Goldschmidt, M., Zesiger, P., Safran, A.B.: Computer-based measurement of letter and word acuity. Ophthal. Physl. Opt. **26**(2), 156–168 (2006)
10. Aparicio, J.A., Arranz, I., Matesanz, B.M., Vizmanos, J.G., Padierna, L., González, V.R., Issolio, L.: Quantitative and functional influence of surround luminance on the letter contrast sensitivity function. Ophthal. Physl. Opt. **30**(2), 188–199 (2010)

Cognitive Ergonomics Applied to the eQRH: Developing an Electronic Quick Reaction Handbook for Use During Aviation Emergencies

Marcia Crosland, Wesley Wang, Jerry Ray, Stuart Michelson and C.J. Hutto

Abstract Technology has transformed much with regard to in-flight use of electronic flight bags (EFBs), but little formal literature addresses guidelines for electronic (non-print) presentations designed to aid in "quick reaction" situations such as emergencies or abnormal conditions in flight. This paper contributes to the body of knowledge by describing the development of an electronic "Quick Reaction Handbook" (eQRH) using content from an aviation platform (the KC-135 aircraft). We describe *design features* that consider (1) technology capabilities and limitations, (2) the operational environment, and (3) human capabilities and limitations in stressful situations. We discuss specifically tailored cognitive engineering/cognitive ergonomics methods for overcoming challenges during development (such as translation of content from paper-based media, subject matter expert knowledge elicitation, and eQRH verification and validation). We describe generalizable cognitive ergonomics *design principles* for the eQRH, and formulate a procedural template (process model) and technical data "content templates" for general eQRH development.

Keywords Human factors · Cognitive ergonomics · Quick reaction handbook

---

M. Crosland · W. Wang · J. Ray · S. Michelson · C.J. Hutto (✉)
Human Systems Engineering Branch, Georgia Tech Research Institute (GTRI), Atlanta, GA, USA
e-mail: Clayton.Hutto@gtri.gatech.edu

M. Crosland
e-mail: Marcia.Crosland@gtri.gatech.edu

W. Wang
e-mail: Wesley.Wang@gtri.gatech.edu

J. Ray
e-mail: Jerry.Ray@gtri.gatech.edu

S. Michelson
e-mail: Stuart.Michelson@gtri.gatech.edu

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_3

1 Introduction

Aviation checklists specify a series of defined procedures, equipment parameters and operator actions intended to achieve peak performance in both normal and emergency flight conditions. This information is developed by airplane manufacturers, and is provided to aircraft owners as procedural manuals and checklists after flight testing and approval. Converting aviation checklists to different formats requires meticulous attention to maintaining accuracy and completeness of the original content.

In many modern aircraft, checklists are displayed on integrated cockpit displays to guide pilots through required steps, which may be ordered by priority and/or color-coded to indicate urgency. As steps are completed, the crew must acknowledge each checklist item, which then disappears from the screen before proceeding to the next item.

In older aircraft, aviation checklists are contained in paper copies of technical user manuals. These manuals are commonly organized according to the aircraft's various engineering systems, and indexed by a table of contents. The aircrew must know (and recall) which system is at fault to solve problems in flight. Often, the procedures themselves are written in paragraphs of prose text, which are overly verbose and fail to prioritize required in-flight procedures over optional, ancillary or ground procedures. This does not support quick and easy access to emergency procedures.

Recent advances in mobile technology have spurred the use of electronic flight bags (EFBs) in the aircraft. Increasingly, aircrews use mobile tablets to display their aviation procedural manuals and checklists to declutter the flight deck. Some aircrews host their full aircraft technical procedural manuals converted into PDFs on their EFBs, while others have some form of interactive PDF.

Electronic presentations of emergency procedures to aid in quick reaction conditions must consider specific design criteria beyond simple conversion of paper manuals into digital representations of those same pages. The design must take into consideration the capabilities and limitations of technology, the environmental conditions in which emergency response action will occur, and the human capabilities and limitations within such stressful situations. Technology capabilities such as automation in human-system interactions and intuitive graphical representations of procedures help to improve effectiveness and efficiency; however, technology limitations such as battery life, susceptibility to damage (e.g., from impact or spilled liquid), and effects of screen lighting on human vision in dark conditions are all important considerations. The environment is often one in which only a short amount of time is available for correct, rapid execution of a critical sequence, and the consequences of mistakes and errors are very high. Human cognitive loads are often quite high, as aircrew members must rapidly integrate information and data from multiple sources in the flight deck. They must simultaneously process the perceptual input of the emergency and attempt to reconcile the observed abnormal situation with the correct emergency procedures (EPs). During execution of the

EPs, aircrew must also maintain a mental map of the procedural flow, as well as their (or their team members') progress through that flow.

Typically, aircrew members memorize a subset of EP steps for the most critical emergencies (the so-called "bold-face" EPs, due to their emphasis in text presentations as bold-style type) [1]. With training and repetition, the paired cognitive and motor activities required to perform the checklist eventually transition to procedural memory, where retrieval is less affected by time pressure [2]. On the other hand, aircrew perform non-boldface EPs when time is not as critical (relative to bold-faced emergencies); nevertheless, non-boldface EPs are still time sensitive. Typically, aircrew practice non-boldface EPs less often, so they would benefit from using a checklist as a memory aid. When multiple systems fail, aircrew depend on procedural flowcharts and decision trees to help direct emergency responses. Reviewing a checklist helps overcome anchoring tendencies (relying too heavily, or "anchoring," on just some part of the information when making decisions) and confirmation bias (interpreting information in a way that confirms one's preconceptions, leading to errors) [2].

The above sets of considerations form the context and landscape for our current work. In the remainder of this paper, we discuss how we apply cognitive ergonomics toward developing an Electronic Quick Reaction Handbook (eQRH) for use in aviation emergencies. We discuss specifically tailored cognitive engineering/ cognitive ergonomics methods for overcoming challenges during development (such as translation from existing paper-based manuals and checklists, subject matter expert (SME) knowledge elicitation, and eQRH verification and validation). We then identify and describe several underlying generalizable design principles for the eQRH.

2 Development Process

For a stand-alone electronic Quick Reaction Handbook (eQRH) to be effective as an emergency resource, it must be accurate procedurally, and be relatively easy and efficient to use on a tablet device under operational conditions. To accomplish this, eQRH development should focus on two areas: streamlining the emergency procedure content contained in the official paper flight manual, and creating an effective visual presentation for an interactive eQRH viewed on a tablet.

The team applied a Human Systems Engineering (HSE) methodology to guide the analysis and development of the eQRH functional and technical requirements. As shown in our HSE process model (Fig. 1), the methodology includes a task analysis, subject matter expert (SME) knowledge elicitation, identification of applicable design guidance, prototype creation and formative evaluation. In a typical HSE analysis, the first step is a task analysis of the operator tasks in the environment of use. This would yield the operators' information needs, task steps and potential sources of errors. In the present challenge, an existing flight manual, "the user manual," published by the aircraft manufacturer thoroughly documents

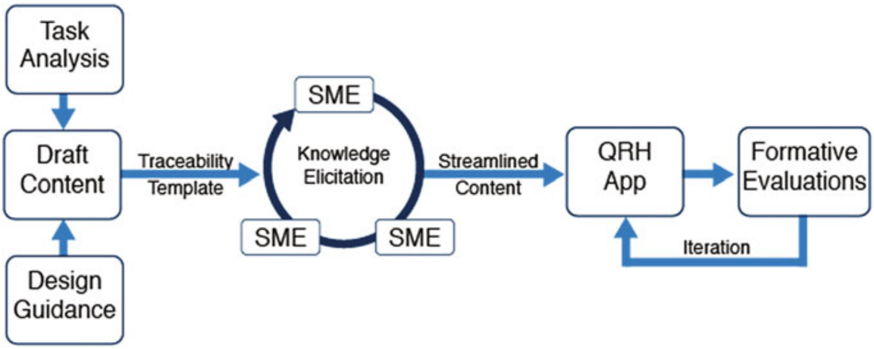


Fig. 1 Human systems engineering process for eQRH development

the operators' procedures for use in emergency situations and is endorsed as the official document. As such, we performed a detailed task analysis on the emergency procedures contained in the user manual and then meticulously converted them into an eQRH format. That is, we transformed both the content and the presentation of these emergency procedures so that they could be accessed (perceptually), interpreted (cognitively), and processed (mentally, physically) rapidly within a stressful time sensitive context without losing meaning.

The task analysis (TA) is a systematic, documentable method to define the decisions and physical tasks that a user has to execute for each of the emergency procedures. To consolidate all of the task analysis information to support the work of the SMEs during review, we generated a traceability template. The traceability template contains the original source material and the TA information organized in a logical order so that it flows from the source material to the detailed task description to the eQRH procedural representation. The traceability template serves as the historical record of changes for later review by a standards evaluation committee and as input to eQRH developers.

The next step in eQRH development was to conduct SME reviews of the traceability templates for the purpose of content verification. The SME reviews focused on two areas: verifying the accuracy of the task analysis, and restructuring the content into an improved organization and streamlined format. In verifying the task analysis accuracy, SMEs judged the content and sequence accuracy, completeness, decision branching structure, links to supporting procedures and appropriateness of the placement of warnings and cautions. In restructuring the content, SMEs recommended ways to reorganize, reduce and merge content into integrated checklists where appropriate to facilitate rapid execution of procedures in an unfolding emergency.

In a parallel work stream, we produced design guidance for presenting procedural content in an interactive form. The initial design research identified a list of human factors guidelines and standards from MIL-STD-1472G [3] and the Federal Aviation Administration's Human Factors Design Standard (HFDS) [4] that pertain to the design of a tablet-based emergency procedure manual. These guidelines covered such features as touch zone size and spacing, font size and use of color. A second literature review identified guidelines specific to the design of stand-alone electronic checklists for abnormal and emergency situations. These guidelines cover very general characteristics such as color usage, legibility, layout, auditory alerts, titles, checklist order, branching, content, tracking and evaluation [1, 5–8].

To assist in selecting a presentation approach for the eQRH, we compared the design features of two tablet-based electronic approaches to the baseline paper procedural manual (Table 1). The goal was to select an eQRH information technology approach that would be easy to use during an emergency. This requires that the technology support quick performance, reduce cognitive and physical workload and minimize errors in the aircraft.

A final step in the eQRH development process was to conduct formative evaluations to validate the design. Formative evaluation began with soliciting input about the procedural representation and the visual presentation in early prototypes of the eQRH. The specific goal was to determine if the tool meets the operational needs of the user in an emergency in three areas: procedures (are these the right

Table 1 Comparison of design features for different quick reaction handbook formats

Design feature	Text-based paper user manual	Text-based Electronic manual (Interactive PDF) Mobile tablet	Electronic quick reaction handbook (eQRH) Mobile tablet
Organization	Groups procedures by systems[1], supports engineering coherence	By systems, or functional—organized around decision trees, supports cognitive coherence[2]	Functional—organized around decision trees, binary decisions, supports cognitive coherence[2]
Navigation	Table of contents	Table of contents, touchable tabs, scroll	Table of contents, hyperlinks, decision buttons, scroll, back button, multi-pathing
Level of detail	Full disclosure of all procedures, notes, warnings, cautions, background, and rationale within the visual field across physical pages	Full disclosure of all procedures, notes, warnings, cautions, background, rationale within the visual field across scrollable pages	Progressive disclosure of information required for specific conditions
Layout	Page, paragraph layout, inconsistent location of all information across pages	Pages, flowcharts extending over multiple screens and page breaks, inconsistent location of content and links	Consistent display of steps in response to user selections, consistent location of interactions and procedures

(continued)

Table 1 (continued)

Design feature	Text-based paper user manual	Text-based Electronic manual (Interactive PDF) Mobile tablet	Electronic quick reaction handbook (eQRH) Mobile tablet
Search	n/a	Keyword	Keyword
Touch zones	n/a	Generally does not meet MIL-STD-1472 size requirements	Meet MIL-STD-1472 size req's
Typography	Fixed	Adjustable size with zoom	Meets MIL-STD-1472 size req's, can select a larger size
Color	Yes	Yes	Yes
Wording	Mix of action oriented and verbose explanation, engineering speak	Action oriented, concise	Action oriented, concise
Sequencing	n/a	Higher priority steps can be presented first for conditions	Higher priority steps can be presented first for conditions
User history, visual feedback	n/a	n/a	Maintain historical list of user actions, visual highlight of user path, mark completed steps (optional)
Potential technology errors	n/a	Accidental activation of navigation controls, markup tools, and overlays; accidental scaling	Accidental button presses (recoverable)
General technology considerations	n/a	Battery life, requires touch-compatible flight gloves	Battery life, requires touch-compatible flight gloves
Maintenance	Edit a document file and print	Edit a document file and publish, combine with an interaction layer	Edit and publish new interactive content

[1]*Engineering Coherence* Match the engineering system; *complete* and accurate [8]
[2]*Cognitive/Logical Coherence* Match user's need in situ; *specific* and accurate [8]

ones for the emergency?); procedural layout and presentation order; and interaction design paradigm. The early evaluations followed a semi-structured interview format to elicit responses about specific aspects of the eQRH design. We used formative evaluation results to make iterative design changes. Once the design develops to a relatively mature state, we will conduct formative evaluations in a simulator to gain a deeper view of the usability of the eQRH. After design and development concludes, we will conduct a summative evaluation to assess whether the eQRH meets performance criteria when executing emergency tasks.

3 Discussion

In this paper, we describe a process model for translating aviation emergency procedures from paper format to a stand-alone tablet-based eQRH. Here, we list a number of specific design features that should be considered during the design and development of an eQRH. During development, we identified eight generalizable design guidelines and their corresponding underlying cognitive principles that inform the design for the eQRH (Table 2).

Table 2 Guidelines and cognitive principles for the design of an electronic quick reaction handbook (eQRH) for use in emergency situations

Guidelines	Cognitive ergonomics principle	eQRH application
1. Make it quick and easy to find the correct procedure	High cognitive workload	Streamline content. Minimize depth of procedural hierarchy (to 2 levels). Use meaningful titles and labels. Provide keyword search and auto-scroll
2. Provide decision support	Decision making, working and prospective memory	Organize content around decision trees. Display explicit decision alternatives, such as binary yes-no questions. Visually highlight user path and decision history
3. Prioritize and preview steps	Direct priming	Show steps to get safe first. Preview alternative procedures for later execution that impact early decisions and preparation
4. Display only information needed	Focus attention	Display only information needed at the time. Do not show steps that do not apply at the time. Provide supporting data such as charts and tables as collapsible panels
5. Keep phrasing direct and object-oriented	Comprehension, conceptual priming	Display concise wording with the object followed by the procedural action. Use a consistent lexicon
6. Minimize errors and create error tolerance	Cognitive aiding	Back button. Simple yes and no selections. Show user history. Avoid controls that induce error. Follow human factors guidelines for touch target sizes and legibility
7. Create multiple ways to get information	Different mental models	Support multi-pathing, i.e., multiple ways to find procedures. Provide search, TOC, links between procedures
8. Maintain design consistency	Procedural learning and memory	Consistent placement of content and common controls. Use standard design conventions

3.1 Guideline 1. Make It Quick and Easy to Find the Correct Procedure

During high cognitive workload, support the aircrew by making it quick and easy to find the correct procedure(s). Design the user interface to reduce the number of interactions (keystrokes, button presses, page turns and scrolls) to facilitate locating the correct procedure more quickly. The design should minimize (within allowable bounds) distance between touch targets to reduce the time to physically move between targets [9]. The design should reduce reading time by using the least amount of text needed to convey meaning. Titles and labels should be sufficiently descriptive so the aircrew can select between procedures with similar initiating conditions [10]. Providing a keyword search that accommodates a wide variety of problem-oriented terms speeds access when users do not see the exact title or label they are seeking. Never ask the user to take an action that the application can do for them. A good example is auto-scrolling: when a user presses a button, such as yes or no in response to a diagnostic question, the display should automatically scroll so that as much newly-displayed content as possible is displayed.

3.2 Guideline 2. Provide Decision Support

Providing decision support in an emergency essentially means leading the aircrew through the process of diagnosis and recovery. Aircrews often use decision tree flowcharts when multiple systems fail. When explicit decisions are required, one approach is to show only the correct procedural path [1, 7]. To accomplish this in the eQRH, we display decision content as explicit alternatives (where possible, we use simple binary yes-no questions). The eQRH binary question presentation allows users to answer simple questions and displays only the appropriate actions for the response. Additionally, where appropriate, we provide quick references that allow users to view cognitive aids such as diagrams and illustrations. These quick reference aids are implemented with links or collapsible panels that the aircrew can open and close for fast and easy access.

Another technique to support decisions is by visually distinguishing procedures that have been visited or performed—indicating progress by differentiating completed steps and selected choices. This alleviates demands on users' working memory to recall where they are in the procedure. Maintaining a user's history also eases the burden on prospective memory to remember skipped steps that still require action. Examples of visual differentiation in an eQRH include showing different colors (or differently styled text, such as bold) for (non-) visited user decision/action paths, maintaining a user history that the user can consult, and allowing users to explicitly mark completed steps.

3.3 Guideline 3. Prioritize and Preview Steps

When displaying an emergency procedure, it is essential to prioritize the display of those steps that will remove immediate danger [7, 11, 12]. However, it is also critical to preview alternative procedures for later execution that impact early decisions and preparation. For these later procedures, the aircrew may need to consider alternative procedural paths and their requisite steps, such as landing with a manual flap extension or crossover. This allows the aircrew to consider the options more fully and to better prepare. Preview can be accomplished with hyperlinks or collapsible panels. By viewing the alternative information early, there is a positive effect similar to direct priming [13] that will make EP execution more operationally effective overall, with increased speed and efficiency in later steps.

3.4 Guideline 4. Display Only the Information Needed

During an emergency, there are many demands on attention. The eQRH should focus attention on a small set of information within the visual field of the display [7]. As a default display, only show steps that apply for the current condition. This technique will minimize distractions when cognitive load is high. However, do provide easy access to supporting information in hidden panels. For example, place supporting charts and tables in collapsible panels and display them only when needed.

3.5 Guideline 5. Keep Phrasing Direct and Action-Oriented

The wording for emergency procedural steps should be concise, direct and object-oriented [7]. Display the object first followed by the specific action required by the aircrew. For example, display "Thrust—INCREASE (until forward separation)" rather than "Tanker Pilot—Increase thrust as required (full throttle if necessary) to obtain forward separation; ensure that the aircraft does not descend." Aircrew are generally very well trained on emergency procedures for the aircraft. Upon viewing the word "Thrust," the pilots will quickly bring to mind the procedures associated with it even before seeing the word INCREASE. This cognitive prompting is a form of conceptual priming [14]. The word "increase" is associated with a number of aircraft objects, so displaying it first would have little benefit.

The use of direct, positive phrasing and a standard lexicon [7] will improve comprehension of procedures. For example, "Is the generator working?" is easier to understand than a negative phrase "If the generator is not working." Another technique to improve understanding is the consistent use of standard word pairings. For example, the phrase "If the GCB CIRCUIT OPEN light remains out" is easier to understand when phrased “Is the GCB CIRCUIT light off?” In this example, a light is either on or off.

3.6 Guideline 6. Minimize Errors and Create Error Tolerance

In the eQRH design, avoid error-prone controls and lay out the controls to reduce errors. Follow human factors guidelines for touch target sizing and spacing [3]. Avoid the use of controls that induce errors particularly on touch screen tablets, such as slider bars and scaling features. Provide mechanisms for recovering from errors, such as a “Back” button and a user history.

3.7 Guideline 7. Create Multiple Ways to Get Information

Different users have different cognitive and interaction styles. The eQRH user interface should be flexible enough to support different mental models [15] for information access and usage. An effective eQRH design will support shortcuts [7] and multiple ways to find procedures, i.e., multi-pathing. To implement multi-pathing, provide a table of contents, search capability using a variety of terms and links between related procedures.

3.8 Guideline 8. Maintain Design Consistency

Design consistency aids procedural learning and performance and is a “hallmark of good checklist design” [12]. Design consistency can be achieved by the placement of content and common controls in the same location on each page. This technique helps train eye and hand movements through repetition, which positively impacts procedural memory retrieval [16]. For example, display a procedure title at the top of the page that does not scroll off the screen, and display the steps in a consistent area of the screen. Common controls (such as home, back and next buttons) should always appear in fixed locations. Other common controls such as selection buttons (yes and no) and jump buttons that link to other procedures may appear in different rows on the page but should appear in the same column location each time they appear.

3.9 Practical Challenges with the Development of an EQRH

We have identified several practical challenges with the development of the eQRH. First, there will be a need to take a printed copy of the QRH onto the aircraft in addition to the electronic version (as a backup). Developers will need to consider how to print interactive content. As listed earlier, a paper format of a user manual has a certain organization and layout that differs from an interactive app based on decision tree logic with binary decisions. One approach is to create a simple set of the eQRH flowcharts that follow the decision tree logic for printing. A second challenge is the need to periodically update and maintain the electronic content. We have considered a simple eQRH editor application that would allow for modifying content as needed. Although this paper has focused on specific design guidelines for the eQRH development, it highlights the importance of the design guidance for interactive applications used on electronic flight bags (EFBs).

References

1. Degani, A., Wiener, E.L.: Cockpit checklists: concepts, design, and use. hum. Factors J. Hum. Factors Ergon. Soc **35**, 345–359 (1993)
2. Clay-Williams, R., Colligan, L.: Back to basics: checklists in aviation and healthcare. BMJ Qual. Saf. **24**, 428–431 (2015)
3. Department of Defense: Design Criteria Standard Human Engineering. Government Printing Office, Washington DC (2012)
4. Federal Aviation Administration: Human Factors Design Standards (HFDS). Goverment Printing Office, Washington DC (2003)
5. Chandra, D.C., Mangold, S.J.: Human Factors Considerations in the Design and Evaluation of Electronic Flight Bags(EFBs)- Version 1: Basic Functions. USDOT Volpe Center, Washington D.C (2000)
6. Chandra, D.C., Yeh, M., Riley, V., Mangold, S.J.: Human Factors Considerations in the Design and Evaluation of Electronic Flight Bags (EFBs) Version 2. Office of Aviation Research, Washington DC (2003)
7. UK Civil Aviation Authority: Guidance on the Design, Presentation and Use of Electronic Checklists. The Stationery Office, United Kingdom (2005)
8. Burian, B.K.: Design guidance for emergency and abnormal checklists in aviation. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **50**, 106–110 (2006)
9. MacKenzie, I.S.: Fitts' law as a research and design tool in human-computer interaction. Hum.-Comput. Interact. **7**, 91–139 (1992)
10. Burian, B.K.: Aeronautical emergency and abnormal checklists: expectations and realities. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **50**, 101–105 (2006)
11. Burian, B.K.: Human factors in aviation-the good, the bad, and the ugly: Some lessons from NASA for medicine (2015)
12. Degani, A., Wiener, E.L.: Human factors of flight-deck checklists: the normal checklist. NASA Ames Research Center (1990)
13. Forster, K.I., Davis, C.: Repetition priming and frequency attenuation in lexical access. J. Exp. Psychol. Learn. Mem. Cogn. **10**, 680–698 (1984)
14. Tulving, E., Schacter, D.L., Stark, H.A.: Priming effects in word-fragment completion are independent of recognition memory. J. Exp. Psychol. Learn. Mem. Cogn. **8**, 336–342 (1982)

15. Kitchin, R.M.: Cognitive maps: what are they and why study them? J. Environ. Psychol. **14**, 1–19 (1994)
16. Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O.T., Oitzl, M.S.: Stress effects on memory: an update and integration. Neurosci. Biobehav. Rev. **36**, 1740–1749 (2012)

The Impact of Chinese Cultural Elements on the User Experience in Government Website Design

Tian Lei, Xu Liu, Lei Wu, Ziliang Jin and Yuhui Wang

Abstract E-government has been recognized the most effective way of improving the government's public management and service efficiency, but at the present stage of China's government web site is a big gap from the expectations of the public. A number of studies have compared the differences of government website design in the style of image under the different cultural atmosphere. Culture plays an important role in interface design, so in the atmosphere of Chinese traditional culture, there are a lot of cultural identification characteristics affecting the user's cognition and habits. This paper will study the mapping relationship between design elements and the cultural mental image, as well as the relationship between the government websites design integrated into the cultural elements and the user experience.

Keywords Government website · Cultural mental image · User experience

1 Background

E-government has been recognized around the world as the most effective way to improve the government's public management and service efficiency, but China's government website still has a long way to reach the expectations of the public at the present stage. For example: in order to conform with the characteristic of national, city, regional, many governments embedded different culture elements into website through a simply copying or grafting manner to build a traditional style website. Although this method will make user feel the traditional culture, produce a sense of familiarity,but it will also bring rigid sense to users. This will make differences with existing cognitive habits. A good user experience needs UI elements and website style match their user's cognitive and preference under the culture background.

T. Lei (✉) · X. Liu · L. Wu · Z. Jin · Y. Wang
Department of Industrial Design, Huazhong University of Science and Technology, Wuhan, China
e-mail: andrew.tianlei@hust.edu.cn

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_4

2 Literature Review

2.1 Culture

Culture is a very abstract concept that relate to every aspect of society. The research of culture in academic world has been never stopped. Many research had compare the difference of government website culture image style which building in different culture environment, those research proved that culture is playing an important role in user interface design, culture background will impact user's cognitive of a website style [1]. For example, Mushtaha and De Troyer [2] found that the cultural background would influence users' feeling, choice and perception of the websites, based on the data from subjects. Smith et al. [3] used the Taguchi Method to study the difference between British and Chinese satisfactions and preferences of websites, based on Hofstede's study on the generic cultural differences. This study indicated that the attitude towards e-financial websites was very different. For example, Chinese users preferred to browse the website in a more general way compared to British users. These studies showed that the cultural context indeed affected the users' cognition and behaviors. This was because Chinese have holistic thinking style, which was possibly resulted from the deeply rooted Confucian philosophy.

In China, the culture is a kind of reflects on the characteristics of national style, it is overall characterization of the national history of ideological culture and ideology, Chinese traditional culture has various forms, according to the content, they can be classified as: exponents, piano chess calligraphy and painting, literature, traditional festivals, Chinese opera, traditional Chinese architecture, Chinese characters, Chinese, traditional Chinese medicine, philosophy of religion, folk art, Chinese Kung-fu, and many other types [4]. Traditional culture has a variety of forms, and have a deep, profound impact on the user, how form cultural into our interface design from inside to outside, and make it more in line with user's cognitive habits and visual style under the influence of Confucian culture, has been one of hot issues in Chinese interface design field.

There are a lot of traditional culture elements interface design cases, they provide us with a good idea: designer use cultural elements in a government website interface design, expresses the unique style of the culture image; Users can convey the cultural connotation according to the interface designer's visual elements, so find the mapping relationship between culture image and cultural elements, can not only help designers expression the culture image more accurately, but also make users feel more familiar, and have a more enjoyable visual experience [5].

2.2 Mental Image and Style

Mental image is an important concept in art and design field, but in a long term, we still cannot use a unified concept to define it, and it means under different cultural backgrounds are different, in contemporary industrial design field, mental image has been used to express the concept of product, it make direct association with product form via the user's visual sense. Because those elements have similar characteristics and rules, and they represent the same mental image, and style is the best word to describe their properties [6]. As a method of the art research, style was originally produce in the art field, and it is used to distinguish different art forms, and different works in same art form, style was affected by the culture and geography.

We can think of style as a kind of benchmark, it can determine the time and place of the original creative, and also it can used to track inter-linkages between artistic groups, it can be used to describe the characteristics of cultural or material goods. Each user has a different style evaluation benchmark, under the same cultural background, there are also differences in cognitive of culture element, what mapping relationship between different design elements and culture image style? If we can find the corresponding relationship between the design elements and style imagery, this relationship must be able to help designer's expression the style imagery more accurate on interface, and selected elements to render the interface style in a faster way.

2.3 User Experience

Norman propose that the user experience is a key indicator to measure the quality of products, it can be considered that the user experience is based on cognitive and emotional experiences [7]. Moelinger [8] think that user experiences is composed of three part: practical experience, sensory experience, emotional experience. Lucas Daniel [9] said: user experience is user's practices, ideas and feelings during interaction, including the functionality experience and emotional experience that product provided to the user, user experience is highly subjective.

User experience is process that cognition something from outside to the inside, it is a mixture of behavior and emotions, so it is emotional. For the users that in the cultural background, if there is accurate mapping relationship between interface design elements and cultural mental image, whether this visual effect will let users feel more familiar, pleasure and produce a better user experience feelings, which it is worth to exploring.

3 Experiment I—Test Between Culture Mental Image and Design Elements

The first stage is to collect sample and culture image vocabulary, we extensively collect government websites at domestic and abroad, classify the sample according to the style, color, layout, characteristics, and the reference expert select typical sample in each category for the final sample. Next is to deal with the samples, the purpose is to set a one-to-one correspondence between web design elements and sample.

Each factors including different categories and forms, the main performance of the web elements including page layout, color, page content, etc., in this paper, each design elements can be divided into three categories, such as A design elements for the “layout”, About three kinds of features are: A1, frame type, A2. Top and bottom frame type, A3. Integrated framework, such as these, these are the design elements of the specific characteristics of the corresponding to each sample. And then the collection and selection the vocabulary which can reflect culture image.

3.1 Material: Mental Image Vocabulary and Samples

Widely collect vocabulary related to web design and cultural image, and selected 32 groups of vocabulary, then print them into the image form and make a 5 point scale questionnaire, from 1 to 5 points in turn represent “completely unable to express web interface style”, “can’t express web interface style”, “general”, “comparison can express web interface style”, “completely can express web interface style”. The participants according to image vocabulary evaluate the web interface style with scale level.

Eight male subjects, twelve female subjects, they are between the ages of 21 and 30, 15 participants from industrial design major, the other five students from non-industrial design professional, after test statistics, we select the following 12 groups of mental image vocabulary as shown in Table 1.

We collect many website interface at home and abroad, including government agencies, research institutions, universities and other sites, finally collected 118 screen-shots of the home page. The four kinds of layout of the interface as classification criteria, 118 websites can be divided into four categories, each category selected four typical samples, at last we select 16 different style websites, used as a research sample, as shown in Fig. 1.

Table 1 Groups of mental image vocabular

No.	Mental image	No.	Mental image	No.	Mental image
1	Elegant–Vulgar	2	Mechanical–Lively	3	Modern–Classic
4	Plain–Gorgeous	5	Popular–Personality	6	Monotonous–Rich
7	Rough–Delicate	8	Conflict–Harmonic	9	Local–International
10	Crowded–Sparse	11	Tolerance–Single	12	Heavy–Light

Fig. 1 Select part of the samples

3.2 *Sample Analysis*

The elements of web design has been summarized and classified by a lot of professional designers, each factors including different categories, each category has different forms. The main forms of the web interface elements including page layout, color, page content, banner, each design element has different characteristics, and these specific category can correspond to any samples (Table 2).

We analyze the design elements of the 16 samples, summing up the categories of them in each design element, the category with the letters and Numbers, the following is a former seven samples (Table 3).

Table 2 Website design element category

Design elements	The specific category		
A. Layout	A1. Left-right structure	A2. Up-down structure	A3. Comprehensive structure
B. Color hue	B1. Cold	B2. Warm	B3. Neutral
C. Page content	C1. More	C2. Less	C3. Moderate
D. Banner	D1. Large image	D2. Small image	D3. No banner

Table 3 Sample design element analysis

Design element / Sample	A. Layout	B. Color hue	C. Page content	D. Banner
1	A2	B1	C1	D2
2	A2	B1	C2	D1
3	A1	B1	C3	D2
4	A3	B2	C3	D1
5	A1	B1	C1	D1
6	A3	B2	C2	D1
7	A3	B1	C2	D1

3.3 Task and Procedure

Screening out of 12 images of vocabulary and 16 representative samples as experiment material, made 7 levels of semantic differential method questionnaire, for example, "Classical–Modern", participants choose to the left of the rating indicates that the samples are more inclined to "Classical" style, more to the right rating indicates that the sample more inclined to the style of "Modern".

3.4 Result

We put the questionnaire data input SPSS statistical software, the following Table 4 is the average scores of each sample about cultural image words (score range: −3 on behalf of the left image, 3 on behalf of the image on the right side, 0 for moderate), the following is a former seven samples.

Based on the average score of cultural mental image, establish a corresponding relationship between design elements and the cultural mental image. The purpose of analysis is to establish multiple linear regression equation of each mental image. Equation: represents a certain image prediction score value (that is the mental image number axis location on [−3, 3]), represents the value of each design element (0 or 1), the coefficient of is looking for. Data analysis method is Quantitative I theory and multiple regression analysis.

Establish reaction matrix [1 representative belong to this kind of item, 0 representative does not belong to this kind of item (Table 5)].

Establish Normal system of equations to a certain mental image, by the least square theory, available equation:

$$X^T XB = X^T Y \tag{1}$$

Table 4 The average score of cultural mental image

Sample / Mental image	1	2	3	4	5	6	7
Elegant–Vulgar	0.1875	0	−0.937	−1.562	−0.125	−2.562	−0.432
Mechanical–Lively	−1.25	1.25	0.187	0.687	1.562	1.437	−0.062
Modern–Classic	0.0625	1.562	0.937	0.687	1.562	2.125	1.125
Plain–Gorgeous	−1.125	0.812	−0.25	0.687	0.875	−0.18	−0.125
Popular–Personality	−1.687	0.812	−0.5	0.875	1.25	1	−0.375
Monotonous–Rich	−0.875	0.812	−0.187	0.625	1.687	−0.875	0.375
Rough–Delicate	−0.187	1.5	0.625	1.437	1.312	1.437	0.937
Conflict–Harmonic	0.5	1.375	1	1.562	1.625	1.812	1.437
Local–International	−1.687	1.812	1.125	1.5	1.75	1.75	1.062
Crowded–Sparse	0	−0.687	−0.312	−0.437	−0.75	−1.062	−0.562
Tolerance–Single	1.5	−0.187	1	1.125	0.812	0.812	0.875
Heavy–Light	−0.687	0.5	0.687	0.937	1.437	2.062	0.312

Table 5 The reaction matrix

Design element / No.	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3
1	0	1	0	1	0	0	1	0	0	0	1	0
2	0	1	0	1	0	0	0	1	0	1	0	0
3	1	0	0	1	0	0	0	0	1	0	1	0
4	0	0	1	0	1	0	0	0	1	1	0	0
5	1	0	0	1	0	0	1	0	0	1	0	0
6	0	0	1	0	1	0	0	1	0	1	0	0
7	0	0	1	1	0	0	0	1	0	1	0	0

X is the above reaction matrix, Y is the sample of this image for experimental measurement vector, and B is the need of solving the regression equation coefficient vector.

This system of equations with an infinite set of solutions, it can be find any a set of feasible solution and do the normalized processing, the result was the only answer. Get coefficient vector B, and then we can selected biggest coefficient of all which in the category of each design element, it is one of the largest coefficient of design elements impact on the mental image. Experiment found the mapping relationship between cultural mental image and design elements. Take following three groups of cultural mental image vocabulary as examples.

"Plain–gorgeous" mental image vocabulary, the regression equation is:

$$\hat{Y} = 0.0491 + 0.1840A_1 + 0.5918B_3 + 0.2449C_3 + 0.5606D_1 \quad (2)$$

Meaning neutral or mixed color, big Banner, moderate page content, left-right structure can lead to more "gorgeous" mental image.

"Popular–Personality" mental image vocabulary, the regression equation is:

$$\hat{Y} = 0.0893 + 0.1920A_1 + 0.4997B_3 + 0.0231C_2 + 0.7474D_1 \quad (3)$$

Meaning great Banner, mixed color, left-right structure, less page content will cause more "personality" mental image.

"Heavy–Light" mental image vocabulary, the regression equation is:

$$\hat{Y} = 0.4955 + 0.4044A_1 + 0.3864B_3 + 0.1694C_2 + 0.5934D_1 \quad (4)$$

Meaning big banner, up-down structure, neutral or mixed color, more less page content will cause "light" mental image.

The data illustrated that: the style image experiment makes subjectivity, randomness and fuzziness of implicit cognition into dominant design direction, experiment found the mapping relationship between cultural mental image and design elements, every culture mental image vocabulary has the corresponding design elements of the form. This can make a design element to match the user's cultural cognition, finally meet user's expectations of the cultural connotation and style of the website image.

4 Experiment II—User Experience Study

The second stage is user experience experimental study, first we tested in accordance with the scale of semantic differential method, subjects according to the mapping relationship of design elements and the cultural mental image to evaluate. Then combined with the results of the mapping of cultural images and design elements to utilizing the cultural elements into government websites design project in varying degrees. The independent variable are "No Utilization", "Natural Utilization" and "Strong Utilization", the control variable are "interface layout", "page color", "page content volume", the dependent variable are user experience and feedback time.

We only accept references written using the Latin alphabet. If the title of the book you are referring to is in Russian or Chinese, then please write (in Russian) or (in Chinese) at the end of the transcript or translation of the title.

4.1 Material

According to the analysis of the government website features and task flow, we determine the user's main operations and functions in government Web site. The results are as follows: when Users browse government websites, the operation and functions includes browsing news, viewing policy information, doing business and participating discussion. This results will serve as material of experimental II. Participants need to complete 4 tasks on 3 different kinds(which is "No Utilization", "Natural Utilization" and "strong Utilization") of government website interface, and then fill in usability model test subject for recording data of the user experience (Fig. 2).

4.2 Result

The way cultural elements utilizing have a significantly affect relationship with "satisfaction of the way cultural utilizing," "naturalness of the way culture utilizing," "interest of interactive", and we also found that "comprehensible of the Feedback" and "memorable" is irrelevant.

Five coefficients perform best in the Natural Utilization condition, perform worst in the no Utilization condition, (Fig. 3). This shows that despite the integration of cultural elements leads to clarity of feedback reduced, but enhance the other aspect experience of participants.

Fig. 2 The experiment task type

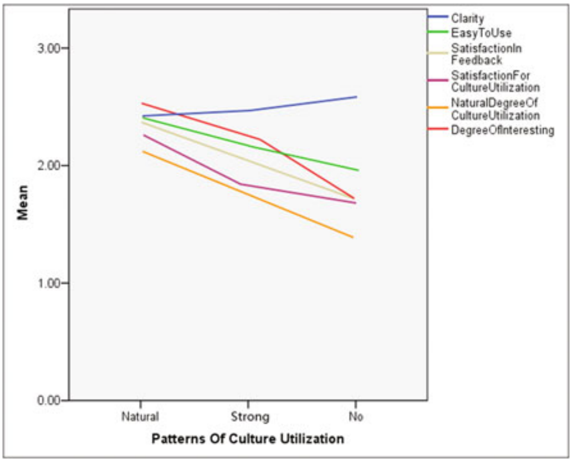


Fig. 3 Relationship between ways of culture utilization

5 Conclusion

The results showed that the mental image experimental transform 'subjectivity of users', 'randomness' and 'fuzziness tacit knowledge' into explicit design direction, so we can get the mapping relationship between design elements and government culture mental image, this result can be able to help designers selected design elements more accurately and quickly to maintain the user's cultural needs. At the same time, experiments also proved that naturally utilizing cultural elements into the government website interface design will enhance the user's awareness of website properties. And it is possible to make the user feel more pleasant, more interesting, enhance the user experience of the site. Part of the data also show that, cultural element utilizing have some negative effects on user experience, such as adding a certain amount of cognitive load which caused by the unfamiliar.

Acknowledgments This paper is supported by the HUST high-level international curriculum projects and the Fundamental Research Funds for the Central Universities HUST: (2014QN017).

References

1. Marcus, A.: Cross-cultural user-interface design. In: Smith M.J., Salvendy G. (eds.) Systems, Social and Internationalization Design Aspects of Human-Computer Interaction, pp. 502–505. Volume 2 of the Proceedings of HCI International 2001, New Orleans, Louisiana, USA, 5–10 Aug 2001
2. Mushtaha A., De Troyer, O.: Cross-Culture and Website Design: Cultural Movements and Settled Cultural Variables (2009)
3. Smith, A., French, T., Chang, Y., McNeill, M.: eCulture: Acomparative study of eFinance web site usability for Chineseand British users. Designing for global markets. Conference (6th 2001). In: Day, D., Duckley, L. (eds.) Proceedings of the 3rd International Workshop

Internationalisation of Products and Systems, pp. 87–100. The Open University, Buckinghamshire (2001)
4. Jianghua, X., Zhang, M.: Reconstruction of Chinese traditional cultural symbol in product design. Packaging engineering, classification codes TB472(2007)
5. Luo, S., Pan, Y.: The perceptual imagery theory in product design, technology and application research. Chin. J. Mech. Eng. (2007)
6. Yang,Y., Tian, L., Zhang, Y.: Handheld mobile device interface design based on user mental models. J. Zhejiang Univ. (2007)
7. Hoft, N.L.: Developing a cultural model. In: Del Galdo, E.M., Nielsen, J. (eds.) International Users Interface, pp. 41–73. Wiley, New York (1996)
8. Moeslinger, S.: Technology at Home: A Digital Personal Scale. CHI 97 Electronic Publications: Formal Video Program (1997)
9. Daniel, L.: Understanding user experience. Web Tech. Press **5**(8), 42–43 (2000)

Research on the Influence of Abstract Knowledge to the Individual Cognitive Behavior and Innovative Design Thinking

Yongjia Zhou, Qing Xue, Jia Hao and Min-xia Liu

Abstract Earlier studies have shown that the degree of abstract knowledge has a significant impact on the result of innovative design, which plays an important role in the process of design thinking, thereby affecting the generation of innovative design. This paper presents a design experiment that was conducted to study the effects of abstract knowledge to individual's design thinking and the cognitive behavior. In the experiment, subjects were given different levels of abstract knowledge, respectively from three aspects of the product: Function-Behavior-Structure, and then they were asked to complete a design task. Then analyze the role of abstract knowledge's form of cognition, and explore the mechanism of inspired design thinking further. The paper finally provides a theoretical basis for the design of innovative knowledge-oriented services.

Keywords Abstract knowledge · Innovative design · Cognitive behavior · Design cognition · Cognitive experiment

1 Introduction

Innovation is a key factor in the survival and development of enterprises, as well as the potential power in the future progress. Designer's innovative thinking depends on the knowledge and technical support. A lot of studies have shown that abstract knowledge could help designers generate innovative thinking [1], and the study scope of the explicit knowledge content is wide. Previous studies have mainly based on proprietary knowledge assisted innovation system and biological analogies on innovation concept. Innovation design based on patent is the use of large patent knowledge, combined with the corresponding heuristic rules, and assist designers making innovative creation [2]. And for biological analogies is mainly based on the biological structure, associated with functional attributes to our

Y. Zhou · Q. Xue (✉) · J. Hao · M.-x. Liu
School of Mechanical Engineering, Beijing Institute of Technology,
100081 Beijing, China

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_5

real-life problems, analogy biological characteristics, assisted innovation and product development [3]. These studies are trying to find an optimal mode of knowledge inspired thinking, in order to achieve quick access to design innovative concepts purposes. Both methods need to have the abstract knowledge processing, abandon the original form of restraint and evolve to the consciousness level, ultimately to think recreation. Studies show that the abstract knowledge has a significant influence on the novelty of innovative design results [4]. This paper puts forward the concept of the knowledge abstraction degree, focusing on the process that how abstract knowledge assisted a re-creation process. Studying the connotation and presence form of abstract knowledge and then decompose it according to the function-behavior-structure (FBS) method [5]. Using the cognitive experiment to explore how different abstraction knowledge will impact the design thinking.

2 Research on the Effect of Abstract Knowledge to Design Thinking

2.1 Design Thinking Model

The individual's cognitive mechanism is the key to explore the essence of innovation design rules, since human is the main part of the design. Studies have shown that the process of creative thinking was generated by the brain taking in task information from the outside, to turn it into working memory, and invoked previous design experience to inspire a new generation of memory, then after the integrated processed by the brain output the program information [6]. The role of different forms of creative thinking strategies in innovative design is shown in Fig. 1:

Logic thinking uses the methods of induction and deduction, through the way of composition and extraction to make the cognition in perceptual knowledge stage abstracted into concepts, make judgment using the concept and take reasoning according to a certain logic, resulting in new thinking activities.

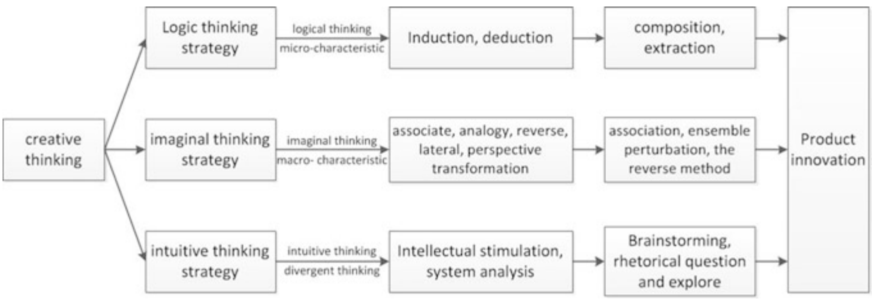


Fig. 1 Creative thinking strategies

Imaginal thinking is the use of imagination and experience associated with emotions, conduct thinking activity through visual and presentative expression mode. Both the nature of the abstract, but also to maintain the image of sensible species.

Intuitive thinking uses the unique intuition as an intermediary instead of logic, and perceive things essential directly from experience. This process requires both the existing design experience and domain knowledge to be applied flexibly, but also needs to capture in time and utilize emotional understanding and awareness generated effectively. Research shows that the most creative and valuable design ideas often hidden in these vague and non-quantitative emotional understanding and awareness [7]. Intuitive thinking is the soul of creative thinking and cognitive activity in the formation of product design and creative solutions

2.2 Definition of Abstract Knowledge

Product design depends on information, particularly the knowledge of the message contains. Knowledge is the basis of innovation. Whether enterprises have the abilities of creation, accumulation and sorting absorption or not will directly determine their competitive capacity [8]. Therefore, knowledge inspired innovative design has become the focus of academic research. Different knowledge properties and presentation forms have different effects on an inspired design. In traditional concept, abstract knowledge refers to the specific things and the relationships between abstract things reflect to one-sided consciousness [9]. In innovative design, abstract knowledge mainly performs as vagueness and generality in the program.

Marx believed that “Thinking is expanded according to the form of concrete-abstract-concrete”, therefore abstraction reduce the number of categories of data processing by creating an appropriate concept in the cognitive process, which save the data content, and seize the core information to create the conditions for the diffusion of knowledge. Rational abstraction beyond direct observation of the original knowledge background, high or low level of abstraction may lead to the loss of knowledge [10]. According to the versatility and ambiguity properties, the abstract knowledge is formed by extracting a series of common knowledge, namely through the integration to form a new kind of guidance universal knowledge. And extend the scope of semantic representation, increase one-sided abstract things by replacing specific concept into a more ambiguous expression.

2.3 Abstract Knowledge Assisted Design Innovation Policy

In product innovation design, knowledge plays a role in the designer’s innovative by a certain rule. In the present study, the method that abstract knowledge affect design thinking and innovative problem-solving, mainly based on TRIZ theory [11] and FBS [12] methods and so on (Fig. 2).

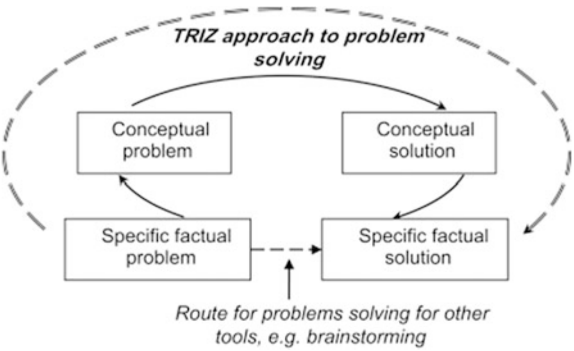


Fig. 2 TRIZ theoretical model

Theory of inventive problem solving (TRIZ) method started in 1946, led by G. S. A1tshuller's team, they analyzed 400,000 patents integrated with multi-disciplinary knowledge, which was condensed into abstract knowledge, and put forward a series of innovation principles [13]. TRIZ theory regard design as an innovative problem-solving process, and designer as a problem solver, achieved innovative product design by identifying, understanding, exploring, defining and solving problems [14]. Later, scholars have been digging through proprietary knowledge to further improve the TRIZ method, build design patent knowledge management system supporting for product innovation.

Howard believed that the engineering design combined with linear creative thinking would lead to creative design process [15]. He used Gero's Function-Behavior-Structure (FBS) model as the engineering design process model, combined with cognitive psychology analysis, generation, evaluation, linear innovation process model to describe an innovative thinking and design process which could help generate different types of design products. Solving process shown in Fig. 3,

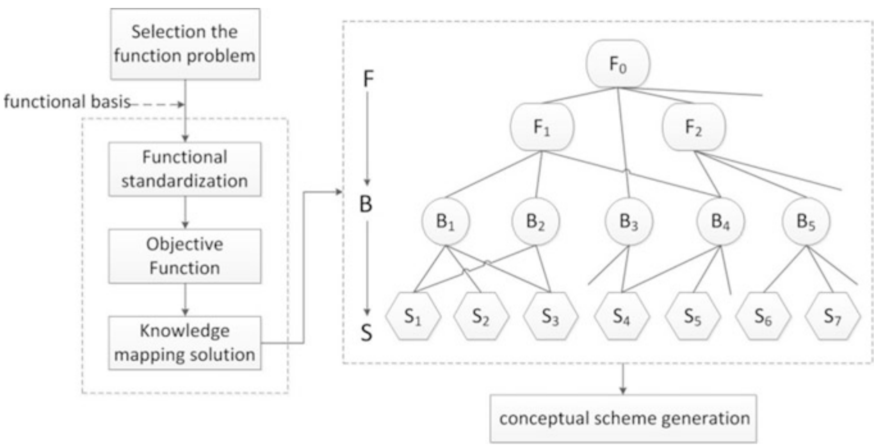


Fig. 3 FBS mapping solution process

which combines functional groups on the standardized function issue, bonded with the available knowledge resources, makes knowledge retrieval on target of function, thereby to obtain the corresponding concept by mapping program.

This framework, that involves the relation between function, behavior, and structure of a design, can be applied to any engineering discipline.

FBS framework has the following elements [16].

Function (F) The set of functions expressing the requirements and objectives that must be realized by the object.

Structure (S) Describe the components of the object and their relationships.

Behavior (B) The set of expected behaviors to fulfill the function F and the structure S.

These elements are connected in the framework by processes

The measure of the abstraction knowledge level is not formed a consistent concept in currently study, and the mechanism that abstract knowledge affects creative thinking is unclear. In this paper, divide the knowledge into the function, behavior, structure levels based on the theory, to explore the effects that different levels of abstract knowledge to innovation design results.

3 Experiments Based on the Levels of Abstract Knowledge

According to the research status of abstract knowledge, combined with individual cognitive-behavioral and cognitive rules, this paper explore the effects of different levels of abstract knowledge on individual cognitive and innovation design according to the appropriate innovative design experiments. In consideration of the subjects' design experience, the tests were divided into ordinary products design and the professional equipment design.

3.1 Participants

There were 12 participants, all university students formed with 3 females and 9 males, with an age range of 22–24 years old, without any design experience, and didn't participate in a similar experiment. According to the abstraction level, the participants were divided equally into three groups.

3.2 Design

The test was a single factor and three levels experiments. The dependent variable was the degree of knowledge. The knowledge was classified into non-indication, high degree of abstraction and low degree of abstraction these three parts according to the theory of FBS. The non-indication test proposed product function, indicated what tools need to be designed directly; high degree of abstraction test described the product behavior, pointed materials, methods and other characteristics from the functional perspective; and low degree of abstraction test further described the details of product structure combined with the current equipment pictures. The experimental independent variable was the properties of the design product, specific effectiveness measures were set forth below.

Through the knowledge of the above experimental materials, participants take the corresponding product innovative design combined with their daily experience. After the design, the participants need to make a description of the product design details through a retrospective approach.

3.3 Material

Test 1: Ordinary Products Design.

1. Non-indication test: Functional description

> Please design a tool to clean the inner wall of the cup

2. High degree of abstraction test: Behavior Description
   Fishbone diagram of cleaning tools design (man. machine, materials, method, environment) (Fig. 4)
3. Low degree of abstraction: Structural description
   Four pictures of glass cleaning tools (Fig. 5)

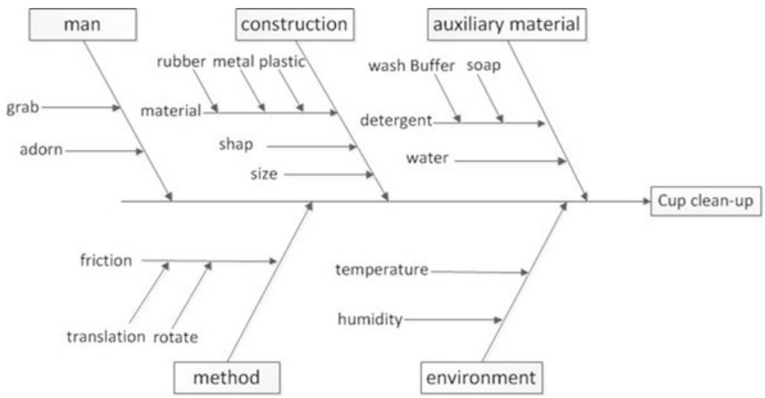


Fig. 4 Fishbone diagram of cleaning tools

Fig. 5 Glass cleaning tools

Test 2: Professional Equipment Design.

1. Non-indication test: Functional description

Please design a an efficient tennis picking device

2. High degree of abstraction test: Behavior Description
   Fishbone diagram of equipment to pick up the ball in tennis court (Fig. 6)
3. Low degree of abstraction: Structural description
   Two devices now commonly used (Fig. 7)

3.4 *Procedure*

- The subjects confirmed the nature of the experiment, and the task was to design a product.
- Subjects read different levels of abstraction knowledge according to the corresponding group, and then made a consideration of the design content.

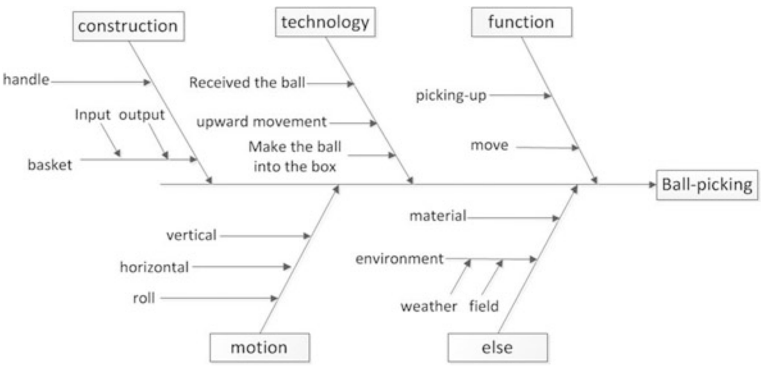


Fig. 6 Fishbone diagram of equipment

Fig. 7 Tennis picking device

- Subjects began to sketch the products drawing within 15 min, and indicated the product properties simultaneously.
- After the end of the design, the subjects need to briefly describe their design process and the characteristics of the product to the tester.

3.5 Results and Discussion

Indicator Selection. Based on the design results, three effective measures are proposed: quantity, quality, and variety of the ideas generated for innovative solutions.

Quantity (M_1) is the sum number of ideas generated.

Quality (M_2) is a measure of the idea feasibility, to create a framework for analysis.

Different indicators and weights can be assigned according to the importance of each function to the problem's solution, making weighting statistics according to the specific content of the design. For example, indicators of test.1 set as follows (Table 1).

Divided among three grades in excellent-8, medium-5, inferior-2. The formula:

$$M_2 = \sum_{k=1}^{n} s_k p_k \tag{1}$$

Variety (M_3) is used to measure the explored solution space during the idea generation process. Analyze all the innovative solutions of a single test and add the score according to the following criteria. The test.1 as an example (Table 2)

Table 1 Computing quality of an idea

Indicator(s)	Cleanliness	Cleaning speed	Operation flexibility	Cost
Weight(p)	0.3	0.1	0.1	0.2
Scheme one				

Table 2 Computing variety of an idea

No. (i)	Score (q_i)	Innovative solutions standard (f_i)
1	5	Ideas use the different washing principle
2	4	Ideas use the different structure
3	3	Ideas use the different material
4	1	Ideas use the different form

$$M_3 = \sum_{i=1}^{n} \sum_{k=1}^{m} f_{ik} q_{ik} \quad (2)$$

Result Analysis.

Test 1: Ordinary Products Design. Summarize all the experimental results, the score of quantity, quality, and variety for each participant shown in the Table (Table 3).

Analyze each indicator separately (Figs. 8, 9, 10).

From a design number, it could be found when there was no indication provided, the subjects would have more innovative designs. The result may be due to the provided product information limited the subjects' thinking to a certain aspect, thereby reducing the amount of designs. From indicators of quality, the effect of the product was the best while providing a high degree of abstraction knowledge to

Table 3 Summary statistics

Indicator	Non-indication test				High degree of abstraction test				Low degree of abstraction test			
Quantity	5	5	5	5	4	4	4	2	2	4	4	3
Quality	4.9	5	4.9	5.1	7.2	5.2	5	6.8	5	4.9	6.1	4.6
Varity	6	7	8.6	7.6	12	7.8	8.3	16	14.5	8.5	9	12.3

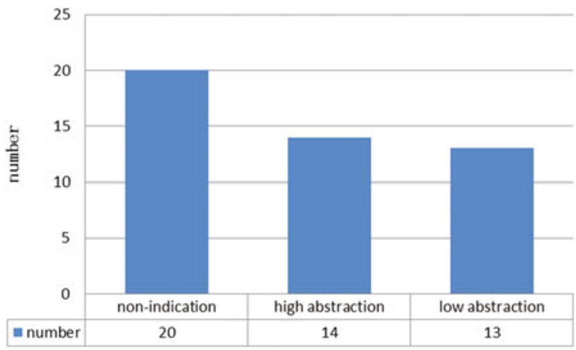


Fig. 8 Number of design results at different abstraction degrees

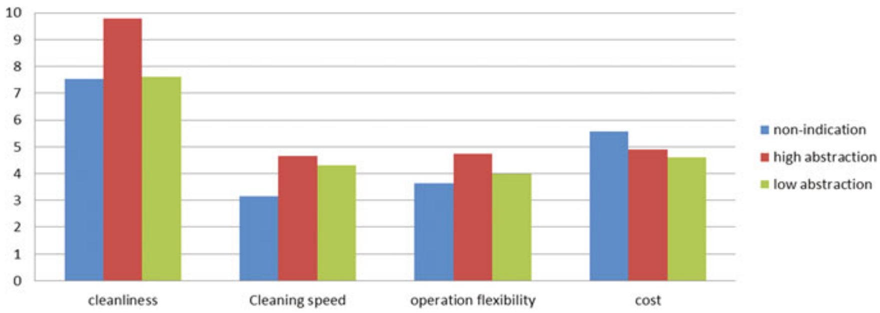
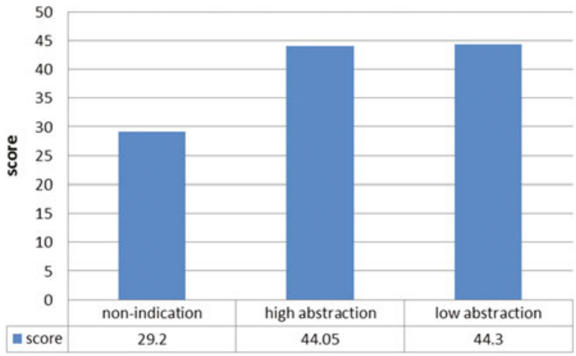


Fig. 9 Quality statistics at different levels of abstraction

Fig. 10 Variety statistics at different levels of abstraction



subjects except the costs indicator. Probably because the fishbone contained material, structure, and various auxiliary materials etc., a lot of available information helped the subjects stimulate the design creativity and consider the content more complete, so the design products have the highest quality. For the factor of product cost, it would be highest when did not offer any information to the subjects. Design verity has a correlation with quality. When providing knowledge to subjects, although the number of results reduced, but increased the product diversity.

Test 2: Professional Equipment Design. Quantity (M_1) the sum number of ideas generated.

Quality (M_2) the indicators: picking speed (0.3) operation efficiency (no manual) (0.3) cost (0.3) feasibility (0.1)

Variety (M_3) the innovative solutions standards: different picking principles, different functions, different structures and different materials.

Summary statistics and indicators statistical are shown as follows (Table 4, Figs. 11, 12 and 13):

Similar to the results of test.1, when there was no available knowledge provided, the participants designed most products. For the quality indicators, a high degree of

Table 4 Summary statistics

Indicator	Non-indication test				High degree of abstraction test				Low degree of abstraction test			
Quantity	3	3	3	4	4	2	2	2	2	3	1	2
Quality	6.2	5.4	5.6	5.2	5.8	6.5	7.6	6.5	6.7	6.1	6.2	6.5
Varity	9.7	5.3	11	7.8	10.3	14	10.5	11.5	14	8.7	16	17.5

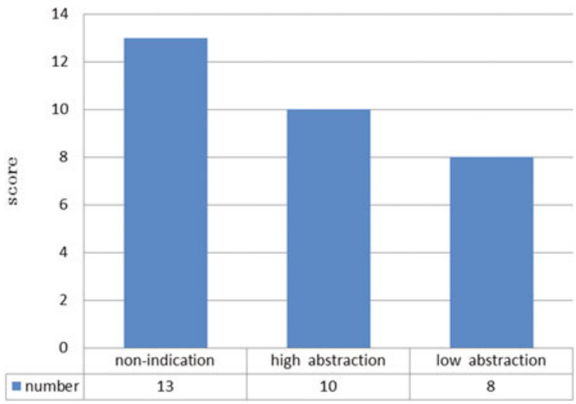


Fig. 11 Number of design results at different abstraction degrees

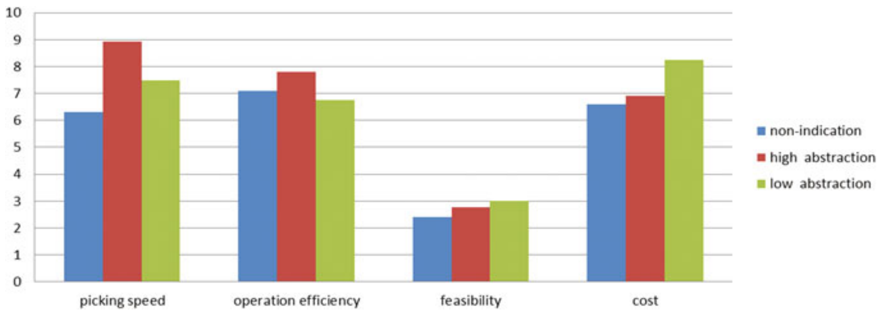


Fig. 12 Quality statistics at different levels of abstraction

abstraction knowledge would motivate subjects to design products better. And for diversity indicator, the design result was better when provided abstract knowledge than no-indicated knowledge. In contrast to test.1, because the content of test.2 was to design specialized equipment, so the design products would be more simple when there was no-indicated knowledge, as well as the product's cost was lowest.

For the quantity and variety, the results of the gradient effect were more obvious when the level of abstraction knowledge decreasing, which showed the amount of

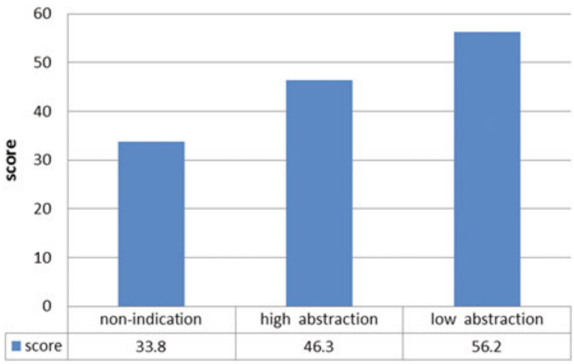


Fig. 13 Variety statistics at different levels of abstraction

designs gradually reduced but the variety enhanced. The reason was that the subjects need to take a reference by the knowledge provided rather than the daily experience could help them get results.

Discussion. After analyzing the results of these two design experiments, this paper find when there was no indication information provided to subjects, they would design largest amount of products for the test. However the design content presented higher repetition, lower quality and variety. When subjects were given a certain degree of abstract knowledge, it could help to stimulate innovative design, improved design quality and product diversity.

In terms of the information content, when participants were provided some aspects of the design concept directly, they would focus more on those parts of the design process, and the effects were better than the subjects explored the existing product design details by themselves. For example, in this study, the fishbone specified the structure, materials, methods, and other information directly. The product pictures also contained a series details, but the participants did not get a high level of attention, so the design quality was inferior to the effect of fishbone diagram.

As for the design product attributes, subjects took more reference from daily experiences when they were asked to design the ordinary product and the intuitive thinking appeared more obvious. On this occasion, different levels of abstraction knowledge had minimal impact on innovative thinking. While designing specialized equipment, the design was more dependent on knowledge and information provided, and the impact of different abstraction knowledge was more significant.

4 Conclusion

Product innovation design inevitably requires internal or external knowledge, which determines the important role knowledge plays on the design thinking process, thereby affecting generate innovative design solutions [17]. Abstract knowledge presents supports for individual innovative design. Also different levels of abstraction information affect the results of the product subject design. The mechanism of abstract knowledge is associated with individuals' cognitive behavior, which is inseparable from the experience and knowledge. Intuitive thinking comes more from the process of daily experience, while creative inspiration is a combination of knowledge and experience. Overall, the abstract knowledge will assist designers making interaction between working memory and new information, and then will contact and generate new innovative solutions.

Acknowledgments I would like to thanks to my tutors for their suggestions and those students participated in the experiments.

References

1. Goldschmidt, G., Sever, A.L.: Inspiring design ideas with texts. Des. Stud. **32**(2), 139–155 (2011)
2. Qingying, Q., Chi, X., Yue, J.: Auxiliary machinery products innovation system based on patent knowledge (in Chinese). Comput. Integr. Manuf. Syst. **19**(2), 354–361 (2013)
3. Vattam, S.S., Helms, M.E., Goel, A.K.: A content account of creative analogies in biologically inspired design. Artif. Intell. Eng. Des. Anal. Manuf. **24**(4), 467–481 (2010)
4. Chen, Q.S., Pendurti, S., Prasad, V.: The idea exposure paradigm in design idea generation. J. Cryst. Growth **18**(18), 93–102 (2007)
5. Gero, J.S., Kannengiesser, U.: The Situated Function-Behaviour-Structure Framework// Artificial Intelligence in Design '02, 373–391. Springer, Netherlands (2002)
6. Ma, G.L., Li, Y., Li, W.Q.: A process model and method of idea generation for conceptual design. Proc. Inst. Mech. Eng. Part B J. Eng. Manuf. **225**, 568–586 (2011)
7. Xu, Q.: Research on integrated design method for creative design of modern mechanical product. Trans. Chin. Soc. Agric. Mach. (2005)
8. Jiashao, P.: Knowledge management—business management topic in 21 century (in Chinese). Group Technol. Prod. Modernization **1**, 27–28 (2000)
9. Fuan, L.: On abstract and concrete (in Chinese). J. Lit. Hist. Philos. **2**, 94–97 (1981)
10. Xiaoyong, X., Huanfeng, Z.: Research on transfer mechanism between abstraction levels and drive types based on organizational knowledge–knowledge of Japanese management to divert Case (in Chinese). J. Intell. **11**, 52–54 (2006)
11. Ilevbare, I.M., Probert, D., Phaal, R.: A review of TRIZ, and its benefits and challenges in practice. Technovation **33**(s2–3), 30–37 (2013)
12. Gero, J.S., Kannengiesser, U.: The Situated Function-Behaviour-Structure Framework// Artificial Intelligence in Design '02, 373-391. Springer, Netherlands (2002)
13. Yan, L., Jie, W., Xianglong, L.I.: Creative thinking and computer aided product innovation design (in Chinese). Comput. Integr. Manuf. Syst. **12**, 1092–1096 (2003)
14. Biju, Y., Yan, L.I., Yan, X.: Design thinking research situation and development trend (in Chinese). Comput. Integr. Manuf. Syst. **19**(6), 1165–1176 (2013)
15. Howard, T.J., Culley, S.J., Dekoninck, E.: Describing the creative design process by the integration of engineering design and cognitive psychology literature. Des. Stud. **29**(2), 160–180 (2008)
16. Diertens, B.: The refined function-behaviour-structure framework. Eprint Arxiv (2013)
17. Howard, T.J.: Information management for creative stimuli in engineering design. University of Bath (2008)

How the Color Matching and Arrow Position Affect the Design of the Bus Stop Board: An Eye Movement Experimental Research

Chuanyu Zou, Na Lin and Yunhong Zhang

Abstract An eye movement study was conducted to study the visual factors which influencing the searching efficiency of bus stop board. 31 ordinary adults were measured to investigate the effects of the color matching, the positions of the arrow indicating bus route direction on the searching efficiency of the different sexual passengers at different age stage by Eye-tracking. The eye movement experiment took the bus stop names as material, set simulated bus route, and made series of bus stop boards with different color combinations and graphic designs. The result shows that the difference in search time and fixation times between the positions of the direction arrow of the bus name list is significant. When the arrow was below of the bus stop name list, it costed much less time for distribute alignment bus stop boards than for top alignment ones in searching destination bus stop; and when the bus stop name list was whether for the white background black text or green background white text, those bus stop boards which the direction arrow was below the bus stop name list had a significant advantage over those that the arrow was above in searching time and fixation times. As a conclusion, the obtained results could be a reference for design the bus stop board.

Keywords Bus stop board · Color matching · Arrow position · Searching efficiency · Fixation time · Eye-tracking

C. Zou
Research Group of Graphical Symbols and Way Guidance,
China National Institute of Standardization, Beijing, China

N. Lin · Y. Zhang (✉)
Human Factor and Ergonomics Laboratory,
China National Institute of Standardization, Beijing, China
e-mail: zhangyh@cnis.gov.cn

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_6

1 Introduction

Bus station board is an important guiding vehicle on contemporary era. If bus station board inappropriate, it will cause low searching efficiency for the public, and will result in traffic jams and influence passengers' security. How to design bus station board coincide with passengers' cognition habits and make their travel more conveniently, which become the question that ergonomic researchers focus on.

In previous study, data collection of board visual design showed that small font, a single color [1] and not comprehensive indicate information were inconvenient for passengers to check [2]. Even there were no driving directions and station name, which is difficult to known his/her own location [3]. Han Li made a summarize about the research of Baoding bus station board visual guiding system design. Bus station board was consist of color, layout graphology, layout figure, material, display method and definite distance [4]. The standard analysis of bus station board visual guiding elements could transmit public traffic information quickly and without error. For color requirements, as we all known, visual is the start of organ to form information and then know them. Because color has emotion convey magic and visual impact force, it become the chief visual element considering. GB/T 5845.3 stipulated color standard of bus station board: The number of color should not beyond 4 kinds, color collection should not influence achromate's and tritanope's recognizing and the backboard should be white. For layout words requirements, in guiding system design, normative words ensured the veracity of instantaneous guiding information. GB/T 5845.3 [5] stipulated layout graphology standard of bus station board: Chinese characters, numbers, bopomofo and English letters should be showed in Fangsong type and bold face letter, typeface should regular and distinct, well arranged and distributed according to simplified character by official formal announcement and bopomofo spelling rules. The size of graphology should not less than 10 mm $\times$ 10 mm. National minorities should use Chinese characters, bopomofo and their own characters [5]. Based on people's searching reaction and correction rate, Xie Hui and his partners found that passengers' visual searching performance would be improved by increasing prominent that arrows and colors were better than big font [2]. Chen Zhengzhe, who wrote improvement design of bus station board indicate information, had a study of single/double direction route could not be recognized, location hint of return bus station, sketch map of transfer bus station route and moving towards relative to the city directions. He put some suggestions about the improvement of the information on bus station board designing [6]. Based on the psychology and visual perceiving, Zheng Yi had a research on static station board design, then put forward a schedule and new conception about design style of font, chart and imaging [7].

Recently, problems of bus station board design was collected by several method, including questionnaire, behavioral tests and theoretical analysis, they were used in the research, but not reach the purpose of bus station board information processing. With the development of eye movement theory and tracking technology, eye movement is widely used in different field, such as cognition research of web sites, ads, graphics. Two basic phenomenon were eyes' movement (saccade) and fixation (relative to static). Keep eyes static relatively was called fixation, fixation time was 200–300 ms, and people get reading information in this way. Saccade, also called eyes jump, eye balls jump from one point to another, movement was controlled by central nervous system regularly at will, it showed suddenly change of fixation point or fixation direction. Because of the limitation of anatomical structure of retina and visual acuity of peripheral scope of central fovea. Saccade was necessary in reading. When someone was reading, he/she divided a line into three parts: central fovea region (2° view scope of central vision), parafoveal region (2°–5° view scope of central vision) and peripheral region (all scopes outside of parafoveal). There was a high visual acuity of central fovea, central parafoveal was lower, peripheral scope was the lowest, therefore, when someone was reading, he/she had to move eyes for presenting new content into central fovea. Saccade was reflection of time planning and executing. Saccadic latency, was the target location of coding visual scope and trigger time of eye movement, the average of data was about 175–200 ms, and different with task demanded. Saccadic duration was eye movement time of two fixation points, it was relevant to distance of saccade, the longer distance, the more time used. Study found that 2° visual angle of saccade in reading at about 30 ms, 5° visual angle of saccade in scene perception spent 40–50 ms. In reading eye movement research, according to the demands, researchers confirmed the target region to analyze, that was the area of interest (AOI). AOI was a single character or one word, maybe further big region, such as a phrase or a sentence [8].

Wang Keqin pointed there were significantly difference on characters and patterns on fixation time, fixation count, and average fixation duration and pupil area in the eye movement study about print ads [9]. Bus station board was transmitted processing by graph and characters, different reading/watching material caused different cognition process machining, they could reveal interior machining effectively by eye movement. In this paper, we try to use the eye-tracking technology to study the influence of various highlighted visual interfaces on the bus stop boards' visual search efficiency and collect experimental data for improving the humanistic design level of the bus station board.

2 Method

2.1 Experiment Design

Within-subject and between-subject crossed four factors design was conducted to investigate the effects of the color matching, the positions of the arrow indicating bus route direction on the searching efficiency of the different sexual passengers at different age stage. A 2 (the positions of the arrow indicating bus route direction: above vs. below the bus stop name list) × 2 (the color matchings of the bus stop name list: the white background black text or green background white text) within-subject factorial design with two additional control condition were used in this experiment. The additional independent variables were: age (young vs. older) and gender (male and female). In this experiment, the search time and fixation times were measured by eye-tracking to assess and compare which type of bus stop board design is better.

2.2 Participants

Thirty-one ordinary adults from 20 to 66 years old (14 male and 17 female, mean age = 40.42, standard deviation of age = 15.83) were recruited and paid to participate in the experiment. All subjects were divided into two groups, one group was young from 20 to 44 (9 male and 8 female, mean age = 27.18, standard deviation of age = 5.34); the other group was senior from 45 to 66 (8 males and 6 female, mean age = 56.50, standard deviation of age = 5.96). All had normal or corrected-to-normal visual acuities and healthy physical conditions, without ophthalmic diseases. They did not have any history of neurological and mental diseases. And the participants were divided into two groups, one group were younger group which included nine male and eight female, and the other group were senior group which included eight male and six female.

2.3 Stimulus

The eye movement experiment took the bus stop names as material, set simulated bus route, and made series of bus stop boards with different color combinations and graphic designs. From Beijing city bus station name library selects the bus station name and form 11 virtual circuit made of green bottom mispronounced character matching bus stop pictures, of which each bus stop containing 46 bus station. And the alignment patterns of the bus station names list is top alignment. One of them was as the practice trial. Each route has 4 target station to search which represent four experimental level and the position effect of the target stations were balanced. 40 search trials were randomly divided into 7 groups, each group included of 4–8 trials. The 4 target search trials of the same line were evenly distributed in each group. The 4 bus stops of the same line were not adjacent for each other. Each experimental level has designed 10 standard trials to emphatically analyze which there were 20 stations between the target station and the current station, and the rest are filled trial. The standard trial of each experiment level was evenly distributed in each experimental group, and their locations were not in the first or in the last of each experimental group. The positions of the standard trials in each experimental group were pseudo randomly arranged.

As generating the virtual circuit bus stop, the station name is selected to avoid the special place (such as Zhongnanhai, Badaling station) and the very familiar

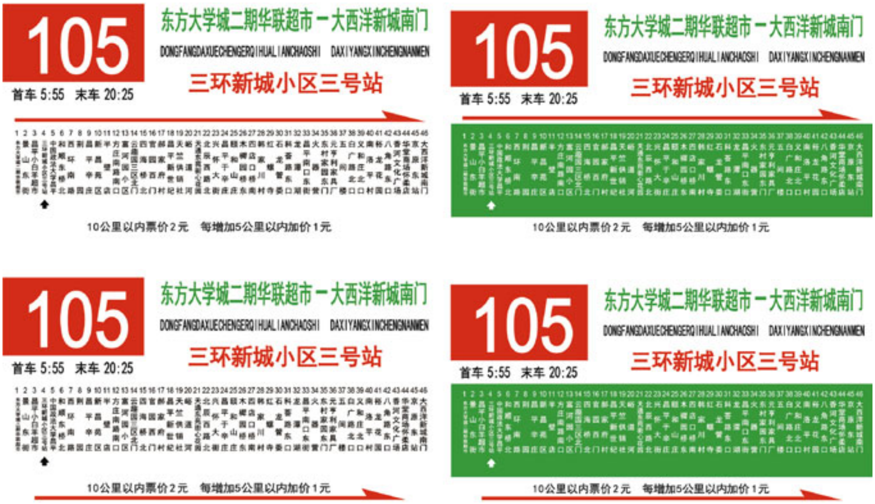


Fig. 1 The stimulus example

place (such as Sanyuan Bridge). And the station names was arranged by pseudo random and the station names with different words (including 2 words, 3 words, 4 words, 5, 6 words, 7 words) was evenly distributed throughout the virtual circuit. The station order was arranged by considering the geographical location to avoid a clear violation of the common sense. The locations of the initial stations were distributed random. The station name with 2 words was not as the initial station and terminal station. Target stations were selected by taking into account the word number, the location and other factors. The terminal station are not the target station. The image resolution was 1284 × 812 and the image example was as follow (Fig. 1).

2.4 *Apparatus*

The eye-tracker was SMI iView X RED made by SMI Corporation and its sampling rate is 60 Hz. And DELL Latitude D620 was as the experimental host with 2116 Hz core frequency, 2G memory card, independent graphics card. The host was connected to two monitors, one monitor was 19 in. display which was presented stimuli for subjects and subjects responded through the mouse and keyboard, while the eye tracker below the display recorded the subjects' eye movement data; the other display device showed the subjects' eye movement for the objects during the experimental process. And the resolutions of all the display terminals were 1280 × 1024 and their refresh rates were 60 Hz.

2.5 Procedures

Experiments were conducted in a quiet and bright environment which simulated outdoors condition. It was installed in the laboratory in the Institute of Human Factors and Ergonomics in China National Institute of Standardization. The eye movement experiment on the design of bus stop board was conducted in the Institute of Human Factors and Ergonomics lab in China National Institute of Standardization. After arriving at the laboratory, participants signed the informed consent and completed a general survey about their demographic information. The participants were asked to sit into the simulator to get ready for the test. After that, the participants were required to complete the visual search task by simulating reading the bus stop boards. The experiment are vertically arranged in the height adjustable special experimental table. Before the experiment, the height of the table and display position were adjusted to make the participants' eyes and display center on a line. The viewing distance is about 60–70 cm, and the head tracking range is 40 * 40 cm. Before viewing the bus stop board, an eye movement calibration was done to ensure data precisely. Then, participants read experimental instruction and did some exercises to ensure participants understand the instruction and conduct the visual search task correctly. If participants don't know the experimental process, then let him/her do practice again. After practice, the experimenter would validate simply the search strategy which the subject used, but do not make any judgments and tips. Then, they entered the formal experiment. During viewing the bus stop boards, the participants' eye movement data were recorded. There was three to five minutes resting between each group. After viewing all the pictures, the participants were required to tell which search strategy he/she used. Each participant spent about one hour finishing the experiment. The searching time and fixation times were recorded by SMI eye-tracker as efficiency indexes when the participants reading different designed bus stop boards.

2.6 Data Analysis

According to the experimental design, the data of standard trials were analyzed through the BeGaze software. The standard trial was divided three regions of interest which were current station, target station and the middle region between the two stations. The time that first enter the target station was as the search time, and the fixation times in the middle regions were as the fixation times. They were as indexes for further data analysis. According to the behavior data and the search strategy that subjects used, the first fixation time data were corrected, and the trials with incorrect behavior data were excluded. The eye movement data of all standard trials were exported from the SMI BeGaze software and analyzed by IBM SPSS 20 Statistics software (IBM-SPSS Inc. Chicago, IL). The repeated-measure ANOVA was applied to analysis the data of searching time and fixation times.

3 Results

3.1 *The Means of First Fixation Time to Enter the Target Station Region and Fixation Times for Each Condition*

Eye-tracking test was conducted to compare the effect of the Bus Stop Board's Design with different color matching and route direction. The first fixation time to enter the target station region and fixation times data were analyzed. A 2 × 2 × 2 × 2 mixed-measure ANOVA was conducted to evaluate age group (above 40 years old and below 40 years old), gender group (male and female), color matching (white background black text or green background white text) and route direction (above the bus stop names and below the bus stop names). The first fixation time to enter the target station region and fixation times for each condition are summarized in Table 1.

3.2 *Repeated-Measure ANOVA of First Fixation Time to Enter the Target Station Region*

With regard to the search time that is first fixation time to enter the target station, a repeated-measure ANOVA was applied to the search time data of the different conditions, a significant main effect of color matching was found ($F(1, 30) = 11.20$, $p < 0.01$), a significant main effect of route direction was found ($F(1, 30) = 41.92$, $p < 0.001$), a significant main effect of age group was found ($F(1, 30) = 6.99$, $p < 0.05$), and a significant interaction was found, $F(1, 15) = 9.29$, $p < 0.01$. The simple effect analysis showed that when the route direction was above the bus stop names the search time of the white background black text condition was remarkably

Table 1 The mean search time and mean fixation times of the four conditions. Standard errors are given in parentheses

	Conditions	Mean first fixation time	Mean fixation times
Color matching	White background black text	5.29E3(594.27)	19.17(2.04)
	Green background white text	8.27E3(857.46)	23.25(2.78)
Route direction	Above the bus stop names	9.37E3(916.83)	26.23(2.96)
	Below the bus stop names	4.20E3(433.72)	16.18(2.05)
Gender group	Male	6.96E3(796.44)	18.42(2.93)
	Female	6.61E3(885.19)	24.00(3.26)
Age group	Above 40 years old	8.36E3(885.19)	26.75(3.26)
	Below 40 years old	5.21E3(796.44)	15.67(2.93)

Table 2 Repeated-measure ANOVA results of first fixation time to enter the target station region

	SS	df	MS	F value	Sig. (two-tailed)
Age group	3.01E8	1	3.01E8	6.99	0.013
Gender group	3.76E6	1	3.76E6	0.09	0.770
Color matching	2.69E8	1	2.69E8	11.20	0.002
The positions of route direction	8.13E8	1	8.13E8	41.92	0.000
The positions of route direction × color matching	8.95E7	1	8.95E7	10.90	0.003

shorter than that of the green background white text condition ($p < 0.05$), and when the route direction was below the bus stop names the search time of the white background black text condition was remarkably shorter than that of the green background white text condition ($p < 0.01$); and whether the white background black text condition or the green background white text condition, the search time of the condition that the route direction was below the bus stop names was remarkably shorter than that of the condition that the route direction was above the bus stop names ($p < 0.001$). The planned comparisons revealed that the search time of the condition that the route direction was below the bus stop names was remarkably shorter than that of the condition that the route direction was above the bus stop names ($p < 0.001$), and the search time of the white background black text condition was less than that of the green background white text condition ($p < 0.01$). And the planned comparisons revealed that the search time of younger condition was remarkably shorter than that of senior condition ($p < 0.05$), but the main effect of gender group was not significant ($p > 0.05$) (see Table 2). It means that the search time was affected by age, but not by gender.

3.3 Repeated-Measure ANOVA of Fixation Times in the Middle Station Region

With regard to the fixation times, a repeated-measure ANOVA was applied to the fixation times data of the different conditions, a significant main effect of route direction was found ($F(1, 30) = 15.07$, $p < 0.01$), a significant main effect of age group was found ($F(1, 30) = 6.38$, $p < 0.05$), and a significant interaction was found, $F(1, 15) = 5.72$, $p < 0.05$; and the main effect of color matching was critical significant ($F(1, 30) = 3.64$, $p = 0.067$). The simple effect analysis showed that when the route direction was above the bus stop names the fixation times of the white background black text condition was less than that of the green background white text condition ($p = 0.059$); and whether the white background black text condition or the green background white text condition, the fixation times of the condition that the route direction was below the bus stop names was remarkably

Table 3 Repeated-measure ANOVA results of fixation times in the middle station region

	SS	df	MS	F value	Sig. (two-tailed)
Age group	3719.34	1	3719.34	6.38	0.018
Gender group	942.70	1	942.70	1.62	0.215
Color matching	503.76	1	503.76	3.64	0.067
The positions of route direction	3063.24	1	3063.24	15.07	0.001
The positions of route direction × color matching	774.83	1	774.83	5.72	0.024

less than that of the condition that the route direction was above the bus stop names ($p < 0.05$). The planned comparisons revealed that the fixation times of the condition that the route direction was below the bus stop names was remarkably shorter than that of the condition that the route direction was above the bus stop names ($p < 0.01$), and the fixation times of the white background black text condition was less than that of the green background white text condition, but the difference did not reach the significant level. And the planned comparisons also revealed that the search time of younger condition was remarkably shorter than that of senior condition ($p < 0.05$). The results indicated that the main effect of gender group was not significant ($p > 0.05$) (see Table 3). It means that the search time was not affected by gender.

4 Discussion

Eye movement study was used in web-pages, books, city landscape designing in recent years. There were various design elements on books' layout, researchers had found out that different design elements influenced reading speed and reading efficiency [10]. In the eye movement observation indexes, by the task of searching performance, one of the main index was the fixation time. Fixation time reflected the material disposing ability and processing ability of participants [10]. More fixation time indicated low efficiency of searching, it may be caused by wrong format [11]. Hendrickso used the fixation times as one of the indexes, and assessed different type menu projects to choose performance in early 1989 [11]. Another index was the first fixation time to enter target region. The longer time was used when entering target region for the first time, the longer time was used to searching the target, the lower efficiency for searching activities.

In the study, it was highly accordance about the first fixation time to enter the target region and the result of fixation times, no matter direction arrows located in upper or below of the station name lists, compared black text on white background with white text on green background, searching time was shorter and central fixation points was less as using black next on white background, when direction arrows located under the station name lists, searching efficiency was lower as using white text on green background. In addition, searching time was shorter and central fixation points was less when direction arrows located in under the station name lists. Compared with direction arrows located under station name lists, no matter black text on white background or white text on green background, they had the same phenomenon. It means that, by contrast, at the condition of black text on white background, processing difficulty of material was lower, there were less interference for users, and attention distribution and cognitive processing were more efficient, so they could save passengers' cognitive resources. Direction arrows located under the station name lists. As a sign of judging driving route direction, direction arrows were not the final searching purpose, therefore, when direction arrows under the name lists, it could be a hint when necessary, and it could not absorb more passengers' attention to decrease searching efficiency, it helped to save passengers' cognitive resources and improve searching efficiency.

In conclusion, adopting the color of black text on white background and composing style of direction arrows under the station name lists, they were good method to re-duce searching interference and improving searching performance.

5 Conclusion

The research indicated that the difference of the search time to target station between white background black text and green background white text color matching conditions was significant, and the difference of the search performance between different positions of route direction was significant, the location below the name lists was more efficient. It also revealed that ageing effect was found in the bus station searching performance. The study results offered data support for bus stop humanistic design. In follow-up work, researchers will continue researching bus station board and relevant facilities from cognitive processing angle to improve the bus station board design. Also, we will research for special group such as ageing people, and finally keep the design of bus station board better suit people's needs.

Acknowledgments The authors would like to gratefully acknowledge the support from the National Key Technology R&D Program of the Ministry of Science and Technology (2012BAK28B03, 2014BAK01B03), China National Institute of Standardization through the "special funds for the basic R&D undertakings by welfare research institutions"(522014Y-3344).

References

1. Ding, H., Zhu, W., Guo, J.: Investigation and Analysation of Shanghai bus station board. Urban Public Traffic **12** (2009)
2. Zhang, Z.: Visual design research of urban bus station board. Space Arts **1** (2014)

3. Xie, H., Fang, W., Ding, L.: Cognitive prominence ergonomic experiment research of bus station board. Int. Conf. 12 (2007)
4. Han, L., Rong, N.: Baoding visual guiding system design research of bus station board. Hebei University (2011)
5. GB/T 5845.3.: The 3rd part of urban public traffic sign: station board and guideboard of public transportation(bus, electric bus). Standards Press of China (2009)
6. Chen, Z.: Improvement design of bus station board indicative information. Highw. Adm. **3** (2014)
7. Zheng, Y.: Several noticed aspects of static bus station board design. Mod. Arts Des. **5** (2005)
8. Yan, G., Xiong, L., Zang, C., Xu, L., Cui, L., Bai, X.: Main eye movement indexes recitation in reading study. Psychol. Sci. Evolvement. **21**(4), 589–605 (2013)
9. Wang, K., Gao, X.: Eye Movement Research of Different Layout Style in Print Ads. Shanghai Normal University (2005)
10. Xu, J.: Eye Movement Study in Psychological Field. Beijing Union Uni. J. **7** (2013)
11. Yan, B.: Application of eye movement study on website usability test. Beijing Uni. Post Telecommun. (2011)

Measuring the Amplitude of the N100 Component to Predict the Occurrence of the Inattentional Deafness Phenomenon

Eve Florianne Fabre, Vsevolod Peysakhovich and Mickaël Causse

Abstract In the field of aviation, a significant amount of accidents are attributable to a phenomenon called inattentional deafness, defined as "the propensity to remain unaware of unexpected, though fully perceptible auditory stimuli such as alarms". The present study aimed at testing the impact of cognitive load on the perception of auditory information unrelated to the piloting task at stake in an ecological flight context. Pilots had to perform simultaneously a piloting task (i.e., approach and landing) in a A320 flight simulator and a passive auditory oddball task, with standard (80 %) and deviant (20 %) tones played. Lower N100 amplitudes were found in response to deviant tones when the piloting task was associated with a high cognitive load than a low cognitive load, demonstrating that cognitive load disrupts the perceptual processing of auditory stimuli, which is likely to trigger inattentional deafness in pilots.

Keywords Human factors · Aeronautics · Inattentional deafness · N100 · Passive oddball task

1 Introduction

In the field of aviation, a significant amount of accidents are due to a phenomenon called inattentional deafness (i.e., the propensity to remain unaware of unexpected, though fully perceptible auditory stimuli such as alarms). For instance, in the famous crash of Eastern Air Lines Flight 401 in the Everglades, the pilots were obsessed by a burnt-out landing gear indicator light and did not perceive the auditory alarm indicating the disengagement of the autopilot [1]. This crash among

E.F. Fabre · V. Peysakhovich · M. Causse (✉)
Département Conception et Conduite des véhicules Aéronautiques et Spatiaux,
Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse, France
e-mail: Eve.FABRE@isae.fr

M. Causse
Ecole de Psychologie, Université Laval, Québec, Canada

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_7

others (e.g., First Air flight 6560), suggests that an unexpected event associated with a high cognitive load is likely to increase the occurrence of inattentional deafness.

Various experiments investigated the inattentional deafness phenomenon in the laboratory [2–4]. To study the inattentional deafness phenomenon in the context of aeronautics, Giraudet and colleagues [4] used an aviation-like task in which the level of cognitive load varied. This aviation-like task was coupled with an alarm detection task (i.e., active oddball task, for a review see [5]). The results showed that an increase in cognitive load leads to a decrease in alarm detection performance and is associated with a decrease in P3b amplitude. However, the amplitude of the N100 component was not affected. Taken together these results demonstrate that tasks associated with high level of cognitive load may enable the perceptual processing of the alarms (i.e., no modulation of the N100 amplitude) but may disrupt cognitive processes, which is likely to trigger inattentional deafness. In this study, the piloting task was very simplified and not totally ecological. Moreover, pilots were explicitly informed that they had to attend auditory information in order to complete the alarm detection task, which may have artificially enhanced the sensitivity to the auditory alarms.

The present study aimed at testing the impact of cognitive load on the perception of auditory information unrelated to the piloting task in an ecological flight context. Fifteen pilots performed two approaches/landings in the A320 flight simulator of ISAE-Supaéro. In the first flight scenario, the approach/landing procedure was associated with a low cognitive load (i.e., Ceiling And Visibility OK, normal functioning of the flight instruments), while the second flight scenario was more complex with a covered weather and a malfunction of flight instruments to manage (i.e., high cognitive load). In addition to the piloting task, pilots had to perform a passive auditory oddball task (i.e., participants had not to react to the tones), with standard (80 %) and deviant (20 %) tones played in the flight simulator. Brain electrophysiological measurements (i.e., ERPs) in response to deviant sounds were measured in both scenarios. An important amount of studies demonstrated that both the N100 component and the P3b component indexed the processing of auditory stimuli. The N100 component is a negative-going ERP, peaking in adults between 80 and 120 ms after the onset of a stimulus, and distributed mostly over the frontal-central region of the scalp [6]. It indexes the perceptual processing of the stimulus and was also found to be larger in response to non-targets and infrequent stimuli [7]. The P3b is a positive-going ERP, observed in a time window between 300 and 600 ms at the central-parietal region of the scalp and known to reflect the occurrence of cognitive and attentional processes (for a recent overview, see [8]). If cognitive load disrupts the perceptual processing of infrequent/deviant auditory information at an early stage, we predicted to observe lower N100 amplitudes in response to deviant tones in the high-load scenario compared to the low-load scenario. While in the case it disrupts later attentional and cognitive processes, we may observe lower P3b amplitude in the high load scenario than in the low load scenario.

2 Materials and Methods

2.1 Participants

Sixteen healthy participants (*MAge* = 32 years old, *SD* ± 10), all native French speakers, participated in this study. They were recruited on the ISAE-supaéro campus. All were right-handed (as assessed by the Edinburgh Handedness Inventory, [9]) private pilots with a valid Private Pilot License. They had normal auditory acuity and normal or corrected-to-normal vision. None of the participants reported a history of prior neurological disorder. All participants were informed of their rights and gave written informed consent for participation in the study, according to the Declaration of Helsinki. The research was carried out fulfilling ethical requirements in accordance with the standard procedures of the University of Toulouse.

2.2 Material

Oddball Sounds. One of two 50 dB (SPL) tone types lasting 100 ms was randomly played. The tone was either standard (frequency = 1900 Hz, $p = 0.9$) or deviant (frequency = 1950 Hz, $p = 0.1$).

The Flight Simulator. The experiment took place in the PEGASE simulator (i.e., an A320 simulator; see Fig. 1) at the ISAE-Supaéro. PEGASE simulator lies on pneumatic jacks that enable to recreate realistic accelerations. It comprises: a cabin equipped with various types of display existing in aircraft cockpits, a 3-axis platform for movement restitution, 3D display of the outside world, the whole set-up (i.e., management, simulation and graphics) is managed by 18 PC type microcomputers.

Fig. 1 Illustration of a participant in the flight simulator

2.3 Procedure

First, participants were sat on a chair while the EEG cap and the electrodes were placed on their hands. Participants were then invited to take the commandant's place in the simulator. They performed a flying training session of 15 min in the simulator. They were then given the flight instructions. They were asked to perform two flights from Bordeaux (France) to Toulouse (France) of 20 min each and were informed of the runway they would have to land on. During the flight, they were continuously given ATC instructions they had to follow. They were informed that they could not reply to these instructions or communicate with the air controllers. In the low load flight, the cognitive load associated with the flight was low with Ceiling And Visibility OK (CAVOK, no clouds below 5000 ft above aerodrome level visibility is at least 10 km no current or forecast significant weather such as precipitation, thunderstorms, shallow fog or low drifting snow) and normal functioning of the flight instruments. In the high load flight, the visibility was very low making visual flight impossible, and the flight instruments provided fluctuating information (i.e., a fluctuant δ was added to the correct flight parameters making the piloting task more complex). Half of the participants started with the low load flight scenario, while the rest of them started with the high load flight scenario. Participants were told that two types of tones would be played along the flights but they would not have to do anything special in response to these sounds.

2.4 Electroencephalography

EEG was amplified and recorded with an ActiveTwo BioSemi system (BioSemi, Amsterdam, The Netherlands) from 32 Ag/AgCl active electrodes mounted on a cap and placed on the scalp according to the International 10–20 System (FP1, FP2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, CP5, CP1, Cz, CP2, CP6, P7, P3, Pz, P4, P8, T7, T8, PO3, PO4, O1, Oz, O2) plus two sites below the eyes for eye movements monitoring. Two additional electrodes placed close to Cz, the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode, were used to form the feedback loop that drives the average potential of the participant as close as possible to the AD-box reference potential. Electrode impedance was kept below 5 kΩ for scalp electrodes, and below 10 kΩ for the four eye channels. Skin-electrode contact, obtained using electro-conductive gel, was monitored, keeping voltage offset from the CMS below 25 mV for each measurement site. All the signals were (DC) amplified and digitalized continuously with a sampling rate of 512 Hz with an anti-aliasing filter with 3 dB point at 104 Hz (fifth-order since filter); no high-pass filtering was applied online. The triggering signals to each word onset were recorded on additional digital channels. EEG data were off-line re-referenced to the average activity of the two mastoids and band-pass filtered (0.1–40 Hz, 12 dB/octave), given that for some subjects the low-pass filter was not effective in completely removing the 75 Hz artifact. Epochs were time locked to the onset of the tones and extracted in the interval from −200 to 800 ms. Segments with excessive blinks and/or artefacts (such as excessive muscle activity) were eliminated off-line before data averaging. The lost data (due to artefacts) were equal to 45 %. A 200 ms pre-stimulus baseline was used in all analyses.

3 Results

A 2 × 3 [Scenario (Low Load, High Load) × Electrode (Fz, Cz, Pz)] repeated measures ANOVA was conducted to assess mean amplitudes in the 100–180 ms time window. The analysis revealed a main effect of Electrode [$F(2, 15) = 7.79$, $p < 0.01$, $\eta p2 = 0.34$], with greater negativities at Fz ($M = -1.47$ µV, $SD = 1.43$) than at Cz ($M = -1.01$ µV, $SD = 1.29$) and Pz ($M = -0.82$ µV, $SD = 1.35$), but significant differences between Cz and Pz ($p < 0.50$). The analysis also revealed a main effect of scenario [$F(1, 15) = 4.72$, $p < 0.05$, $\eta p2 = 0.24$], with greater negativities observed in response to infrequent tones in the low load scenario ($M = -1.55$ µV, $SD = 0.99$) compared to the high load scenario ($M = -0.65$ µV, $SD = 1.43$). However, the Scenario x Electrode interaction was not significantly different [$F(2, 30) = 0.56$, $p = 0.58$, $\eta p2 = 0.04$]. See Fig. 2 for grand average ERP waveforms.

4 Discussion

The present experiment aimed at testing the "permeability" of the pilots to unrelated auditory stimuli depending on the cognitive load of the piloting task at stake. Subjective measurements revealed that participants were more cognitively charged in the high load flight scenario than in the low load flight scenario, confirming that high load scenario generated greater cognitive load in participants than the low load scenario. Moreover, lower N100 amplitudes were found in response to infrequent tones in the high load scenario than in the low load scenario. The N100 component was found to index the perceptual processing of auditory stimuli but also the early allocation of attentional resources to these stimuli [6]. The electrophysiological results of the present study demonstrate that the perceptual processing of auditory stimuli (i.e., such as alarms) may be disrupted in pilots performing high cognitive load piloting tasks. We conclude that pilots are less likely to process auditory information unrelated to the task at stake in a complex situation (i.e., malfunction of flight instruments), which may lead them to become "deaf" to auditory alarms.

These results are in line with the results of a recent study [10] that showed that when the cognitive workload was high, visually impaired pilots showed lower N100 responses to infrequent auditory stimuli compared to when the cognitive

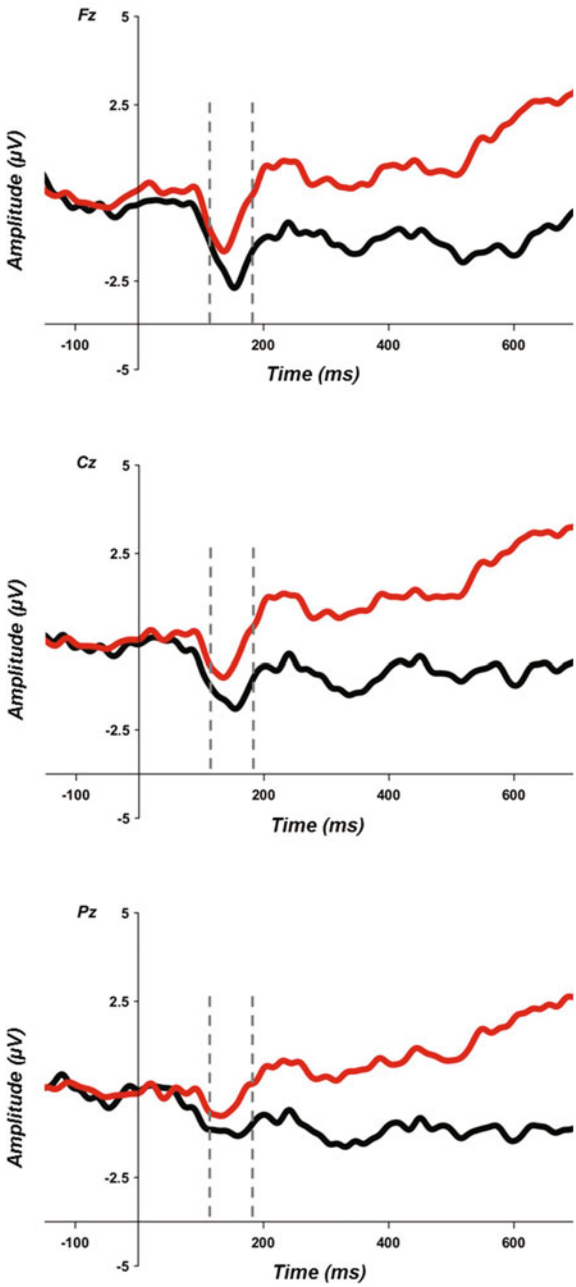


Fig. 2 Grand average ERP waveforms at Fz, Cz and Pz electrodes for infrequent tones in the low load condition (*black line*) and in the high load condition (*red line*)

workload is low. In the contrary, using an active oddball task, Giraudet and colleagues [4] found no modulation of the N100 amplitude, but a decrease in P300 amplitude in response to infrequent tones when the cognitive load associated with the task increased. This difference in results is likely to be due to the nature of the oddball task that was active in the study of Giraudet and colleagues [4] and passive in the present study. Using an active oddball task is interesting in that it enables to measure behaviorally the inattentional deafness phenomenon, by counting the amount of missed hits. However, as mentioned in the introduction section, it may also modify the attentional focus of the pilots, because it increases artificially the attention toward auditory information (i.e., the pilots are expecting the occurrence of the auditory stimuli) which may modify the way auditory stimuli are processed. While Giraudet and colleagues [4] concluded that an increase in cognitive load may lead to disrupting the processing of auditory alarms at a late stage (i.e., cognitive process), we argue that when pilots do not attend to the auditory stimuli, an increase in cognitive load may disrupt the processing of these auditory stimuli at an early stage (i.e., perceptual). Future studies should confirm this hypothesis.

Various solutions can be used to prevent the occurrence of the inattentional deafness in the cockpit. First, a recent study has demonstrated using a simulated piloting task that spoken distractors (i.e., words) are more likely to be intensively processed than simple tones [11]. We argue that critical information such as the disengagement of the auto pilot should be indicated using a spoken alarm like "autopilot disengaged" and not simple tones as has been the case up to now. Second, now two main ERP components associated with the occurrence of the inattentional deafness phenomenon (i.e., the N100 and the P3b components) have been identified, we argue that more work should be done to implement Brain Computer Interfaces enabling the detection of decreased N100 and P3b responses to auditory alarms to inform the pilots they may suffer from inattentional deafness.

References

1. NTSB.: Aircraft Accident Report Eastern Air Lines, Inc. L-1011, N310EA, Miami Florida 29 Dec 1972. http://www.ntsb.gov/investigations/AccidentReports/Reports/AAR7314.pdf (1973)
2. Koreimann, S., Strauß, S., Vitouch, O.: Inattentional deafness under dynamic musical conditions. Atten. Percept. Psychophys. **73**(6), 1780–1789 (2009)
3. Macdonald, J.S., Lavie, N.: Visual perceptual load induces inattentional deafness. Atten. Percept. Psychophys. **73**(6), 1780–1789 (2011)
4. Giraudet, L., St-Louis, M.E., Scannella, S., Causse, M.: P300 event-related potential as an indicator of inattentional deafness? PLoS ONE **10**(2), e0118556 (2015)
5. Bennington, J.Y., Polich, J.: Comparison of P300 from passive and active tasks for auditory and visual stimuli. Int. J. Psychophysiol. **34**(2), 171–177 (1999)
6. Scannella, S., Causse, M., Chauveau, N., Pastor, J., Dehais, F.: Effects of the audiovisual conflict on auditory early processes. Int. J. Psychophysiol. **89**(1), 115–122 (2013)
7. Kramer, A.F., Trejo, L.J., Humphrey, D.: Assessment of mental workload with task-irrelevant auditory probes. Biol. Psychol. **40**(1), 83–100 (1995)

8. Luck, S.J., Kappenman, E.S.: The Oxford Handbook of Event-Related Potential Components. Oxford university press, Oxford (2011)
9. Oldfield, R.:. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, **9**(1), 97. doi:10.1016/0028-3932(71)90067-4
10. Valéry, B., Peysakhovich, V., Causse, M.: Hear me flying! Does visual impairment improve auditory display usability during a simulated flight? Procedia Manuf. **3**, 5237–5244 (2015)
11. Causse, M., Peysakhovich, V., Fabre, E.F.: High working memory load impairs the processing of linguistic stimuli during a simulated piloting task: an ERP and pupillometry study. Front. Hum. Neurosci. (in revision)

Part II
Cognitive Computing

Application of a Simulation-Based Software Tool for the Prospective Design of IT Work Places

Nico Feller, Andreas Amann, Ulf Müller, Michael Schiffmann, Oliver Kurscheid and Markus Gorzellik

Abstract The following article presents an approach on the extension of WorkDesigner—a simulation-based software tool for the strain-based staffing in industrial manufacturing—for the prospective design of IT work places. After a short introduction of WorkDesigner, the common economical and technical need for the individual design of IT work places is described in the following chapters. Here the current mega trend Digital Transformation takes center stage. Chapter 4 presents additional parameters for the adaption of WorkDesigner to the drafted "digital" needs. Finally, the results and the future developments are discussed.

Keywords Workdesigner · IT work place · Prospective design · Digital transformation · Internet of things · Industry 4.0

N. Feller (✉) · U. Müller · M. Schiffmann
Cologne UAS, Laboratory for Manufacturing Systems, Betzdorfer Straße 2, 50679 Cologne, Germany
e-mail: nico.feller@th-koeln.de

U. Müller
e-mail: ulf.mueller@th-koeln.de

M. Schiffmann
e-mail: michael.schiffmann@th-koeln.de

A. Amann
Detecon International GmbH, Pharma & Healthcare, Sternengasse 14–16, 50676 Cologne, Germany
e-mail: Andreas.Amann@detecon.com

O. Kurscheid · M. Gorzellik
Cologne UAS, Betzdorfer Straße 2, 50679 Cologne, Germany

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_8

1 Introduction

Digital Transformation, Internet of Things or Industry 4.0 are all representative terms for a new reality: the global crosslinking between humans, products and services as well as resources. This crosslinking has the potential to release enormous power of innovation. Here the information technology takes center stage and will be the enterprises' key to success and competitiveness. A high level of flexibility and speed of the IT is demanded for the long-term assurance of profitability and competitiveness. Common IT approaches can not sufficiently guide the processes of the digital transformation. A change of the IT paradigm is needed to enable application- and user-specific work modes (cf. [1]).

In 1995 Landauer [2] already argued that efficient and ergonomic software design can increase the productivity of IT-based work by about 700 %.

In this context a field study by System Concepts LTD. System Concepts LTD [3] identified three primary reasons for the respectively low performance of IT software solutions: confusing dialog elements (12 %), inconsistent design of graphic user interfaces (25 %), work processes mismatched to dialog sequences (60 %) and others (3 %) (see also [4, p. 1077]).

To assess and improve innovative IT software solutions and IT work places with regard to productivity and workload, WorkDesigner, a simulation-based software tool for the strain-based staffing and design of work processes, is proposed to adapt to the current "digital" need for action.

WorkDesigner is based on the approach established by Feller and Müller [5] using an age- and stress-based simulation model for the development and assessment of work systems for employees in industrial manufacturing; Fig. 1 shows some impressions of WorkDesigner.

It is important to note that in this approach employees and work places are the main parameters. Based on the specific age, sex and an absolute term for the individual's ability the employee's physical ability level and the corresponding work ability level is determined for every discrete simulation step. Every work place is defined by five parameters (stress factors rated from excellent to deficient): lighting, climate, noise, work posture and work intensity. Considering age-related changes in the employees' abilities, all stress factors are individually weighted for every interaction between an employee and a work place. Based on the formulas provided by Feller and Müller [5] the overall stress is calculated for every work process, which leads at the end to the determination of the employee's utilization of his or her work abilit. (cf. [6]).

In the following chapters the common economical and technical need for action is described as well as an approach on the extension of WorkDesigner for the prospective design of IT work places is drafted.

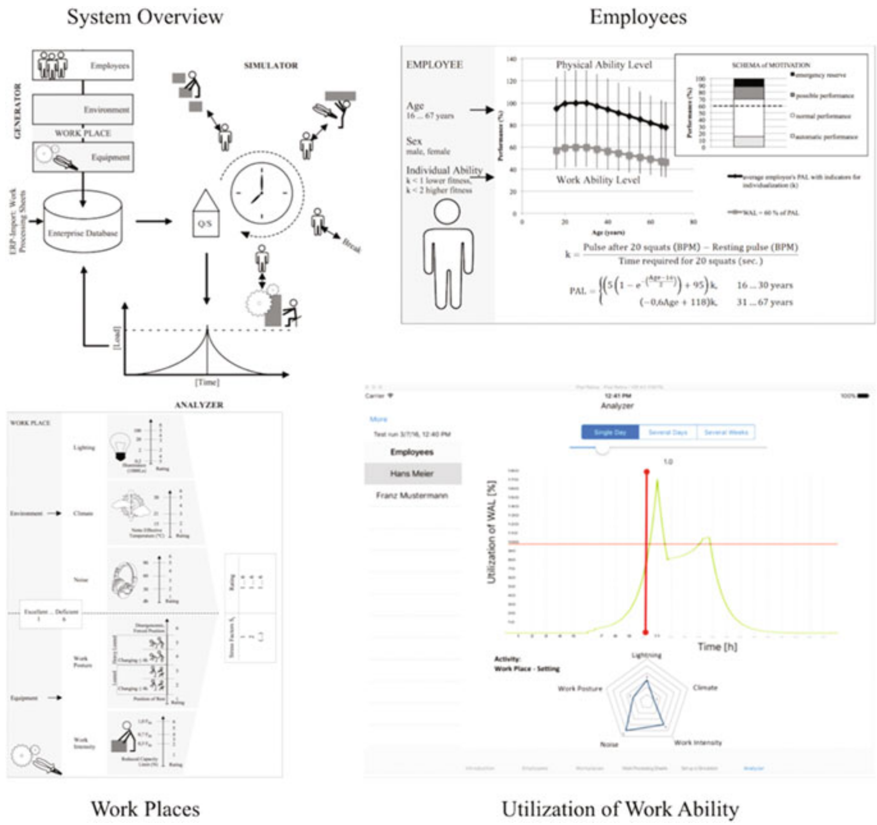


Fig. 1 Impressions of WorkDesigner (cf. [5])

2 The Role of IT in the Context of Digital Transformation

The Digital Transformation has become the ultimate challenge for almost all industries. It's leading to significant changes within the company with impact on the operational structure, strategic positioning and the organization. The use of technology to radically improve performance, enhance the customer engagement and create new digital business models is the overall task and a substantial challenge for the IT organization.

In the past the IT departments focused mainly on supporting the business by offering the needed technical capabilities for highly optimized processes. Running a stable and reliable infrastructure and offering adequate services were the primary focus. A sufficient business and IT alignment was the goal.

This has changed radically through the Digital Transformation. Suddenly the digital services are a substantial part of the offering. The requests towards the IT are no longer coming from the internal organization, but from the external market and the customer itself. Not stability but agility and speed are the critical capabilities of today. To perform this tremendous mind shift is the major challenge of today's IT departments.

To be directly involved in the product design, the development process and the digital service offering is a new situation for the IT department and causes sustainable discomfort.

Through the enormous pressure the most IT departments do not see the great chances they have been given by moving into the strategic center of the core business. The shift from product orientated business models towards service focused business models will make the IT a substantial part of the value chain itself.

The integration of digital services will not only have a tremendous impact on the IT department but on almost all areas of a company. The number and the complexity of IT workplaces will grow dramatically in the next years, and at the same time the flexibility of the workplace infrastructure will change as well. Flexible and mobile workplace concepts within the company and outside must be supported. The integration of smartphones and tablets will take place and will have a tremendous impact.

A seamless IT infrastructure, optimized towards the needs of the operational procedures will be substantial for the overall success of a company. New ways how to increase the efficiency of IT workplaces is a task most industries haven't tackled yet. Through the Digital Transformation and the new role the IT will take in the near future, the ergonomic aspects of IT workplaces will become a critical success factor for companies.

3 Common Approaches for the Design of IT Work Places

Many different influences impair the achievement of great software. Some programs are more compelling to increase or decrease the user efficiency than others. The quality of software should not only be selected by the visual appeal of the graphical user interface but more influenced by the quality of the programming of the core object model.

As shown in Fig. 2 visual design affects the user experience by about 10 % which is by way of comparison, three times less than the interaction of techniques and design. The most important influence is the object model with about 60 %, which contains the task model and the consideration of significant user groups.

There are principles of software development placed in norms like ISO 9241 which the subject matter is the Ergonomics of human-system interaction, including dialog principles (cf. [8]), guidance on software individualization (cf. [9]), principles for the presentation of information (cf. [10]) and many more. There is also a standard about the forms (cf. [11]) as well as the norm for the software ergonomic design principles and framework for multimedia user interfaces (cf. [12]).

The general purpose is to recommend how the software structure be realized. It can be comprehended that parts of the ISO 9241 as well as elaborations (cf. [13]),

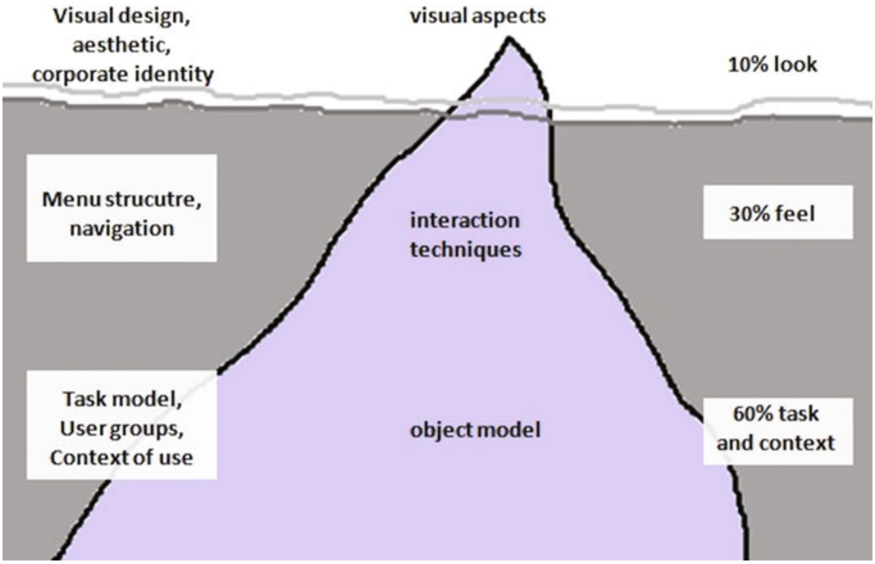


Fig. 2 Hidden influences of usage quality (cf. [7])

which are mostly based on the ISO 9241, are aiming to accomplish the following characteristics: adequacy of tasks, self-description capability, expectation compliance, learnability, controllability, fault tolerance, and customizability.

According to the ISO 9241, the software should maintain the specifications of each part to reach this characteristic. There are some known basic approaches such as grouping similar information to receive clarity (cf. [10], Chap. 6.3.4). But it is quite surprising these standards should be considered as the publication date is more than ten years since its original publication. This is an inherent problem of the ISO 9241. Most parts are just too old for the extremely fast rising and especially changing information technology sector.

The flow of the user dialog today is an example, which is not comparable to the entire principles of 2006 (cf. [8]). There is no sense to use outdated specifications that do not fit with the features of today's information technology. Especially when considering the entire domain of connected devices.

Furthermore, there is a considerable number of specifications that in some cases concern an overstated level of detail, as well as most specifications, are still not entirely regulated. Fortunately, there are no regulations that dictate how to design a Button (cf. [11], Chap. 5.4.8).

A significant problem is the neglecting of user groups. Different knowledge of users can affect the software-usability and acceptability in a strong way. A well developed user-adaptable software program can generate a high efficiency because it is providing the user new skills and knowledge in an optimal way. Only fundamental characteristics can be inferred from the literature.

Today, great software design means not only to accommodate user skills; it also means the adjustment to lifestyle. This implies self-adaption to the software via exterior influences. A nice example of the difference between the dialog principles is Google Glass. Depending on the location and view, Google Glass displays necessary information through an augmented reality interface.

To achieve optimal usage, the software would be required to adapt to any hardware device. There is not one correct way and always best software design; the designer needs to differentiate the product to reach the optimum level of characteristics and general acceptance in any and all situations.

4 Extension of the Simulation-Based Software Tool WorkDesigner for the Design of IT Work Places

As it has been shown in the previous chapters, the impact of IT is increasing. More and more workplaces are turning into mobile offices, which enables employees to work from anywhere. As result of this, their ergonomic requirements are changing as well. Accordingly, the simulation-based software tool WorkDesigner is proposed to be adapted. Based on the considerations of the previous chapters, the following part gives an approach on how the software should be expanded.

To evaluate software regarding its impact on the users' stress level, there are a huge number of different parameters. Most of the common simulation systems are lost in the detail. For companies which want a quick result this is not practical. So additional parameters therefore have to be evaluated quickly, either in a subjective or an objective way.

4.1 User Index

As discussed in Chap. 3, a big mistake in the DIN norm is disregarding the user, because how effective software is depends heavily on him. So as Florin [14] said, the user has to be recorded accurately. The software WorkDesigner already records the user in a sufficient but limited (physical) way. To get more details the user description has to be expanded. An example how this could look like is given in Fig. 3.

The parameters for age and gender are already used in WorkDesigner and now added to this are nationality, area of expertise and a few questions about the personal behavior with software. While the questions about behavior works as an indicator for the experience and attainments with software, the "area of expertise" gives an idea of the affinity to technology. One point often neglected in DIN is "nationality". So the fact that different cultures have different ways of reading and writing are included in this point. With an appropriate rating the user description

Fig. 3 User description

generates a "user index" with which the learning curve and stress level of the user can be derived.

4.2 *Software*

Just as important as the user is the software itself. A highly qualified user can not gain his full potential without having the right software available to him. So it is important to find parameters which allow a professional and also personal way of evaluation. Figure 4 shows some parameters which meet these requirements.

All of these could be evaluated quickly on a scale of one to six. According to Florin [14], who explained that usability and design have to be clearly divided, the four parameters are split into these two categories.

Setup The first impression of software. This contains the optic and presentation for the user. Are objects arranged in a proper way? Do I have all important tools on the first page? These are just a few questions which give a personal impression of the Software and represent a subjective way to evaluate this point. To evaluate the software more objectively, the DIN EN ISO 9241-12 should be used. The DIN gives a huge number of criteria as to how objects have to be arranged and how much information should be given. In addition to that, the knowledge about human nature, as for example Florin [14] explained it, can be important.

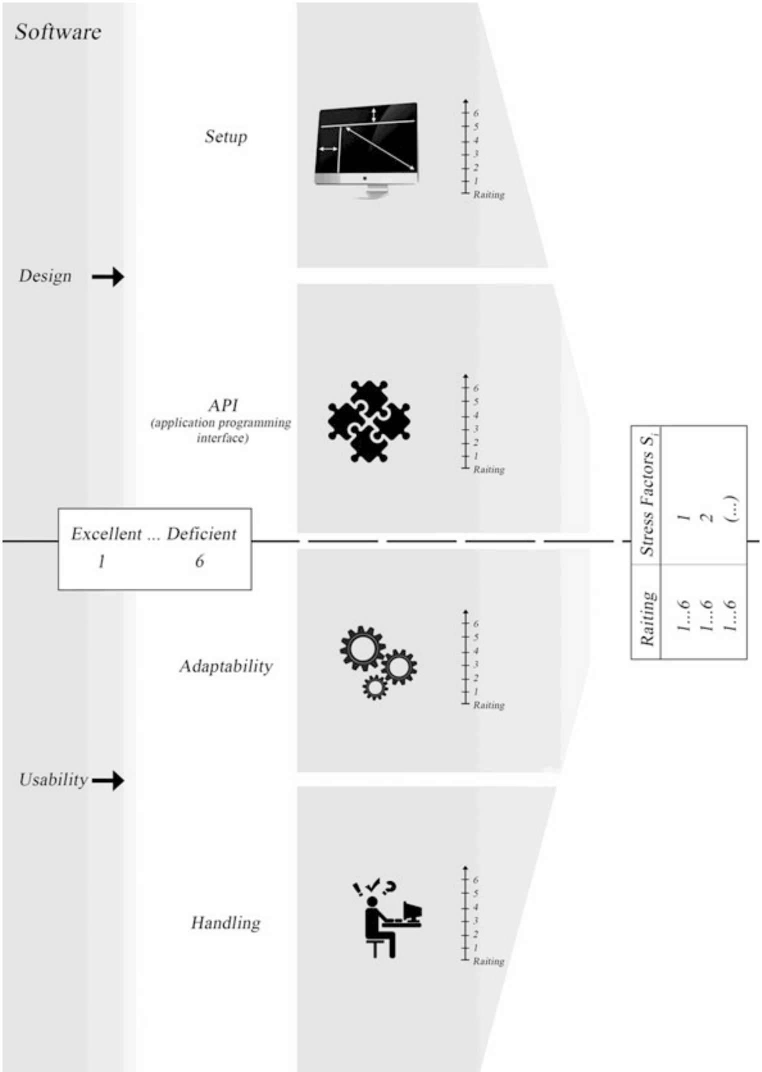


Fig. 4 Software parameters

API (Application programming interface) Apart from the Setup one important aspect of Software is the possibility of sharing data with other programs. This connection is the API. Most workplaces are working with more than one type of Software, so it is important to have good a connection between all programs. How well this works and if all programs are involved is a good way to evaluate this point. Also the possibility to share between different operating systems, like iOS and Windows, has to be in focus. In contrast to Setup it is difficult to evaluate the

API in an objective way. How many and which Software should be supported is very personal and can't be a fixed number.

Adaptability The first parameter which deals with usability. In 1995 Landauer [2] already described that with the right software, you could raise your productivity to nearly 700 percent. In addition to Landauer [2] the System Concepts LTD. Reference [3] figured out that the important part of software working well is the matching between software dialog and work processes. It is therefore important that you can fit the software to your personal requirements. How many different changes can you do? Can you change the design and can you create own software add-ons? These questions could help to evaluate the software more objectively.

Handling As Reiterer and Geyer [15] describe, usability transforms more and more into the user experience so it is important to take a look at this. So the last suggested parameter is handling which describes the user experience and feelings while working with software. Experiences like: How fast can I work with the software? How much help did I get and also how consistent are the different parts of the software? All of this creates the impression of the work and has an influence on the level of stress. The DIN EN ISO 9241-10 gives a few more additional points.

In combination with the user index, the software parameters give a good interpretation for the level of stress. It is a quick way to evaluate software but includes all important aspects of the software ergonomic.

5 Discussion

In the context of Digital Transformation flexible and mobile work place concepts are highly demanded likewise by employees (user experience—the employee as IT's customer) and enterprises (to raise productivity and maintain competitiveness). Here the IT takes center stage and will be the key to economical success.

As described the common approaches for the design of IT work places are not sufficiently adjusted to the individual user's needs, the requirements of the Digital Transformation and the current mobile technologies. So the application of a simulation-based software tool for the assessment and prospective development of IT work systems is proposed. For this purpose, an approach on the extension of the software solution WorkDesigner with additional user/employee parameters as well as software parameters respectively complimentary work place parameters is introduced in Chap. 4.

It has to be mentioned that the presented work is still in progress. The next development steps are the sufficient and quantitative definition of the introduced parameters, the software integration and the validation in a field study.

References

1. Steinbrunn, M., Bubner, N.: IT-Strategien in Zeiten der Digitalen Transformation. Detecon Update, Digital Business News. Detecon Consulting GmbH, GER (2015)
2. Landauer, T.K.: The Trouble with Computers, Usefulness, Usability and Productivity. MIT Press, USA (1995)
3. System Concepts LTD.: When GUIs Fail. London (2008)
4. Schlick, C., Bruder, R., Luczak, H.: Arbeitswissenschaft. Springer, Berlin (2010)
5. Feller, N., Müller, U.: Development and assessment of work systems for elder employees in industrial manufacturing. In: Trzcielinski, S., Karwowski, W. (eds.) Advances in the Ergonomics in Manufacturing, Managing the Enterprise of the Future, pp. 140–151. AHFE Conference ©, USA (2014)
6. Müller, U., Gust, P., Feller, N., Schiffmann, M.: Simulationstool zur alters- und belastungsorientierten Entwicklung und Bewertung von Arbeitssystemen in der industriellen Fertigung. In: Rabe, M., Clausen, U. (eds.) Simulation in Production and Logistics 2015, pp. 469–480. Fraunhofer IRB Verlag, Stuttgart (2015)
7. Rudolf, C.: Handbuch Software-Ergonomie Usability Engineering. Unfallkasse Post und Telekom, Tübingen (2006)
8. Ergonomics of Human-System Interaction—Part 110, Dialogue Principles (ISO 9241-110:2006), German Version EN ISO 9241-110:2006. Beuth, Berlin (2006)
9. Ergonomics of Human-System Interaction—Part 129, Guidance on Software Individualization (ISO 9241-129:2010), German Version EN ISO 9241-110:2010. Beuth, Berlin (2010)
10. Ergonomics of Human-System Interaction—Part 112, Principles for the Presentation of Information (ISO/DIS 9241-112:2015), German and English Version prEN ISO 9241-112:2015. Beuth, Berlin (2015)
11. Ergonomics of Human-System Interaction—Part 143, Forms (ISO 9241-143:2012). Beuth, Berlin (2012)
12. Software Ergonomics for Multimedia User Interfaces—Part 1, Design Principles and Framework (ISO 14915-1:2002). Beuth, Berlin (2002)
13. Oppermann, R.: Einführung in die Software-Ergonomie. http://userpages.uni-koblenz.de/~oppi/SE-EinfuehrungGesamt.pdf (2016)
14. Florin, A.: User-Interface-Design Usability in Web-und Software-Projekten. Books on Demand GmbH, Berlin (2015)
15. Reiterer, H., Geyer, F.: Mensch-Computer-Interaktion. In: Kuhlen, R., Semar W., Strauch, D. (eds.) Grundlagen der praktischen Information und Dokumentation, Handbuch zur Einführung in die Informationswissenschaft und -praxis, pp. 431–440. De Gruyter Saur, Berlin (2013)

Evaluation of User's Affective Engagement While Interacting with Educational Technologies: A Pilot Study

Marvin Andujar, Patricia Morreale, Yerika Jimenez, Luis Jimenez and Juan E. Gilbert

Abstract There are several educational technologies developed to enforce learning in computing. These tools success have generally been studied through subjective measurement. However, subjective data may be inaccurate due to users not providing exact information. Therefore, validity may be of concern. Human-Computer Interaction (HCI) researchers have recently implementing neurophysiological tools towards their user studies for subjective data support. This paper presents an exploratory user study to implement neurophysiological tools to gather objective data to evaluate the usefulness of educational technologies. A between-subject study was conducted comparing Alice and App Inventor user's affective engagement levels. The results showed no statistical significant. Lastly, this paper can serve as a short guideline on how to adapt neurophysiological tools towards their user studies.

Keywords Brain-Computer interfaces · Educational technologies · Affective engagement · Affective brain-computer interfaces · Human-computer interaction

M. Andujar (✉) · Y. Jimenez · J.E. Gilbert
Computer and Information Science and Engineering Department, University of Florida, Gainesville, FL, USA
e-mail: manduja@ufl.edu

Y. Jimenez
e-mail: jimenyer@ufl.edu

J.E. Gilbert
e-mail: juan@ufl.edu

P. Morreale
Computer Science Department, Kean University, Union, NJ, USA
e-mail: pmorreal@kean.edu

L. Jimenez
Mobile Content Solutions, Verizon Wireless, Bedminster, USA
e-mail: luis.jimenezvillanueva@verizonwireless.com

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_9

1 Introduction

There has been a promising interest and need of Brain-Computer Interfaces (BCI) in the Human-Computer Interaction (HCI) and User Experience (UX) community. This recent interest is serving as a path towards the rise of new neurophysiological applications for healthy users. Therefore, since 2001, there has been a significant increase in BCI publications based on non-medical applications [1]. These publications range from studying the affective and cognitive state of users while performing different tasks to utilizing BCI devices as an input to control machines or virtual characters. Another area that BCI may be helpful and has attracted interest is in education. BCI can be used to gather affective engagement from students while they interact with educational technologies. The adaptation of BCIs may contribute towards understanding the design flaws of the educational tools.

This paper presents an exploratory user study of the measurement of affective engagement through a wearable BCI device with high school students while they perform a task in an educational technology tool (Fig. 1). The purpose of the study is to identify the most engaging educational tool for high school students and explore how BCIs can be adapted towards such study. Lastly, this paper may be used as a guideline on how BCI may be adapted for educational studies.

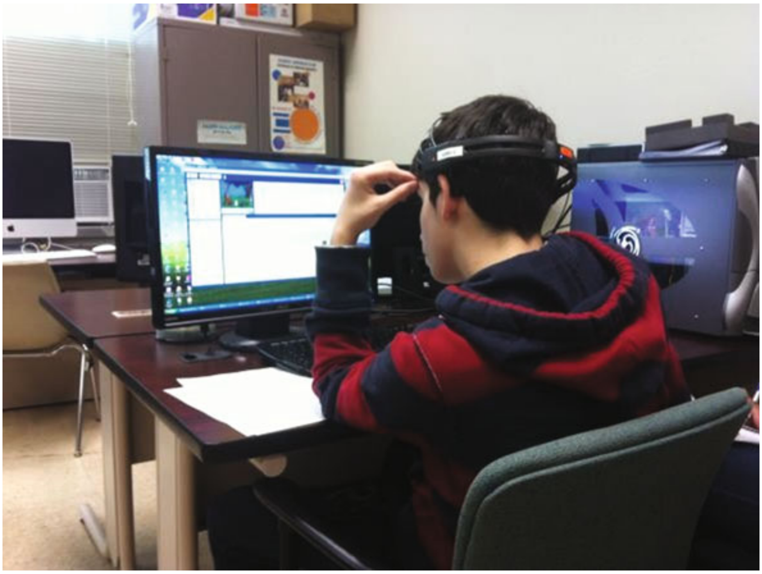


Fig. 1 User completing a programming task while wearing the BCI

2 Background

There is research work on implementing BCI to Human-Computer Interaction studies. Solovey et al. [2] used Functional-Near Infrared Spectroscopy (fNIRS) methodologies to implement them to HCI studies. Their studies show appropriate user posture while adapting physiological devices to a HCI/UX research study. Affective Engagement evaluations in HCI/UX user studies have been adapted as well. Andujar and Gilbert presented a new method to measure engagement through EEG and enhance it with media. If the engagement of the user is low while the user is reading a piece of information, then the application recommends a video related to the reading to increase engagement [3]. Furthermore, in another study Andujar et al. compared two learning techniques (video games and reading a book) by measuring engagement physiologically. The purpose was to observe any correlation between physiological engagement and information retention (measured by test scores after learning) [4]. Another study involving engagement is by Szafir and Mutlu. They quantified engagement by measuring the spectral bands (alpha, theta, and beta) using the non-invasive BCI EEG Neurosky device while the users were interacting with an adaptive agent [5].

Previous work has documented of ways of implementing a BCI technology to HCI/UX research. However, there is no or little work done concentrated on the impact of educational technologies by measuring the user's affective state. Therefore, this paper presents a preliminary study on how a non-invasive BCI can be implemented to measure engagement of users while interacting with educational technologies.

3 Methodology

There were a total of 13 participants (6 Alice group and 7 App Inventor group) in this preliminary study. These participants were high school students participating in an enrichment program for college preparation. This was a between-subject study, the participants were randomly selected to one of the two groups: Alice or App Inventor. The participants were given a task to complete in 10 min. The tasks involved a simple programming task in which the students could accomplish in a small amount of time. While the participants were completing the given task, the Emotiv was gathering engagement levels in real time. The purpose of the study was not just to compare the engagement of two interfaces and determine which one was more engaging, but to explore how BCI can be implemented to evaluate engagement while users interact with these interfaces. The participants were broken into two groups due to the limitation of time they had to conduct the study.

The procedure of the study consisted of the following phases:

- *Mounting Phase*: Before the user started performing the task, the apparatus was mounted on the user's head. This usually takes around 5–15 min depending on the amount of hair on the user's scalp and the head shape.
- *Affective Calibration Phase*: Affective data was recorded for a period of at least of 10 min to obtain accurate data. Therefore, in this phase, the user closed his/her eyes and relaxed (trying not to think about anything) until 10 min have passed (this phase can be manipulated at the experimenter choice and study conditions).
- *User Task/Engagement Measurement Phase*: The user started performing the given task for a period of 10 min, while s/he was performing the task, Emotiv recorded the affective engagement.

3.1 Affective Engagement

Engagement was recorded physiologically with a non-invasive BCI device. The engagement values ranges from 0 (not engaged at all) to 1 (extremely engaged). Figure 2 shows an example of how on engagement changes through time while a user is interacting with the programming environments. For that particular user his/her engagement ranged from 0.55 (lowest engagement) to 0.95 (highest engagement). There is a lot of data collected in a period of 10 min as electroencephalographic (EEG) data is recorded in milliseconds, therefore every two minutes the engagement was averaged to establish the changes in engagement several periods of time.

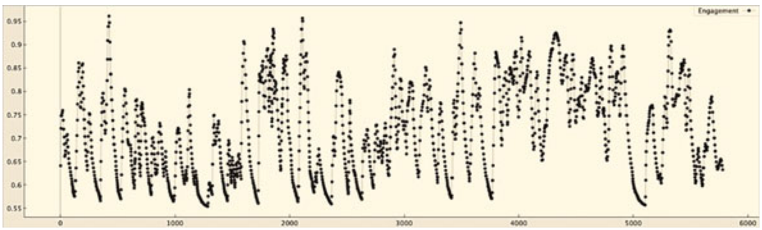


Fig. 2 Example of engagement change through time (unprocessed) (x-axis: time | y-axis: engagement)

Fig. 3 Emotiv EPOC device

3.2 *Emotiv Apparatus*

The Emotiv apparatus (Fig. 3) is a non-invasive EEG wireless device. It processes raw EEG data to determine the user's affective and cognitive and can be implemented as an input to control machines such as: computers, electric wheel chairs, cars, mobile phones, robots, etc. The data measurements and the input control interaction between the device and the user is achieved by the 14 channels (AF3, F7, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) of the device. These devices are based on the international 10–20 systems [6]. These channels are used to read the electrical signal of the human brain, which allow researchers to understand more the state of the user while performing a specific task. This device has been widely used by several researchers for various research studies. A good example is the work of Campbell et al. [7] created an application where the users can make phone calls by selecting the contact using the electrical signals of their brain, where the device translated those signals to commands. These different implementations stated in related work and the work by Campbell et al. represent the popularity and the accuracy of the device while using it as a measurement or as an input, which is the reason this device was used for this particular project.

3.3 *Educational Technologies*

The programming drag-n-drop environments used in the study were the Android App Inventor and the 3D programming environment Alice. These tools are programming environments dedicated to teach programming using a drag-n-drop interface to students without programming background. Both programming environments use a block interface to put together the code parts as a puzzle. App Inventor is an environment developed (Fig. 4) by Google for non-developers to build their own mobile applications without the need of programming.

Alice is a programming environment (Fig. 5) developed by Carnegie Mellon University to increment the number of enrollment in Computer Science and to

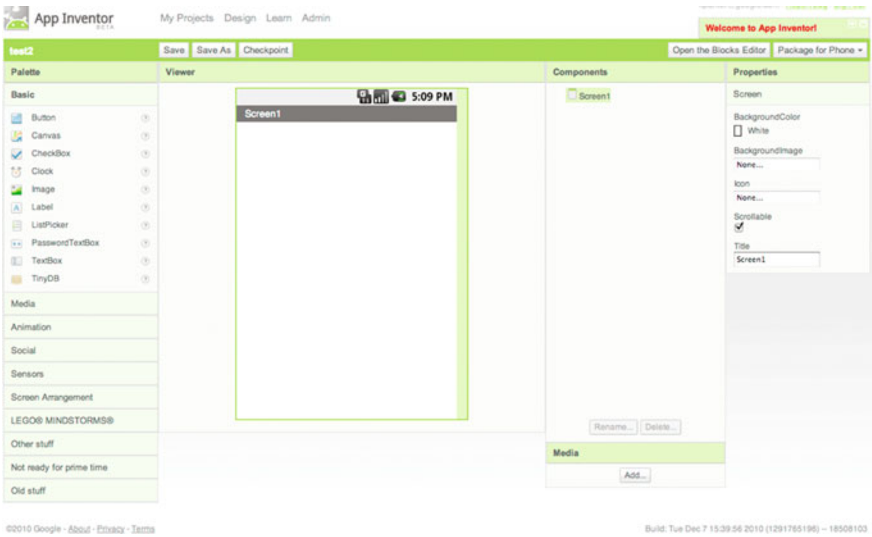


Fig. 4 Android App inventor user interface

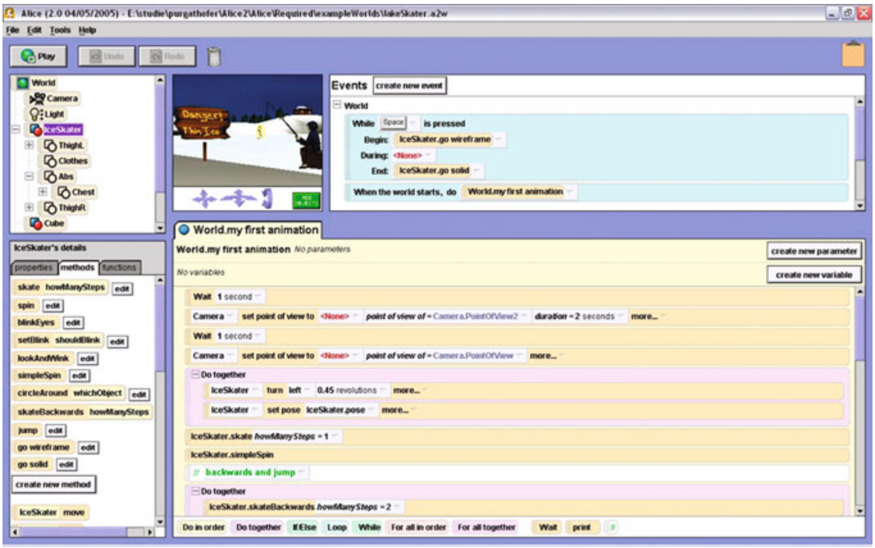


Fig. 5 Alice user interface

improve retention within the major [8]. Alice allows the users to create interactive stories by dragging and dropping 2D objects and programming their function by dragging and dropping code, hence no syntax memorization.

4 Preliminary Results and Discussion

The results presented in this paper are the affective engagement values obtained from the Emotiv apparatus. A two sampled t-test was performed to determine if there is a statistical significant on engagement levels between the two interfaces. Based on the result, there is no statistical significance ($P > 0.05$). As shown in Fig. 1, there is a lot of data collected even in a small amount of time. Therefore, for this particular study engagement was average every two minutes to see the difference in engagement through time.

The Alice group (Fig. 6) had a very similar distribution of engagement for the Alice group (Mean = 0.0608717 | STD = 0.089198 | Min = 0.491171 | Max = 0.716489). It was observed that engagement was quite different across the six participants, except for two. Also, their engagement did not change a lot through time. Although, one of the participant's engagement dropped at the end of the task.

The App Inventor group (Fig. 7) had similar engagement values except for two participants (Mean = 0.0635827 | STD = 0.080102 | Min = 0.549641 | Max = 0.723772). It can be observed the results for this group was different from the Alice group. The engagements for most of the participants were close to 0.7

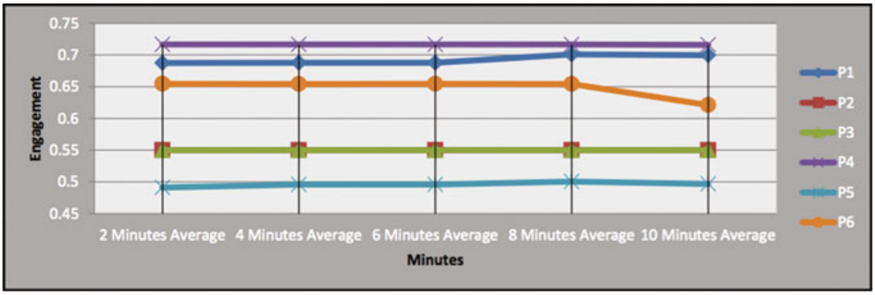


Fig. 6 Alice group: engagement results every two minutes

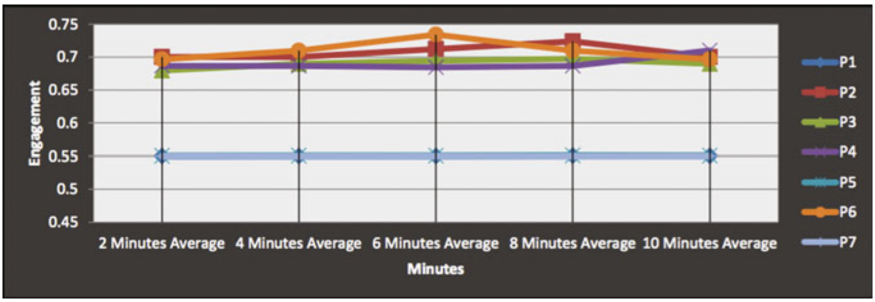


Fig. 7 App inventor group: engagement results every two minutes

except for two participants. The other two participants had identical engagement, which is interesting to observe.

5 Current BCI Challenges

Current wearable BCI devices offer low-cost expenditure, flexibility, and a wider variety of applications for different purposes. However, the channels' qualities of the wearable BCIs are not as good as the traditional BCIs. It requires the user to stand still, restrict muscle movement as much as they can, have shorter or no hair, and be at a room at a decent cold temperature to avoid sweat. These requirements do not allow the user to perform a real-world task in a natural manner and may affect how they feel about BCIs negatively. An ergonomic issue is that these BCIs are not adjustable for all the different types of head shapes and sizes. Each device has their own limitation on how and who can wear them.

The current states do not allow researchers to completely substitute the subjective measurements with BCIs, but the gathered data is still accurate enough for supporting the subjective information obtained from the users. Also, through different research studies, recommendations on how to advance this technology for universal adaptation of users performing real world tasks or laboratory experiments may be beneficial towards its advancement.

6 Summary

This paper describes a study measuring affective engagement physiologically while users interact with educational technologies. This paper also gives a brief introduction on how BCI may be implemented in educational studies. The results demonstrate the differences of engagement in each group, in which there is not a big difference when interacting with theses interfaces. This may also help educators see how BCI may be used to determine which of their students were not actively engaged through the exercise while using an educational tool.

Further studies are needed to develop a standardize methodology to implement neurophysiological engagement methods in research studies to determine the involvement of the users while interacting with educational tools. This may help those who develop the tools to improve them to provide a high level of engagement through the user interaction.

Acknowledgments This material is based in part upon work supported by the National Science Foundation under Grant Number DGE-1315138 & DUE-1060545. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Van Erp, J.B., Lotte, F., Tangermann, M.: Brain-computer interfaces: beyond medical applications. Computer **4**, 26–34 (2012)
2. Solovey, T.E., Girouard, A., Chauncey, K., Hirshfield, L.M., Sassaroli, A., Zheng, F., Fantini, S., Jacob, R.J.K.: Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology (UIST'09), pp. 157–166. ACM, New York, NY, USA (2009)
3. Andujar, M., Gilbert, J.E.: Let's learn!: enhancing user's engagement levels through passive brain-computer interfaces. CHI'13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13), pp. 703–708. ACM, New York (2013)
4. Andujar, M., Ekandem, J.I., Gilbert, J.E., Morreale, P.: Evaluating engagement physiologically and knowledge retention subjectively through two learning techniques. In: Proceedings of 15th International Conference, HCI International 2013, Las Vegas, NV, USA, Part II, vol. 8005, Lecture Notes in Computer Science, pp. 335–342, 21–26 July 2013
5. Szafir, D., Mutlu, B.: Pay attention! designing adaptive agents that monitor and improve user engagement. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, Austin, Texas, USA, pp. 11–20. CHI'12. New York, NY, 5–10 May 2012
6. Pivik, R.T., Broughton, R.J., Coppola, R., Davidson, R.J., Fox, N., Nuwer, M.R.: Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. Psychophysiology **30**(6), 547–558 (1993)
7. Campbell, A.T., Choudhury, T., Hu, S., Lu, H., Mukerjee, M.K., Rabbi, M., Raizada, R.D.D.: NeuroPhone: Brain-Mobile Phone Interface using a Wireless EEG Headset, MobiHeld 2010, 30 Aug 2010, New Delhi, India, pp. 3–8 (2010)
8. Kelleher, C., Pausch, R., Kiesler, S.: Storytelling Alice motivates middle school girls to learn computer programming. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1455–1464. ACM, New York, USA (2007)

On the Modelling an Artificial Cognitive System Based on the Human-Brain Architecture

Olga Chernavskaya, Dmitry Chernavskii and Yaroslav Rozhylo

Abstract The approach to modeling a cognitive system based on the human-brain architecture, called the Natural-Constructive Approach is presented. The key point of this approach is the following: an artificial cognitive system, being a complex multi-level combination of various-type neural processors, should be divided into two subsystems, by analogy with two cerebral hemispheres in a human brain. It is shown that one of them should necessarily contain a random element (*noise*) for generation of information (creativity); it is responsible for *learning*. The other one, being free of noise, is responsible for memorization and processing the *well-known* information. Emotions could be interpreted as the noise-amplitude variation and incorporated into the system by coupling the noise amplitude with the additional variable representing the aggregated value of neurotransmitter composition, which reflects the influence of subcortical brain structures. It is shown that the activity of both subsystems should be controlled by the noise-amplitude derivative.

Keywords Generation of information · Neuroprocessor · Emotions · Learning · Noise amplitude · Hemisphere · Stimulant · Inhibitor

1 Introduction

The scientific area of Artificial Intelligence (AI) covers various approaches to modeling the cognitive process: the Robotics [1, 2], Active Agent systems [3, 4], neuromorphic (neuron-based) models [5, 6], Brain Re-Engineering (BRE) [7, 8], etc. Let us stress that any neuromorphic model (though this approach seems to be the closest to the goal) inevitably meets with the "explanatory gap" between the brain and the mind [9]. This implies that we do have a lot of information on the

O. Chernavskaya (✉) · D. Chernavskii
P.N.Lebedev Physical Institute, Russian Academy of Sciences, Moscow, Russia
e-mail: olgadmitcher@gmail.com

Y. Rozhylo
NGO Ukrainian Center for Social Data, Kyev, Ukraine

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_10

single neuron structure and functioning [10, 11], as well as on manifestation of psychological reactions [12–14], but there is lack of idea on how the first could provide the second.

Another problem is connected with the Nature challenges: any human-like cognitive model should be able to answer the questions: why there are just two cerebral hemispheres? Why any human person is individual? Why there are men and women? Up to our knowledge, these problems are not in the focus of modern researches, but this is strange. Perhaps, this could be explained by the complexity of the problems of such sort. Indeed, it is very difficult to imagine and create an artificial system which being manufactured according to standard procedure, would be strictly individual.

In the papers [15–17], there was proposed and elaborated so called "Natural-Constructive Approach" (NCA) to modeling the cognitive system. This approach is based on neurophysiology data on the human-brain structure [10, 11], the Dynamical Theory of Information [18–20], and neural computing [21] (combined with the nonlinear differential equation technique). It was shown that the cognitive architecture designed under NCA makes it possible to answer the majority of the Nature challenges. Several aspects of applying this approach to AI systems were discussed recently [22, 23]. In this paper, the main points of the NCA-architecture are discussed and compared with the neurophysiology data on the human brain structure.

2 Theoretical Foundation of NCA

2.1 Human Brain Architecture

According to the neurophysiology data (e.g., [10]), the human brain consists of the cerebral cortex and the sub-cortical structures, as it is presented in Fig. 1.

The neocortex is responsible for the high-order cognitive processes, while the sub-cortical structures (thalamus, amygdala, bazal ganglies, etc.) produce, in particular, the emotional bursts and participate in the memory formation.

Neocortex The neocortex itself could be (conventionally) divided into zones ("lobes") which are responsible for the vision (occipital lobe), motor activity (parietal lobe), and auditory activity (temporal lobe), etc. Temporal zone embraces Wernicke's and Broca's areas that are responsible, respectively, for language *hearing* (word perception) and *reproducing* (word production), but not for the *speech* itself. The *speech* function, i.e., coherent and sensible transmission of information, relates to the frontal lobe that is associated with *abstract* thinking. An appropriate model of cognitive process should reproduce and explain the distribution of these functions.

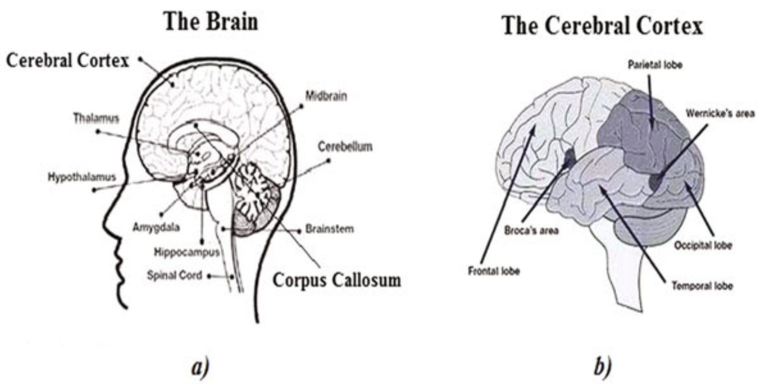


Fig. 1 Schematic representation of the human-brain structures: **a** whole view of the brain; **b** functional zones of the cerebral cortex (extracted from [24])

Sub-cortical Structures The role of sub-cortical structures was considered in [7, 8] and very interesting and completed models connecting the cognitive and emotional aspects of the learning and thinking processes were elaborated. In particular in [7], specific role of amygdala, basal ganglia, and thalamus in the learning process was demonstrated. However, there still remain unclear issues, in particular, the arrangement of the functional cortical zones. As it will be shown further, that arrangement meets quite natural explanation within the NCA architecture.

Elements of Experience Another aspect of experimental data concerns specific modification of those cortical neurons that are involved in the training process, which results in acquisition of a certain "skill" [11]. This effect is not reproduced under BRE approach [7, 8], but could be reproduced under NCA model.

Hemisphere Specialization This refers rather to neural psychology. There is a wide-spread opinion that the right hemisphere (RH) is responsible for intuition and the parallel processing of the information, while the left one (LH) provides logical thinking and sequential information processing [13, 14]. Another hypothesis that does not contradict but complements this one was proposed in [12]: RH is associated with learning new information, while LH does process the well-known one. Both these inferences (with minor revision) will be shown to be inherent for NCA architecture.

2.2 *Elements of Dynamical Theory of Information*

Definition of Information According to Quastler [24], *information is the memorized choice of one version among a set of possible and similar ones*. This definition gives an idea of *how* the information might emerge. The choice could be made as a result of two different processes, namely—*reception* (superimposed choice) and *generation* (free choice) of information. The process of generation of information could proceed only in the presence of chaotic (random) element, commonly called the *noise*.

Objective Versus Conventional Information Depending on *who* makes the choice, there appear: *objective* information (the choice made by Nature, i.e., physical principles) and *conventional* information, i.e., the choice made by certain *collective*. This choice should not be *the best*, but *individual* for a given group. In certain sense, the information that could arise in an ensemble of neurons represents the choice of this ensemble, that is, the very conventional information.

Definition of the Cognitive Process The self-organizing process of recording (perception), storing (memorization), encoding, processing, generating, and propagation of the "self" conventional information.

Main Inference from DTI Since the generation and reception of information represent *dual* (*complementary*) processes requiring different conditions, they should proceed in *two different subsystems.* The generating subsystem should contain the random element (noise), the subsystem for reception should be noise free.

2.3 Neural Computing

Neural Processors: The Concept of Dynamical Formal Neuron The neural processor is treated as a *plate* populated by the *dynamical* formal neurons (*n* being the total number) described by the nonlinear differential equations (see [15, 17]), which represent a particular case of the FitzHugh-Nagumo model [25].

"Image" Information Information about ever met objects of any kind should be recorded and stored within the Hopfield-type [26] processor (*distributed* memory). It could be described by the following equations:

$$\frac{dH(t)}{dt} = \frac{1}{\tau^H}\left\{H - \beta \cdot (H^2 - 1) - H^3\right\} \equiv \frac{1}{\tau^H}\Im_H(H, \beta), \qquad (1)$$

where H_i(t) represents the dynamical variable for Hopfield-type *i*-th neuron with stationary states being $H = +1$ (active) and $H = -1$ (passive); β_i is the parameter that control the activation threshold; τ is characteristic time of activation, i = 1… *n*. The image information is stored in the connections Ω(t), which are to be trained depending on the processor purpose.

Recording the information (learning) requires Hebbian training mechanism: initially weak connection become stronger ("blacker") in course of the learning process [27]:

$$\frac{d\Omega_{ij}^{Hebb}(t)}{dt} \propto \frac{\Omega_0}{4\tau^{\Omega}} \cdot [H_i(t) + 1] \cdot \left[H_j(t) + 1\right], \tag{2}$$

with Ω_0 and τ^{Ω} are the training-process parameters.

Storage and processing the well-known information (recognition, prognosis, etc.) require the training rule proposed by Hopfield himself [29]. This implies that all connections are initially equal and strong; during the training process the "irrelevant" ones are gradually *frozen out* according to the rule:

$$\frac{d\Omega_{ij}^{Hopf}(t)}{dt} \propto -\frac{\Omega_0}{2\tau^{\Omega}} \cdot \left[1 - H_i(t) \cdot H_j(t)\right], \tag{3}$$

what represents the principle of "redundant cut-off".

Encoding The conversion of an "*image*" (a set of M connected neurons) into a *symbol* (single neuron at the higher level of hierarchy) occurs Grossberg-type [28] processor with nonlinear competitive interactions that could be described by equations:

$$\begin{aligned} \frac{dG_k(t)}{dt} &= \frac{1}{\tau_k^G} \cdot \{[-(\alpha_k - 1) \cdot G_i + \alpha_k \cdot G_k^2 - G_k^3] \\ &- \sum_{l \neq k}^{n} \Gamma_{kl}(t) \cdot G_k \cdot G_l\} + Z(t)\xi_k(t); \end{aligned} \tag{4}$$

where G_k are variables for Grossberg-type dynamical formal neurons; k – 1…n; Z (t) being the noise amplitude, $0 < \xi_k(t) < 1$ is random function (obtained, e.g., by the Monte-Carlo method). Note that these equations are written to provide stationary states of neurons $G = +1$ (active) and $G = 0$ (passive). The parameters are: τ^G—characteristic activation time, α_k is the activation threshold (it controls the competitive ability of the k-th neuron). Competitive connections are trained according to:

$$\frac{d\Gamma_{kl}(t)}{dt} = -\frac{\Gamma_0}{\tau^{\Gamma}} \{G_k \cdot G_l(G_k - G_l)\}, \tag{5}$$

where τ^{Γ} being the characteristic time of the winner choosing, Γ_0 the model parameter. Analysis of this model has shown (see [15]) that in the symmetrical case, $\alpha_k(t = 0) = \alpha$ and $\Gamma_{lk} = \Gamma_{kl} = \Gamma(t = 0) = \Gamma_0$, the process of choosing the symbol appears to be *unstable*. This implies that the slight (casual!) advantage of one active neuron does provoke its expansion and suppression of the others (as a result of nonlinear interaction). Thereby, the paradigm of Kohonen [29] is realized: "Winner Take All". It should be stressed that it is impossible to predict in advance, *what exactly* neuron would be a winner for a given image; this choice should be made by the plate itself in the process of symbol formation. This very fact secures the *individuality* of an artificial system. Note that the process of symbol formation represents a typical example of appearance of the *conventional* information within a given system (collective of neurons).

Formation of Generalized Images After the given *G*-type neuron became the *symbol*, it should be eliminated from the competitive struggle and acquires the possibility to cooperate with other neuron-symbols to form the *generalized* image ("image-of-symbols") by means of the cooperative connections training in analogy to Eq. (2). This effect could be provided by *parametric* modification of the neuron-symbol. Actually, at the time scale $t \gg \tau^{\Gamma}$, the neuron-symbols should behave as the *H*-type neurons, while "free" *G*-neurons could compete only.

Semantic Connections Very important role belongs to the inter-plate connections between the symbol and its image (including the generalized one). In the process of symbol formation, these connections are to be trained according to Hebbian principle, i.e., by analogy with Eq. (2):

$$\frac{d\Psi^{0}_{ik}(t)}{dt} \propto \frac{\Psi_0}{2\tau^{\Psi}} \cdot G^{1}_{k} \cdot [H_i(t) + 1]; \quad \frac{d\Psi^{\sigma-1}_{kl}(t)}{dt} \propto \frac{\Psi_0}{\tau^{\Psi}} \cdot G^{\sigma}_{k} \cdot G^{\sigma-1}_{l}, \tag{6}$$

where only active neurons participate in the training process; $\sigma = 2\ldots N$ represents the number of the plate ("hierarchy level"), N is total number of levels; Ψ_0 and τ^{Ψ} are training parameters.

3 Cognitive Architecture Under NCA

3.1 Scheme of the Cognitive System

The scheme of the architecture designed under NCA in [15–17] is presented in Fig. 2. The whole system is divided into two (similar) "hemi-systems": RH (Right Hemi-system) containing the noise, and LH (Left Hemi-system) that is free of noise. The terms are chosen to correlate these "hemi"-systems with cerebral hemispheres, with the cross-hemi-system connections Λ being related with *corpus callosum* that is aimed to ebsure the dialog between subsystems. The noise in RH provides generation process, i.e. production of new information and learning; LH is responsible for reception and processing the already known (learned) information. This specialization, being the theoretical result of DTI principles only, surprisingly coincides with inference of practicing psychologist Godberg [12]. This fact represents a pleasant surprise and indirect confirmation of NCA relevance.

The whole system represents complex multi-level block-hierarchical combination of different-type processors. The lowest level is presented by Hopfield-type processors H^0 and H^{typ}, the other levels embrace the Grossberg-type processors,

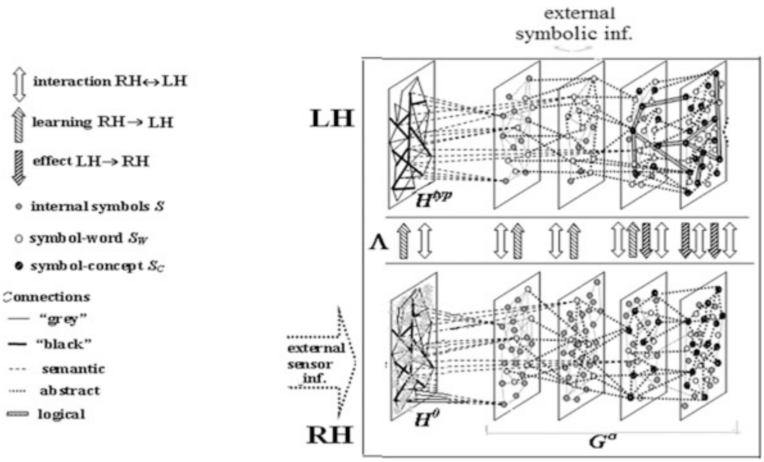


Fig. 2 The scheme of cognitive architecture

which, however, could produce the "generalized images" ("image-of-symbols") by the same mechanism of the cooperative-connection training. The number of G-type plates is neither limited, nor fixed: they appear "as required" in course of the system evolution.

This structure evolves by itself due to the self-organizing principle of "connection blackening" (see below). This implies that all the connections in RH are training according to the Hebb's principle of "connection amplification", as in Eq. (2). In LH, on the contrary, all connections are training according to original version proposed in [26], i.e. the principle of "redundant cut-off", as in Eq. (3). Thus, RH provides the *choice*, while LH performs the *selection,* with RH being the Supervisor for LH.

3.2 Equations for Neuron Ensemble Interaction

The equations describing interactions between neurons of various-type processors could be written in the form (see [15–17]):

$$\begin{aligned}\frac{dG_k^{R,\sigma}}{dt} = &\frac{1}{\tau_G}[\Im_G(G_k^{R,\sigma}, \alpha_k^\sigma(\{\Psi_{ki}^{R,(\sigma-1)}\}, G_{\{k\}}^{R,(\sigma+\nu)}))\\ &+\hat{\Upsilon}\{G_k^{R,\sigma}, G_l^{R,(\sigma+\nu)}, \Omega_{kl}^{R,\sigma}, \Psi_{ki}^{R,(\sigma-1)}, \Psi_{ki}^{R,(\sigma+1)}\}]\\ &+Z(t)\cdot\xi(t) - \Lambda(t)\cdot G_k^{L,\sigma},\end{aligned} \tag{7}$$

$$\begin{aligned}\frac{dG_k^{L,\sigma}}{dt} = \frac{1}{\tau_G}&[\Im_G(G_k^{L,\sigma}, \alpha_k^{\sigma}(\{\Psi_{ki}^{L,(\sigma-1)}\}, G_{\{k\}}^{L,(\sigma+v)})) \\ &+ \hat{\Upsilon}\{G_k^{L,\sigma}, G_l^{L,(\sigma+v)}, \Omega_{kl}^{L,\sigma}, \Psi_{ki}^{L,(\sigma-1)}, \Psi_{ki}^{L,(\sigma+1)}\}] \\ &+ \Lambda(t)\cdot G_k^{R,\sigma}\end{aligned} \tag{8}$$

where $G_k^{R,\sigma}$, $G_k^{L,\sigma}$ are dynamical variables referring to the RH and LH, respectively; σ is the number of symbol's hierarchy level (for the sake of brevity, the image plate H is treated as zero-level plate G^0). The functional $\Im(G, \alpha)$ describes the internal dynamics of a single neuron, the functional $\Im\{\alpha_k, G_k^{\sigma}, G_k^{\sigma+v}, \Omega^{\sigma}, \Psi^{\sigma}\}$ describes the intra- and inter-plate interactions between neurons (for details, see [16]); α_k and τ_G are model parameters. Here, as in Eq. (4), the term Z(t)ξ(t) in (7) corresponds to the random component ("*noise*"), with Z(t) being the noise amplitude. It is presented in RH only.

It is important to stress that the *parametric modification* of those neurons that participated in forming any "information item" (image, symbol, generalized image, etc) $\alpha_k^{\sigma} \rightarrow \alpha_k^{\sigma}(\{\Psi_{ki}^{R,(\sigma-1)}\}, G_{\{k\}}^{R,(\sigma+v)})$ *does explain* the experimental effect [11] on morphological changes of the neurons which have acquired certain "skill".

The last term in Eqs. (7), (8) refers to the cross-subsystem connections Λ(t) which control the activity of each subsystem thus providing their "dialog". Here and below, it is accepted: $\Lambda = +\Lambda_0 = \Lambda^{R \rightarrow L}$, and vice-versa, $\Lambda = -\Lambda_0 = \Lambda^{L \rightarrow R}$. These connections are not trained, but should switch depending on the stage of the problem solving. The mechanism of their switching should be specified (see below).

3.3 Elementary Learning Act: The Principle of Connection Blackening

The elementary act of the new symbol formation is presented in Fig. 3.

This process corresponds to the self-organizing principle of "connection blackening" and proceeds in two steps. At the first step (Fig. 3a), an image formed by the previous-level (σ − 1) symbols in RH, after its cooperative connections Ω^R become strong ("black") enough, is delivered by the direct inter-plate connections ψ to the symbolic plate G^{σ}. Simultaneously, by the inter-subsystem connections $\Lambda^{R \rightarrow L}$, it is transferred to the "image" plate $G^{\sigma-1}$ in LH. At the second step (see Fig. 3b), new symbol is formed due to the winner-choosing process described above by Eqs. (4), (5). Semantic inter-plate connections Ψ^R are trained according to Eq. (6) up to sufficiently strong ("black") state which is defined by the condition:

$$\frac{1}{M_k^{\sigma}} \cdot \sum_{i}^{M_k^{\sigma}} \Psi_{ik}^{R\sigma} - \Psi^{thr} \geq 0; \leftrightarrow i \in \{M_k^{\sigma}\}, \tag{9}$$

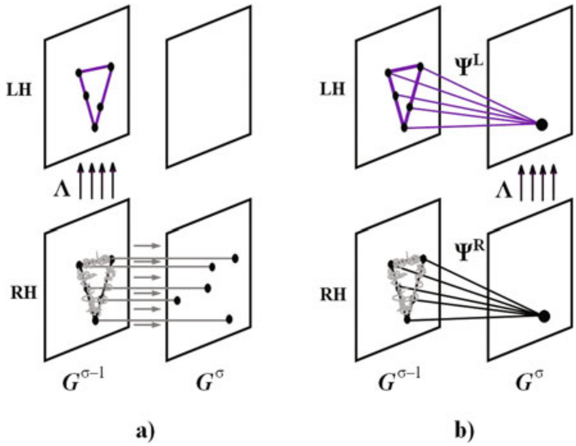


Fig. 3 Elementary learning act: **a** the "image" is created in RH at the level σ and transferred to LH; **b** its "symbol" is formed in RH (with its semantic inter-plate connections Ψ^R) and transferred to LH

where $\Psi_{ik}^{R\sigma}$ are semantic connections between *k*-th symbol at *σ*-th level and its "image" neurons at the level $\sigma - 1$, M_k^σ is the number of connections, Ψ^{thr} being the connection's threshold value (considered as the model parameter); summation proceeds over each *i*-th neuron that belong to the given image. After the condition (7) is fulfilled, the process of symbol formation is completed. Then the inter-subsystem connections Λ^{RL} should switch on to transfer it to the LH, where semantic connections $\Psi_{ik}^{L\sigma}$ are trained by the Hopfield-type rule, as in Eq. (3). Thus, the elementary learning act (or the "cog" according to terminology of Anokhin [5]) is accomplished.

This elementary act is reproduced ("replicated") at every level of the architecture, being a "*cell*" of the whole architecture presented in Fig. 2. In physics, this type of organization is called *scaling,* and the whole structure is described by the term *fractal.*

3.4 Functions of Processors at Different Levels of Hierarchy

The hierarchy levels of the architecture presented in Fig. 2 perform different functions.

Image ("Visual") Information The lowest level σ = 0 is represented by the Hopfield-type plates containing the *image* information. The plate H^0 in RH should carry the *whole* image information received by the "receptors" of a given system. This plate is responsible for *recording new images* (learning). The plate H^{typ} in LH contains images selected for storage (memorization); they are called *typical images* and play main role in recognition of already known objects.

Symbolic (Semantic) Information The next level $\sigma = 1$ is occupied by the *symbols of typical images*, which are formed in RH and transferred to LH. Each symbol does carry a *semantic* content which consists in the *decomposition* into its image. At the same level, the symbols could cooperate and create the *generalized images* ("image-of-symbols") in RH, which acquire their own symbols at the next level $\sigma + 1$. After formation of "black" semantic connections this symbol are to be transferred to LH.

Standard Symbol (Words) At the middle levels $\sigma > 1$, internal symbols (image's "name") in RH are to correlate with external standard conventional names (*words*) S_W, which are obtained as *external symbolic information* (see Fig. 2) directly by LH. After establishing the correspondence between internal and external meaning of the words, the system could express, as well as understand the external symbolic information. Thus, this very part of the scheme should embrace the Broca's area (expressing words) and Wernicke's area (understanding words).

Abstract Information At the higher levels $\sigma \gg 1$, the *abstract information* emerges. This means the infrastructure of symbols and their connections, which are not mediated by real images, i.e., the neuron-progenitors of Hopfield-type plates. Here, the generalized images are converted into the *symbol-concepts* $\boldsymbol{S_C}$ that could not be related to any concrete object (e.g., *conscience*, *infinity*, *beauty*, *number*, etc.). This information appears in the already trained system as a result of interactions of all the plates and present not the "perceptible", but "deduced" knowledge.

Thus, the system as a whole does *grow up* from the lower *image* information levels, over *semantic* information (understandable for a given individual system only), to higher levels of *abstract* information which could be verbalized and *propagated* within given society. Note that the same evolution process is typical for human beings in course of its ontogenesis.

It is important to stress that at each stage of new level formation, some information is not delivered to LH, but remains in RH as the *latent* (*hidden*) individual information for a given system. This very information could be interpreted as an *intuition* (see [15, 17]).

Note that the plates (processors) in Fig. 2 could be arranged not in parallel, but sequentially along some surface. Then the "functional zones" of NCA-architecture represents the "mirror reflection" of that presented in Fig. 1b.

4 Representation of Emotions

4.1 The Problem of Formalization of Emotions in AI

Incorporating the emotions into artificial cognitive system represents really the challenge (see [30] and refs. therein), since there is the same "explanatory gap". From the mind viewpoint (psychology), they represent *subjective self-appraisal* of the current/future state. From the "brain viewpoint", emotions are associated with objective and experimentally measured *composition of neural transmitters*, which is controlled by the sub-cortical structures (thalamus, basal ganglia, *amygdala*, etc., see Fig. 1) [7, 8]. All the variety of neurotransmitters can be sorted into two groups: the *stimulants* (like *adrenalin*, *caffeine*, etc.) and the *inhibitors* (*opiates*, *endorphins*, etc.). Under NCA, this factor could be accounted for by incorporating additional *aggregated variable* $\mu(t)$ to be the effective difference between the stimulants and inhibitors.

In psychology, the self-appraisal (emotion) is ordinarily associated with achieving a certain *goal*. Commonly, they are classified as positive and negative ones, with increasing probability of the goal attainment leads to positive emotions, and vice versa. Furthermore, any *new* (*unexpected*) thing/situation calls up *negative* emotions [12], since it requires additional efforts to hit the new goal.

According to DTI, emotions could be classified as *impulsive* (useful for generating information) and *fixing* (effective for reception). Since the process of generation requires the noise, it seems natural to associate impulsive emotions (*anxiety, nervousness*) with the *growth of noise amplitude.* Vise versa, fixing emotions could be associated with *decreasing* noise amplitude (*relief*, *delight*). Defining the living-organism goal as a *homeostasis* (calm, undisturbed, stable state), one may infer that, speaking roughly, this classification could correlate with negative and positive emotions, respectively.

4.2 Main Hypothesis on Emotion Representation in AI

Based on these reasons, one could propose the following set of hypothesis:

Proposition 1 *The influence of neurotransmitters could be accounted for by the system of equations that link the noise amplitude Z(t) with the aggregated variable μ(t) that represents virtual composition of neurotransmitters (stimulants minus inhibitors).*

Proposition 2 *The emotional reaction of human beings could be interpreted in AI systems as the time derivative of the noise amplitude, i.e., dZ(t)/dt. The absolute value of derivative dZ/dt corresponds to the degree of emotional manifestation:*

drastic change (jump) in Z(t) imitates either panic (dZ/dt > 0), or euphoria (dZ/dt < 0), and so on. Note that this value could be either positive, or negative that could be (very roughly) related to negative and positive emotions, respectively.

Proposition 3 *The same derivative should control the "dialog" between hemi-systems: increasing Z(t) (negative emotions) corresponds to activation of RH, while decreasing Z(t) (positive emotions) switches on LH activity.*

4.3 Equations for Coupling the Activity of Cortex and Subcortical Structures

The equations representing these propositions could be written in the form:

$$\frac{dZ(t)}{dt} = \frac{1}{\tau^Z} \cdot \{a_{Z\mu} \cdot \mu + a_{ZZ} \cdot (\mathrm{Z} - \mathrm{Z}_0) + F_Z(\mu, Z) - \hat{X}\{\mu, G_k^{R,\sigma}, \Psi\} + [\chi(\mu) \cdot D - \eta(\mu) \cdot \delta(t - t_{D=0})]\}, \tag{10}$$

$$\frac{d\mu}{dt} = \frac{1}{\tau^\mu} \cdot \{a_{\mu\mu} \cdot \mu + a_{\mu Z} \cdot (\mathrm{Z} - \mathrm{Z}_0) + F_\mu(\mu, Z)\}, \tag{11}$$

$$\Lambda(t) = -\Lambda_0 \cdot th(\gamma \cdot \frac{dZ}{dt}), \tag{12}$$

where φ, a, χ, η, τ are model parameters, the functional $\mathrm{X}\{\mu, \mathrm{G}_k^{R,\sigma}, \Psi\}$ refers to the process of new symbol formation presented in Fig. 3 (which should decrease $Z(t)$ value). Linear in Z and μ part in Eqs. (10), (11) provides the system's homeostasis: stationary stable state corresponds to $\{Z = Z_0, \mu = 0\}$. The functions $F_Z(\mu, Z)$ in (10) and $F_\mu(\mu, Z)$ in (11) are written to account for possible nonlinear effects (see [16, 22]).

The last term in (8) refers to processing the incoming information. Here, *D* stays for the *discrepancy* between the *incoming* and *internal* (learned and stored) information, which provokes *Z increasing*. This very situation refers to the "effect of unexpectedness", that should give rise to human's negative emotions and, according to Eq. (12), leads to activation of RH: $\Lambda(t) = -\Lambda_0 \equiv \Lambda^{L\rightarrow R}$. Vice versa, finding the solution to the problem ($D = 0$) causes rapid *decline* of Z, which corresponds to positive emotional splash and LH activation ($\Lambda(t) = +\Lambda_0 \equiv \Lambda^{R\rightarrow L}$); then RH gets the possibility to be "at rest". Thus, the model (10)–(12) seems quite reasonable and self-consistent.

5 Conclusion

Thus, it is shown that the cognitive architecture designed in analogy to the human-brain structure really provides the possibility to interpret and reproduce the peculiarities of human cognitive process under NCA. The key point of NCA is the following: an artificial cognitive system, being a complex multi-level combination of various-type neural processors, should be divided into *two subsystems*, like human brain does (two cerebral hemispheres). It is shown that one of them should necessary contain a random element (*noise*) for generation of information (creativity); it is responsible for *learning* (an analogy to the right hemisphere). Another one should be responsible for memorization and processing the *well-known* information (after learning), by analogy with the left hemisphere; this subsystem should be free of noise. It is shown that the noise-reach subsystem could provide an *intuition*, while the noise-free one is associated with *logical* thinking. Both subsystems should be linked by the cross-subsystem connections (by analogy with *corpus callosum*). Those connections should switch on depending on the stage of solving a problem.

It is shown that the human emotions, being a subjective appraisal of the current/future state, and, simultaneously, the product of the neural transmitters, are inherently embedded into the NCA-architecture. Emotions are interpreted as *dynamical variations of the noise amplitude*, and these very variations should *control the activity of two* sibsystems. Accounting for the neurophysiology, this interpretation requires including an additional variable which corresponds to the "effective" composition of neural transmitters. The system of coupled equations on the noise amplitude and the neural-transmitter variable is proposed, which provides reasonable correlation between switching the subsystem activity and the necessity of additional stimulant (or inhibitor) production.

Under NCA, the emotions are classified as *impellent* (stimulating the *generation* of information) and *fixing* (stimulating the *memorization* process) ones. Very roughly, this classification corresponds to common *negative* and *positive* emotions, respectively. Since the generation of information requires rather high level of noise, thus the increase of the noise amplitude does simulate impellent ("negative") emotions. Vice versa, a termination of the learning act, i.e., ertain "*skill*" acquiring, causes the decrease of noise that corresponds to the fixing ("positive") emotions. Thus, this idea seems reasonable and deserves further elaboration.

It is shown that the experimentally observed modifications on the gene-expression level of those neurons that are involved into skill-acquiring process [11] could be reproduced under NCA by *parametric modification* of the "trained" neurons. The latter is possible only within the concept of *dynamical* formal neuron.

Let us point out that NCA actually provides the "bridge" over the gap between physiological ("brain") and psychological ("mind") approaches to cognitive process. The concept of *conventional information*, being based on the material connections between neurons, represents the *free choice* of the neuron ensemble as a whole.

References

1. Chella, A., Mazotti, R.: Machine consciousness: a manifesto for robotics. Int. J. Mac. Conscious. **1**(1), 33–51. ISSN 1793-8430 (2009)
2. Kushiro, K., Harada, Y., Takeno, J.: Robot uses emotions to detect and learn the unknown. Biol. Inspired Cogn. Archit. **4**, 69–78 (2013)
3. Laird, J.E.: The Soar Cognitive Architecture. MIT Press (2012)
4. Samsonovich, A.: Bringing consciousness to cognitive neuroscience: a computational perspective. J. Integr. Des. Process Sci. **11**, 19–30 (2007)
5. Anokhin, K.V.: Cognitom: theory of realized degrees of freedom in brain. In: The Report Given at Fifth International Conference on Cognitive Science, Kenigsberg, June 2012. http://vk.com/Kognitivnaja_Nauka (in Russian)
6. Yakhno, V.G.: Basic models of hierarchy neuron-like systems and ways to analysis of some their complex reactions. Opt. Mem. Neuron NetW. **4**(2), 141 (1995)
7. Koziol, L.F., Budding, D.E.: Subcortical Structures and Cognition. Implications for Neurophysiological Assessment. Springer, Berlin (2009)
8. Adcock, R.A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., Gabrieli, J.D.E.: Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron **50**, 507–517 (2006)
9. Levine, J.: Materialism and qualia: the explanatory gap. Pac. Phil. Quart. **64**(4), 354–361 (1983)
10. Stirling, J., Eliott, R.: Introducing Neuropsychology, 2nd Edition (Psychology Focus). Psychology Press (2010)
11. Svarnik, O.E., Anokhin, K.V., Aleksandrov, Y.I.: Experience of a First "Whisker-Dependent Skill Affects the Induction of c-Fos Expression in Somatosensory Cortex Barrel Field Neurons in Rats on Training the Second Skill". Neurosci. Behav. Physiol. (2015). doi:10.1007/s11055-015-0135-3
12. Goldberg, E.: The New Executive Brain. Oxford University Press (2009)
13. Solso, R.: Cognitive Psychology, 6th ed. Allyn & Bacon (2006)
14. Bianki, V.L.: Parallel and sequential information processing in animals as a function of different hemispheres. Neurosci. Behav. Physiol. **14**(6), 497–501 (1984)
15. Chernavskaya, O.D., Chernavskii, D.S., Karp, V.P., Nikitin, A.P., Shchepetov, D.S.: An architecture of thinking system within the dynamical theory of information. Biol. Inspired Cogn. Arc. **6**, 147–158 (2013)
16. Chernavskaya, O.D., Chernavskii, D.S., Karp, V.P., Nikitin, A.P., Shchepetov, D.S., Rozhylo, Y.A.: An architecture of the cognitive system with account for emotional component. Biol. Inspired Cogn. Arc. **12**, 144–154 (2015)
17. Chernavskaya, O.D., Chernavskii, D.S.: A natural-constructive approach to modeling the cognitive process. J. Biophy. **61**(1), 155–169 (2016)
18. Haken, H.: Information and Self-Organization: A macroscopic approach to complex systems. Springer, Berlin (2000)
19. Progogine, I., Stengers, I.: The End of Certainty. Free Press (1997). ISBN 978-0-684-83705-5
20. Chernavskii, D.S.: The origin of life and thinking from the viewpoint of modern physics. Phys. Usp. **43**, 151–176 (2000)
21. Haykin, S.S.: Neural Networks and Learning Machines. Prentice Hall (2009)
22. Cernavskaya, O.D., Rozhylo, Y.A.: On Possibility to imitate emotions and a "sense of humor" in an artificial cognitive system. In: Proceedings of the Eighth International Conference on Advanced Cognitive Technologies and Applications COGNITIVE 2016, 20–24 Mar 2016, Rome, Italy (in press)
23. Cernavskaya, O.D.: The Cognitive Architecture within Natural-Constructive Approach. FIERCES on BICA 2016. Springer (in press)
24. Quastler, H.: The Emergence of Biological Organization. Yale University Press, New Haven (1964)

25. FitzHugh, R.: Impulses and physiological states in theoretical models of nerve membrane. Biophys. J. **1**, 445 (1961)
26. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. **79**, 2554 (1982)
27. Hebb, D.O.: The Organization of Behavior. Wiley (1949)
28. Grossberg, S.: Studies of Mind and Brain. Riedel, Boston (1982)
29. Kohonen, T.: Self-Organizing Maps. Physica-Verlag, Heidelberg (1995)
30. Hudlicka, E.: Affective BICA: Challenges and open questions. BICA **7**, 98–125 (2014)

Identity Verification Using a Kinematic Memory Detection Technique

Merylin Monaro, Luciano Gamberini and Giuseppe Sartori

Abstract We present a new method that allows the identification of false self-declared identity, based on indirect measures of the memories relating the affirmed personal details. This method exploits kinematic analysis of mouse as implicit measure of deception, while the user is answering to personal information. Results show that using mouse movement analysis, it is possible to reach a high rate of accuracy in detecting the veracity of self-declared identities. In fact, we obtained an average accuracy of 88 % in the classification of single answers as truthful or untruthful, that corresponds overall to 9.7/10 participants correctly classified as true tellers or liars. The advantage of this method is that it does not requires any knowledge about the real identity of the declarant.

Keywords Identity verification · Lie detection · Memory detection

1 Introduction

Nowadays the security concerning the identity has become a very sensitive issue. In particular, the increase of terrorist attacks in the last decades imposes the need to recognize declarants of false identity. Usually migrants from Middle East entering Europe or USA do not have any documents and personal details are frequently self-declared. Among them, a high number of terrorists giving false identities are believed to be hidden. Because terrorists move across countries using fake identities, the identity detection is now a major target in anti-terrorism [1].

M. Monaro (✉) · L. Gamberini · G. Sartori
University of Padova, Human Inspired Technolgy Research Centre, via Luzzati 4, 35122 Padua, Italy
e-mail: merylin.monaro@studenti.unipd.it

L. Gamberini
e-mail: luciano.gamberini@unipd.it

G. Sartori
e-mail: giuseppe.sartori@unipd.it

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_11

Deception is cognitively more complex than truth telling and this higher complexity reflects itself in a lengthening of the reaction times (RT) during a response [2]. According to literature, two memory detection techniques based on RT have been proposed to identify liars. These are the autobiographical Implicit Association Test (aIAT) [3] and the RT-based Concealed Information Test (RT-CIT) [4]. These techniques may be used also as tools for identity verification [5].

RT based techniques have a number of advantages compared to the traditional psychophysiological techniques to detect deception, as the polygraph [6]. First, RT are not subjected to strong individual and environmental changes, such as in the case of physiological parameters. Secondly, these techniques are inexpensive and suitable to be used on large scale. However, these techniques are not without limitations. Even though RTs are implicit measures, during the aIAT or CIT examination the lie detection purpose is explicit (overt detection of deception). Furthermore, RT based techniques only studied the latency in the response, so the liar has to check only this unique parameter to falsify the evidence. Finally, the use of these methods requires a prior knowledge about the information that has to be checked as true or false. In fact, both aIAT and CIT require that the true identity (or the true memory) is available, while in most real applications the true identity, as the migrant's case, is unknown to the examiner. This feature limits the practical application of RT based verifications, even if their efficiency is proved.

The analysis of movements during the response has already been shown to present a series of advantages, since it allows to capture the cognitive complexity in stimulus processing by the registration of a variety of indicators including not only the reaction time. Recently, researchers have shown that kinematic analysis can be used as implicit measure of the cognitive processes underlying a task [7]. Several authors, as [8–10], measured hand movements during choice tasks on a screen to understand the dynamics of a wide range of psychological processes. They described as a simple hand motion can reflects in real-time the progress of the underlying cognitive processing. Therefore, hand-motor tracking can provide a good trace of mind processes.

Because cognition is largely involved in the process of lie [11], it is reasonable to think that the analysis of hand movements can be a good implicit measure to study the cognitive mechanisms involved in lying. A first and precursive study about the kinematic as signatures of deception was presented in [12]. The authors compared motor trajectories while subjects were engaged in an instructed lie task. Participants were required to respond truthfully or lying to the presented sentences by a visual cue. Authors used the Nintendo Wii Remote to record subjects' responses. Results reported that deceptive responses could be distinguished from truthful ones on the basis of several parameters, including the motor onset time, the overall time required for responding, the trajectory of the movement and kinematic parameters such as velocity and acceleration. In Ref. [13], the authors studied mouse movements in an insurance fraud online context. Their results suggest that liars had an increasing in the distance of movements, a decreasing in the speed of movements, an increasing in the response time, and a more number of left clicks. In [14] authors proposed a pilot study to identify guilty individuals involved in specific insider threat activities. They analysed mouse movements while participants compiled an online survey similar to the Concealed Information Test (CIT). Their preliminary observations showed that guilty insiders had a different motion pattern when answering the key-item as compared to the answering of non-key-items, which was indicative of an increased cognitive activity while deceiving.

Concerning the identity verification, there are also several studies in literature that applied mouse movements analysis to biometric user authentication or identification in informatics fields [15]. However, these methods require necessarily a certain level of knowledge about the alleged user and a user-specific training, in order to be able to recognize him/her or the liar.

The goal of this work is to present a new identity check technique based on mouse movements recording, to identify false self-declared identities without knowing anything about the real identity of the declarant. This method consists in a memory detection technique, which investigates the truthful or untruthful nature of the memory for the personal information declared, using implicit measures from mouse movements. In other words, we employed kinematic analysis of the mouse movements to identify implicit signatures of deception.

2 Method

2.1 Participants

40 participants were recruited at the Department of General Psychology in Padova University. The sample consisted of 17 males and 23 females. Their average age was M = 25 (SD = 4.6), and their average education level was M = 17 (SD = 1.8). Because they use the mouse differently, left-handed subjects were excluded. All subjects agreed on the informed consent before the experiment.

2.2 Experimental Procedure

The experiment was implemented using *MouseTracker* software [16].

During the experimental procedure, participants were asked to answer 3 *yes* or *no* questions about their personal information, clicking with the mouse on the correct alternative response on the computer screen (Fig. 1 shows an example). 20 participants answered truthfully, while the others were instructed to lie about their identity according to a false autobiographical profile.

The 20 liars were instructed to learn a false identity from an Italian standard Identity Card, where a photo of the subject were attached, and which contained false personal data (an example of ID Card is reported in Appendix). After the

Fig. 1 Example of the task presented to the subjects

learning phase, participants recalled the information in the ID card for two times. Between the two recalls, they held a mathematical distracting task. On the other hand, the truth tellers performed a mathematical task and revised their real autobiographical data only once before starting the experiment.

During the experimental task, three different kinds of questions were presented to participants, in random order. *Expected questions*: 6 questions about information explicitly trained from liars during the learning and recall phases (e.g. date of birth). *Unexpected questions*: 6 questions related to the identity but not explicitly rehearsed before the experiment (e.g. age). Liars can get this information by applying a reasoning to the learned data. For example, if I know that I was born in April 1989, I can conclude that I am 26 years old. *Control questions*: 4 questions about personal characteristics that could not be denied. These are information regarding evident physical traits that cannot be hidden to the examiner, as the gender.

Each of these 16 questions was presented two times, one time the subject had to answer *yes* and in the other one the participant had to give a *no* response, for a total of 32 questions. In this way, truth tellers answered sincerely at all questions, whereas liars answered lying on *expected* and *unexpected* questions that required a *yes* response. Liar's answers to *control* questions and to *expected* and *unexpected* questions, which required a *no* response, were truthful. An example of questions is reported in Appendix.

To view each question, participants had to click on the Start button in the lower part of the screen. Then they chose the answer clicking on the response boxes positioned in the two top corners of the screen.

2.3 Data Analysis

For each answer, the motor response was recorded using *MouseTracker* software. Because each recorded trajectory have a different length, in order to permit averaging and comparison across multiple trials, each motor response was time-normalized. By default, *MouseTracker* performs a time normalization in 101 time steps using linear interpolation. Thus, each trajectory had 101 time-steps and each time-step had a corresponding x and y coordinate.

We analysed signatures of deception in terms of the shape of each movement trajectory and the location of the trajectory over time. We also quantified the trajectory properties on dimensions of velocity, stability, and direction. In particular, we collected the following features:

- *Number of errors*: number of incorrect answers.
- *Initiation time*: time between the appearance of the question and the beginning of the mouse movement.
- *Reaction time*: time from the appearance of the question to the click on the answer box.
- *Maximum deviation*: the largest perpendicular deviation between the actual trajectory and its idealized trajectory.
- *Area under the curve*: the geometric area between the actual trajectory and the idealized trajectory.
- *Maximum deviation time*: time to reach the point of maximum deviation.
- *x-flip*: number reversals of direction along the x-axis.
- *y-flip*: number reversals of direction along the y-axis.
- *X*, *Y coordinates over the time*: position of the mouse along the axis over the time. Specifically we choose to use for the analysis only Y coordinate data, for time-steps 18, 29, 30. This is because already from a preliminary visual analysis the two experimental groups clearly differed only in position of the mouse along the y-axis over the time.
- *Acceleration over the time*: acceleration of the mouse along the axis over the time. We calculated acceleration along y-axis for time intervals 18–29 and 29–30.

These features were used to train different machine learning classifiers on all subject responses.

3 Results

A preliminary visual analysis showed a significant difference in kinematic responses between liars and truth tellers. The average maximum deviation (MD) for liars is 0.33 (SD = 0.42), while for truth tellers is 0.15 (SD = 0.28). The area under the curve is wider in liars (AUC = 0.6, SD = 1.1), than truth tellers (AUC = 0.22, SD = 0.5). Figure 2 shows the average trajectory for liars and truth tellers. Furthermore, it represents the position of the mouse along x and y-axis during the response time for liars and truth tellers. In addition, liars made a greater number of errors than truth tellers (error frequency for liars = 84, error frequency for truth tellers = 7).

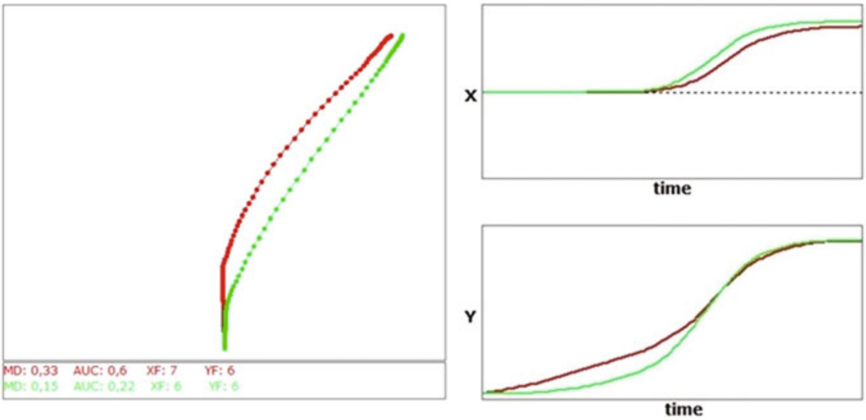


Fig. 2 *Left* average trajectory for liars (in *red*) and truth tellers (in *green*). *Right* Position of the mouse along *x* and *y* axis during the response time for liars (in *red*) and truth tellers (in *green*)

In a first step, a 10-fold cross-validation Random Forest classifier was run on all dataset (1280 stimuli). We obtained an accuracy around 90 % in the classification of the single answer as truth or lie.

Secondly, the efficiency of the classification was evaluated on 10 test sets of 10 subjects each one. The 10 test sets were extracted from the original dataset of 40 subjects using the following rules: each test set contained 5 liars and 5 truth tellers; each subject appeared in the 10 test sets a minimum of 2 and a maximum of 3 times. In this way, each test set included 320 stimuli gathered from 10 subjects. Data from the remaining 30 subjects (960 stimuli) were employed to build the model. Results for the classification of each training-test couple is reported in Table 1. Using a

Table 1 Simple logistic classifier accuracy for 10 training set and 10 test set including all stimuli

Training-test set couples	Accuracy for cross-validation on training set (%)	Accuracy in test set (%)	Number of subjects correctly classified
1	84.37	65.62	7/10
2	77.81	75.31	9/10
3	82.08	72.5	7/10
4	77.7	77.81	9/10
5	81.35	74.68	8/10
6	77.81	75.31	8/10
7	77.06	76.56	7/10
8	80	80.31	8/10
9	8.16	63.75	7/10
10	76.85	73.75	8/10
Mean	79.91	73.56	7.8/10

Table 2 Simple logistic classifier accuracy for 10 training set and 10 test set including as stimuli only *expected* and *unexpected* questions that required a *yes* response

Training-test set couples	Accuracy for cross-validation on training set (%)	Accuracy in test set	Number of subjects correctly classified
1	83.88	74.16	8/10
2	80.55	72.5	9/10
3	79.44	82.5	9/10
4	82.5	80.83	9/10
5	85.55	70.83	8/10
6	80.27	84.16	10/10
7	80	83.33	9/10
8	80.27	80	9/10
9	83.05	73.33	8/10
10	79.44	78.33	9/10
Mean	81.49	77.99	8.8/10

Simple Logistic classifier, we obtained an overall accuracy of 73.56 % in classifying a single stimulus as truthful or untruthful. From the classification of single answer as true or false, according to a majority vote system, we traced the classification of the single subject as liar or truth teller. On the single participant, we reached an average accuracy of 7.8/10 participants correctly classified as true tellers or liars, with a minimum accuracy of 7/10 and a maximum of 9/10.

We repeated this procedure for training and testing the classifier on the answers in which only truth tellers responded sincerely and liars cheating (*expected* and *unexpected* questions that required a *yes* response). 10 training sets and 10 test sets were created as above. This time, each test set included 120 stimuli gathered from 10 participants and each training set included 360 stimuli obtained from 30 participants. Classification results are shown in Table 2. Training a Simple Logistic classifier, we obtained an accuracy around 78 % in the classification of the stimuli as sincere or deceitful, which means that 8.8/10 participants were correctly classified as true tellers or liars, with an accuracy ranging from 8/10 to 10/10.

Finally, we built a model including in the training set also the answers of all 40 participants, in which both liars and truth tellers answered truthfully (*control* questions and *expected* and *unexpected* questions that required a *no* response). Each test set included the answers of 10 participants in *expected* and *unexpected* questions that required a *yes* response. Therefore, each one of the 10 training sets included 1160 stimuli, and each test set included 120 stimuli. Using a Random Forest classifier, we reached an average accuracy of 88.08 % in the classification of single answers as truthful or untruthful, that corresponds overall to 9.7/10 participants correctly classified as true tellers or liars, with a minimum accuracy of 8/10 and a maximum of 10/10. These data are reported in Table 3.

Table 3 Random forest classifier accuracy for 10 training set and 10 test set including *control* questions, *expected* and *unexpected* questions of all 40 participants in the training set

Training-test set couples	Accuracy for cross-validation on training set (%)	Accuracy in test set (%)	Number of subjects correctly classified
1	91.37	87.5	10/10
2	92.58	76.66	8/10
3	90.86	90	10/10
4	91.63	86.66	10/10
5	90.77	90.83	10/10
6	91.63	89.16	10/10
7	91.12	87.5	10/10
8	90.77	95.83	10/10
9	92.15	90.83	10/10
10	92.4	85.83	9/10
Mean	91.52	88.08	9.7

4 Conclusions

This work shows that using mouse movement analysis, it is possible to reach a high rate of accuracy in detecting the veracity of self-declared identities. The accuracy of the classification is very high not only for the single subject, but also for the single answer.

As already shown in literature [17], the presence of *unexpected questions* induce in liars an increase in cognitive load. This increase reflected itself in a different pattern of the kinematic response that became distinguishable from the truth teller pattern.

We believe that this approach can have several advantages compared to the RT based techniques mentioned above. First, kinematic indices can be recorded in a hidden way while the user interacts with the device and not being aware of what we are observing. Secondly, the detection of these indices is inexpensive, easily obtainable and does not require any equipment in addition to what the subject is already using during the interaction with the computer. This method is potentially very well adapted to the detection of deception also in the context of web, because it do not require the presence of an examiner and can be run automatically, quickly and anywhere. Furthermore, the use of mouse kinematic instead of the simple RT pushing a key on keyboard in order to record responses has a number of advantages. While button press may only permit to record RT, to use a mouse allows to capture the cognitive processes and their complexity by the registration of a large set of indicators, which include not only the reaction time. For this reason, the technique is promising also concerning resistance to countermeasures. The large number of characteristics of movement seem, in principle, difficult to control entirely via efficient countermeasures to lie detection.

Appendix

Fake ID Document for Facilitating Rehearsal of Faked Identity by Liars. The document reproduced an Italian standard Identity Card. It contains the following information: last name, first name, date of birth, city of birth, citizenship, city of residence, residence address, marital status, occupation, height, hair color, eye color.

List of Questions Presented to Subjects

Topic	Example for *yes* answer	Example for *no* answer
Control questions		
Gender	Are you female?	Are you male?
Skin color	Is your skin white?	Is your skin brown?
Hair color	Do you have blond hair?	Do you have black hair?
Citizenship	Are you an Italian citizen?	Are you a French citizen?
Expected questions		
First name	Is Alice your name?	Is Maria your name?
Last name	Is Rossi your last name?	Is Bianchi your last name?
Year of birth	Were you born in 1989?	Were you born in 1986?
Month of birth	Were you born in April?	Were you born in August?
City of residence	Do you live in Limena?	Do you live in Caserta?
Residence address	Do you live at Vespucci street?	Do you live at Marconi street?
Unexpected questions		
Age	Are you 26 years old?	Are you 23 years old?
Zodiac sign	Is Aries your zodiac sign?	Is Leo your zodiac sign?
Region of birth	Were you born in Veneto?	Were you born in Campania?
Province of birth	Were you born in Padova province?	Were you born in Caserta province?
Region of residence	Do you live in Veneto?	Do you live in Campania?
Chief town of residence region	Is Venezia the chief town of your residence region?	Is Napoli the chief town of your residence region?

References

1. The University of Texas at Austin. http://news.utexas.edu/2015/12/07/the-direct-link-between-identity-theft-and-terrorism
2. Walczyk, J.J., Igou, F.P., Dixon, A.P., Tcholakian, T.: Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. Front. Psychol. **4**, 14 (2013)
3. Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S.D., Castiello, U.: How to accuratly detect autobiographical events. Psychol. Sci. **19**(8), 772–780 (2008)
4. Verschuere, B., Ben-Shakhar, G., Meijer, E.: Memory Detection: Theory and Application of the Concealed Information Test. Cambridge University Press, Cambridge (2011)
5. Verschuere, B., Kleinberg, B.: ID-check: online concealed information test reveals true identity. J. Forensic Sci. **61** (2015)
6. Ganis, G., Keenan, J.P.: The cognitive neuroscience of deception. Soc. Neurosci. **4**(6), 465–472 (2009)
7. Freeman, J.B., Dale, R., Farmer, T.A.: Hand in motion reveals mind in motion. Front. Psychol. **2**, 59 (2011)
8. Calcagni, A., Lombardi, L.: Dynamic fuzzy rating tracker (DYFRAT): a novel methodology formodeling real-time dynamic cognitive processes in rating scales. Appl. Soft Comput. **24**, 948–961 (2014)
9. Song, J.-H., Nakayama, K.: Hidden cognitive states revealed in choice reaching tasks. Trends Cogn. Sci. **13**(8), 360–366 (2009)
10. McKinstry, C., Dale, R., Spivey, M.J.: Action dynamics reveal parallel competition in decision making. Psychol. Sci. **19**(1), 22–24 (2008)
11. Vrij, A.: Deception Detection: Current Challenges and New Approaches. Wiley, Oxford (2014)
12. Duran, N.D., Dale, R., McNamara, D.S.: The action dynamics of overcoming the truth. Psychon. Bull. Rev. **17**(4), 486–491 (2010)
13. Hibbeln, M., Jenkins, J., Schneider, C., Valacich, J., Weinmann, M.: Investigating the effect of fraud on mouse usage in human-computer interactions. In: International Conference on Information Systems (2014)
14. Valacich, J.S., Jenkins, J.L., Nunamaker, Jr., J.F., Hariri, S., Howie, J.: Identifying insider threats through monitoring mouse movements in concealed information tests. In: Hawaii International Conference on System Sciences. Deception Detection Symposium (2013)
15. Jorgensen, Z., Yu, T.: On mouse dynamics as a behavioral biometric for authentication. In: ASIACCS '11 Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, pp. 476–482 (2001)
16. Freeman, J.B., Ambady, N.: MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. Behav. Res. Methods **42**, 226–241 (2010)
17. Vrij, A., Leal, S., Granhag, P.A., Mann, S., Fisher, R.P., Hillman, J., Sperry, K.: Outsmarting the liars: the benefit of asking unanticipated questions. Law Hum. Behav. **33**(2), 159–166 (2009)

Emotiv-Based Low-Cost Brain Computer Interfaces: A Survey

Naveen Masood and Humera Farooq

Abstract Within the field of Brain Computer Interfacing (BCI), Electroencephalography (EEG) is the most widely applied modality. But most of the EEG based BCI applications use expensive sensors for capturing brain data. In order to make these systems accessible to end-user, it is quite necessary to have low-cost alternatives. The Emotiv EPOC is one of the inexpensive EEG devices that has been increasingly employed. Although the headset has limitations related to signal quality but it is gaining popularity in BCI researches. In this paper, a detailed review of Emotiv based BCI systems is presented along with its comparison with medical grade EEG devices. Classification algorithms and preprocessing techniques used with these systems are also discussed. Its performance is evaluated based on different factors including subjects, stimuli and specific nature of the application. The paper is concluded with the discussion of present challenges and future research possibilities for Emotiv based BCI applications.

Keywords Emotiv · Brain computer interfaces · Electroencephalography · Signal processing · Classification

1 Introduction

A brain-computer interface (BCI) is defined as a communication system for translating the signals from brain of a person into commands interpretable by a machine or a computer [1]. Within the field of BCI, electroencephalography

N. Masood (✉)
Electrical Engineering Department, Bahria University,
13 National Stadium Road, Karachi, Pakistan
e-mail: navinza@hotmail.com

H. Farooq
Computer Science Department, Bahria University,
13 National Stadium Road, Karachi, Pakistan
e-mail: humera.farooq@bimcs.edu.pk

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_12

Fig. 1 Emotiv EPOC headset [2]

(EEG) is the most widely applied modality. But most of the applications are designed using expensive EEG sensors. In order to make these BCI systems accessible to end-user for home use or outside the lab applications, it is quite necessary to have low-cost alternatives with increased portability. The Emotiv EPOC headset is one of the inexpensive EEG devices that has been increasingly employed in BCI applications based on consumer-grade sensors. Although the Emotiv headset does have limitations related to the signal quality and number of electrodes but it is gaining popularity in BCI researches that are focused towards low-cost and outside the lab applications (Fig. 1).

The paper summarizes the results of a survey of different BCI systems using Emotiv as main device for EEG signal acquisition. A detailed review of Emotiv based BCI systems is presented along with its comparison with medical grade EEG devices. The classification algorithms and preprocessing techniques used with these systems are also discussed. Its performance is evaluated based on different factors including subjects, stimuli and specific nature of the application.

2 Non-medical Applications

Because of its accessibility and portability to consumers and researchers, Emotiv EPOC has been employed in a variety of BCI applications. But due to its limited signal quality, the applications usually cover non-medical domain although few studies have been conducted for medical purposes as well. Following five categories of non-medical applications have been identified where Emotiv is used for capturing brain signals.

2.1 Device Control

One of the major reasons for conducting researches in BCI domain is desperate need to give patients or users who lack full control of their limbs access to control different devices and communication systems. Keeping these scenarios in consideration, such users may be facilitated from a device even it bears limited accuracy, efficiency or speed. Emotiv has been used in designing of such applications. Using emotive various device control applications have been designed including control of wheelchair for disabled persons [3, 4], any input device of computers [5], to control in-vehicle systems or drive a car [6, 7].

2.2 Gaming

Emotiv EPOC headset has been increasingly employed for gaming purpose as well. In such applications signals from user's brain are used to play the game and operate the controls. Sourina and Liu designed a gaming application based on Emotiv employing Steady State Visual Evoked Potential (SSVEP) as electrophysiological signal. Along with this, the performance of Emotiv is compared with the research grade EEG sensor and the study suggests that the satisfactory results are obtained from Emotiv [11].

Bernays and Mone have worked on an adaptable 3D video game named as Lost in the Dark: Emotion Adaption. It makes use of player's emotions as input to alter and adjust the gaming controls and environment. Emotiv headset is used here not only to record brain signals but facial expressions and head movement also [13]. Few gaming applications using Emotiv are mentioned in Table 1.

Table 1 List of Emotiv based studies categorized on the basis of type of application

Device control	Gaming	Emotion detection	User brain state detection	Robots
i. Controlling a Wheelchair Fattouh et al. [3], Li et al. [4] Chowdhury and Shakim [8] Bahri et al. [9] Abiyev et al. [10] ii. Controlling a computer input device Lievesley et al. [5] iii. Controlling an in-vehicle system Nisar et al. [6], Cernea et al. [7]	Sourina and Liu [11], Thomas et al. [12], Bernays et al. (2012) [13], Qiang, Sourina et al. [14]	Jatupaiboon et al. [15], Liu et al. [16], Pham and Tran [17], Aspinallet al. [18], Wang [19], Fattouh et al. [3]	Wang et al. [20], Purawijaya and Fitri [21], Ben Dkhil et al. [22], Ekanayake et al. [23]	Szafir and Signorile [24], Vourvopoulos and Liarokapis [25], Vourvopoulos and Liarokapis [26], Grude et al. [27], Guneysu and Akin [28]

2.3 Emotion Detection

Human emotion refers to a complex psychological state comprised of three components i.e. user experience, his physiological response along with behavioral or expressive reaction [29]. Different categories of emotions are disgust, pride, satisfaction, anger etc. [30]. Various studies have been conducted to find how the EEG signals correlate to human emotions. Emotiv has also been used for this purpose. Pham, Duy et al. have used Emotiv to capture EEG data while users are watching movies to induce emotions. Oscillatory brain rhythms with different frequency bands filtered from the recorded brain signals are used as input to different machine learning classifiers [17].

2.4 User Brain State Detection

While performing any task, state of brain of user can be analyzed using BCI devices. Emotiv is used for this purpose also. Wang et al. developed a virtual driving mechanism to perform driving experiments for the collection of subjects' brain data for mental fatigue. Technique of Wavelet-packets transform (WPT) was applied for continuous feature extraction [20]. McMahan and Parberry compared three different engagement indices during various video game modalities using the Emotiv device. From the results, it is concluded that Emotiv can be used to measure a player's varying levels of engagement as they play a video game [31].

Drowsiness is one of the major reasons behind road accidents especially while driving on motorways and highways. Ben Dkhil, Neji et al. worked on a drowsiness detection system based not only eye blinking but physiological signals also. A smart video camera is used to capture face images and eye blinks and to record the brain signals Emotiv has been used [22].

2.5 Robots

One of the major application of BCI systems is to control robots by means of EEG brain signals. For this area also, Emotiv has successfully been employed. Vourvopoulos and Liarokapis worked on a project for controlling a robot in both the real and virtual world. The whole set up is performed with two prototypes based on the headset type used. One is the Neurosky headset that has been tested with 54 users. Other one is performed with Emotiv EPOC headset. Results indicate that using commercial grade headsets, robot can be navigated effectively [25].

Table 1 lists the studies and researches categorized on the basis of type of applications mentioned above.

Table 2 Signal processing algorithms considered in Emotiv based BCI applications

Signal processing algorithms	Research studies
PCA (principal component analysis)	Elsawy et al. [32], Turnip et al. [33]
ICA (independent component analysis)	Turnip et al. [33]
CSP (common spatial pattern)	Bialas and Milanowski [34]
Wavelet transform	Abdalsalam [35]

3 Signal-Processing Methods Used in Emotiv-Based BCIs

Signal processing algorithms are applied on EEG signals to remove artefacts. Emotiv based BCI systems have employed different approaches for pre-processing. In Table 2, we have categorized the research papers according to the signal processing techniques and algorithms used.

4 Feature Classification Algorithms Used in Emotiv-Based BCIs

In order to identify different brain activity patterns produced by a user during any BCI experiment, machine learning algorithms for classification are mostly applied. To achieve maximum performance, suitable algorithm must be selected so that it could aim at correct estimation of the class of data represented by the feature vector. Table 3 lists the classification algorithms used in design and development of Emotiv based BCI systems.

Table 3 Classification algorithms considered in Emotiv based BCI applications

Classification algorithms	Research studies
SVM (support vector machine)	Zhang et al. [36], Vamvakousis and Ramirez [37], Wang [19], Pham and Tran [17]
LDA (linear discriminant analysis)	Wang et al. [38], Duvinage et al. [39], Vamvakousis and Ramirez [37]
Decision tree	None (to the best of our knowledge)
KNN	Pham and Tran [17], Mampusti et al. [40]
Naïve bayes	Pham and Tran [17]
Ada boost	Pham and Tran [17]
Multi layer perceptron	Abdalsalam et al. [35]

5 Discussion

Emotiv offers a cost-effective solution for design and development of BCI applications. It is quite necessary while having the advantage of low cost, Emotiv based applications could not be compromised on performance and quality. To address this issue, researchers have conducted studies by comparing Emotiv with other medical or research grade devices. Badcock, Mousikou et al. tested if auditory Event Related Potentials (ERPs) measured using Emotiv are comparable to those by a medical grade widely-used EEG system from Neuroscan. The study suggests that the consumer grade EEG system may prove a valid alternative for medical grade Neuroscan system for recording late auditory ERPs over the frontal cortices [41]. Liu conducted a research based on SSVEP physiological signal using Emotiv headset. Based on video stimuli, SSVEP are recorded from Emotiv. In this study, canonical correlation analysis (CCA) is used for feature extraction. Furthermore, the performance of g.tec EEG equipment and Emotiv EPOC is also compared. The classification accuracy and Information transfer rate (ITR) of g.tec are 94.79 ± 1.94 % and 35.66 ± 5.71 bits/min while for Emotiv, accuracy is 82.99 ± 4.98 % and the ITR is 28.06 ± 6.45 bits/min [42]. Table 4 lists some of the studies with classification accuracies obtained with Emotiv based sensor. Some studies are conducted specifically for comparison of Emotiv with other devices. From the table, it is evident that the low cost Emotiv give satisfactory results as the classification accuracy is not significantly degraded using this equipment. Other than classification percentage, other performance parameters are also used for making such comparisons like reaction time, information transfer rate, *p*-value etc. Due to limitation of length, these parameters are not covered in this paper.

5.1 Limitations for Emotiv Based BCI Applications

BCI applications have general limitations like subject dependency, experiment scenario, research paradigm etc. for each type of EEG equipment whether consumer

Table 4 Classification accuracies obtained using Emotiv

Application	Classification accuracy with	
	Emotiv (%)	Other medical grade device (%)
Visual stimulator [42]	82.99	94.79 (gtec)
Controlling video game [43]	82–84	87–89 (IMEC)
P300 speller [32]	86.29	–
Diagnosis of major depressive disorder [38]	89.66	–
Emotion detection [15]	75	–
SSVEP visual stimulator [44]	76.6	–

or research grade. Therefore in case of Emotiv based BCIs, these limitations also exist. As in the study of Van Vilet in which it is observed that the tendency to control a game mainly depends on the subject rather than the device. During out of the lab public event, using EPOC headset, 36 % users achieved good accuracy for game control while 52 % found it difficult and challenging whereas 12 % users could not achieve any control [43]. Similarly, Lin et al. have worked on SSVEP based visual stimulator system such that subjects walking on the treadmill are provided with Emotiv EPOC headest to record their brain activity with different speeds of treadmill. Highest accuracy 76.7 % is achieved with standing position while accuracy decreases with increasing speed. This is the general limitation for BCI applications as the best results are produced when the subjects are in still position [44]. Jatupaiboon et al. found that using emotiv, pair of channels at temporal region produces better result than the other regions for the happiness detection system [15].

Liu et al. have implemented SSVEP based application using Emotiv EPOC. As per the research, to connect Matlab with Emotiv using existing software is difficult and to some extent little inconvenient. In future, extensive work could be performed to provide a better solution for a stable connection between Matlab and EPOC [42].

Hariston et al. worked on performance comparison of three wireless EEG sensors including B-Alert X10, EPOC and Quasar's dry sensor with conventional wired BioSemi's ActiveTwo EEG sensor. The study elicited that some subjects found lack of comfort with EPOC since the weight of electrodes is focused on certain locations of the scalp instead of being distributed all over the scalp surface. So, the problem for uneven weight distribution is observed in EPOC [45].

In order to compare Emotiv with other devices, one of the major issues arises due to inconsistency and difference of electrode placement using different sensors. In order to conduct true comparison, except the EEG sensor, all other parameters should ideally be uniform especially the electrode placement [45].

6 Conclusion

Major objective of this paper is to bring together pieces of information from research studies to summarize the situation and give beginners and researchers in this field a preliminary but clear and concise overview of where we are standing for Emotiv based BCI systems. In the paper, a detailed review of Emotiv based BCI systems is presented along with its comparison with medical grade EEG devices. The classification algorithms and preprocessing techniques used with these systems are also discussed. Its performance is compared based on classification accuracy. Five non-medical applications have been identified where Emotiv is being employed successfully. In a BCI context, the results they obtained, have been analysed and compared with other medical grade devices. Although limited number of studies are conducted so far to make this comparative analysis. One major difficulty that generally encountered in such comparative studies concerns the lack of published comparisons between algorithms, classifiers, devices etc. Ideally, to compare EEG sensors, the BCI experiments should be performed within the same context, that is with the same subjects, using the same feature detection technique and the same protocol for pre-processing.

The signal processing and classification techniques that are not applied so far in Emotiv based systems could be explored. Currently, this is one of the crucial problems for for BCI research and not specifically for Emotiv based systems. Based on the studies covered in this paper, it is concluded that although Emotiv does not perform as effectively as research grade equipment but it still offers the option to provide a user's brain wave signature.

Acknowledgments This work is funded by Higher Education Commission (HEC), Pakistan and is being conducted and supervised under the 'Intelligent Systems and Robotics' research group at Computer Science (CS) Department, Bahria University, Karachi, Pakistan.

References

1. Lotte, F.: Study of electroencephalographic signal processing and classification techniques towards the use of Brain-computer interfaces in virtual reality applications, INSA de Rennes (2008)
2. Emotiv.: Emotiv. 2015. Available from: https://emotiv.com/
3. Fattouh, A., Horn, O., Bourhis, G.: Emotional BCI control of a smart wheelchair (2013)
4. Li, M., et al.: An EEG based control system for intelligent wheelchair. In: Applied Mechanics and Materials. Trans Tech Publ (2013)
5. Lievesley, R., Wozencroft, M., Ewins, D.: The Emotiv EPOC neuroheadset: an inexpensive method of controlling assistive technologies using facial expressions and thoughts? J. Assist. Technol. **5**(2), 67–82 (2011)
6. Nisar, H., et al.: Analysis of brain activity while performing cognitive actions to control a car. In: The 15th International Conference on Biomedical Engineering. Springer (2014)
7. Cernea, D., et al.: Controlling in-vehicle systems with a commercial eeg headset: performance and cognitive load. In: OASIcs-OpenAccess Series in Informatics. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2012)
8. Chowdhury, P., Shakim, S.K.: Optimizing cognitive efficiency of emotiv EPOC and controlling wheelchair through it. BRAC University (2014)
9. Bahri, Z., Abdulaal, S., Buallay, M.: Sub-band-power-based efficient brain computer interface for wheelchair control. In: World Symposium on Computer Applications and Research (WSCAR). IEEE (2014)
10. Abiyev, A.H., et al.: Brain based control of wheelchair. In: Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (2015)
11. Sourina, O., Liu, Y.: EEG-enabled affective applications. In: Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII). IEEE (2013)
12. Thomas, K.P., Vinod, A., Guan, C.: Enhancement of attention and cognitive skills using EEG based neurofeedback game. In: 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE (2013)
13. Bernays, R., et al.: Lost in the dark: emotion adaption. In: Adjunct Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology. ACM (2012)

14. Qiang, W., Sourina, O., Khoa, N.M.: A Fractal Dimension Based Algorithm for Neurofeedback Games. CGI 2010
15. Jatupaiboon, N., Panngum, S., Israsena, P.: Real-time EEG-based happiness detection system. Sci. World J. **2013** (2013)
16. Liu, Y., Sourina, O., Nguyen, M.K.: Real-time EEG-based human emotion recognition and visualization. In: 2010 International Conference on Cyberworlds (CW). IEEE (2010)
17. Pham, T.D., Tran, D.: Emotion recognition using the emotiv epoc device. In: Neural Information Processing. Springer (2012)
18. Aspinall, P., et al.: The urban brain: analysing outdoor physical activity with mobile EEG. Br. J. Sports Med. **49**(4), 272–276 (2015)
19. Wang, J.: A Framework for the Integration of the Emotiv EEG System into the NeuCube Spiking Neural Network Environment for Robotics Control. Auckland University of Technology (2014)
20. Wang, F., et al.: Design of driving fatigue detection system based on hybrid measures using wavelet-packets transform. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE (2014)
21. Purawijaya, D.A., Fitri, L.L.: Evaluation of blue light exposure to beta brainwaves on simulated night driving. In: The 5th International Conference On Mathematics And Natural Sciences. AIP Publishing (2015)
22. Ben Dkhil, M., et al.: A new approach for a safe car assistance system. In: 2015 4th International Conference on Advanced Logistics and Transport (ICALT). IEEE (2015)
23. Ekanayake, H.B., et al.: Comparing expert driving behavior in real world and simulator contexts. Int. J. Comput. Games Technol. **2013**, 2 (2013)
24. Szafir, D., Signorile, R.: An exploration of the utilization of electroencephalography and neural nets to control robots. In: Human-Computer Interaction–INTERACT 2011, pp. 186–194. Springer (2011)
25. Vourvopoulos, A., Liarokapis, F.: Robot navigation using brain-computer interfaces. In: 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE (2012)
26. Vourvopoulos, A., Liarokapis, F.: Brain-controlled NXT Robot: Tele-operating a robot through brain electrical activity. In: 2011 Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES). IEEE (2011)
27. Grude, S., et al.: Controlling mobile Spykee robot using Emotiv neuro headset. In: Control Conference (CCC), 2013 32nd Chinese. IEEE (2011)
28. Guneysu, A., Akin, H.: An SSVEP based BCI to control a humanoid robot by using portable EEG device. In: 2013 35th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC). IEEE (2013)
29. Hockenbury, D.H., Hockenbury, S.E.: Discovering Psychology. Macmillan (2010)
30. Ekman, P.E., Davidson, R.J.: The Nature of Emotion: Fundamental Questions. Oxford University Press (1994)
31. McMahan, T., Parberry, I., Parsons, T.D.: Evaluating Electroencephalography Engagement Indices During Video Game Play (2015)
32. Elsawy, A.S., et al.: Performance analysis of a principal component analysis ensemble classifier for Emotiv headset P300 spellers. In: 2014 36th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC). IEEE (2014)
33. Turnip, A., Setiawan, I.R., Junaidi, E.: An experiment of ocular artifacts elimination from EEG signals using ICA and PCA methods. J. Mech. Electr. Power, Veh. Technol. **5**(2), 129–138 (2014)
34. Bialas, P., Milanowski, P.: A high frequency steady-state visually evoked potential based brain computer interface using consumer-grade EEG headset. In: 2014 36th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC). IEEE (2014)

35. Abdalsalam, M.E., et al.: Mental task motor imagery classifications for noninvasive brain computer interface. In: 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS). IEEE (2014)
36. Zhang, Y., Zhu, X., Luo, Y.: EMG assisted EEG intelligent wheelchair control system based on multi-class SVM. J. Central S. Univ. (Science and Technology) S2 (2013)
37. Vamvakousis, Z., Ramirez, R.: Towards a low cost mu-rhythm based BCI. In Proceedings of the Fifth International Brain-Computer Interface Meeting (2013)
38. Wang, D., et al.: Auditory evoked potentials in patients with major depressive disorder measured by Emotiv system. Bio-Med. Mater. Eng. **26**(s1), 917–923 (2015)
39. Duvinage, M., et al.: Performance of the Emotiv Epoc headset for P300-based applications. Biomed. Eng. Online **12**(1), 56 (2013)
40. Mampusti, E.T., et al.: Measuring academic affective states of students via brainwave signals. In: 2011 Third International Conference on Knowledge and Systems Engineering (KSE). IEEE (2011)
41. Badcock, N.A., et al.: Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. PeerJ **1**, e38 (2013)
42. Liu, Y., et al.: Implementation of SSVEP based BCI with Emotiv EPOC. In: 2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS). IEEE (2012)
43. Van Vliet, M., et al.: Designing a brain-computer interface controlled video-game using consumer grade EEG hardware. In: Biosignals and Biorobotics Conference (BRC), 2012 ISSNIP. IEEE (2012)
44. Lin, Y.-P., Wang, Y., Jung, T.-P.: Assessing the feasibility of online SSVEP decoding in human walking using a consumer EEG headset. J. Neuroeng. Rehabil. **11**(1), 1 (2014)
45. Hairston, W.D., et al.: Usability of four commercially-oriented EEG systems. J. Neural Eng. **11**(4), 046018 (2014)

Part III
Physiological Monitoring and Interaction

Test-Retest Stability of EEG and Eye Tracking Metrics as Indicators of Variations in User State—An Analysis at a Group and an Individual Level

Jessica Schwarz and Sven Fuchs

Abstract EEG- and eye tracking metrics have been investigated for their potential as indicators of user state in a variety of studies. However, their stability over time has rarely been assessed and findings are reported predominantly on a group level. In this paper, we report a test-retest analysis that aimed to investigate—at group and individual levels—the temporal stability of fixation duration, pupil dilation, and two built-in metrics from the Emotiv EPOC EEG sensor, namely Engagement and Frustration. The retest confirmed the temporal stability of most physiological metrics at the group level. But analysis at an individual level revealed that outcomes differ strongly between and also within individuals from test to retest. The divergent results between individual and group level illustrate that group level findings are of limited value for applications such as adaptive systems requiring individual user state diagnosis.

Keywords Adaptive systems · User state assessment · Temporal stability · Test-retest analysis · Physiological measures · EEG · Eye tracking

1 Introduction

Research has demonstrated that physiological measures can serve as sensitive indicators of various cognitive and affective states of the user (c.f. [1–3]). Physiological user state assessment appears to be particularly valuable in adaptive systems design. Adaptive technical systems dynamically adapt their behavior to the current state of the human operator in order to mitigate critical user states and performance decrements. This requires user state to be measured continuously and in real-time which is supported by most physiological measures (e.g. heart rate,

J. Schwarz (✉) · S. Fuchs
Fraunhofer FKIE, Fraunhoferstr. 20, 53343 Wachtberg, Germany
e-mail: Jessica.Schwarz@fkie.fraunhofer.de

S. Fuchs
e-mail: Sven.Fuchs@fkie.fraunhofer.de

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_13

electrodermal responses, brain activity, facial skin temperature, respiration, pupillary response, and eye movements). Furthermore, with technology pushing forward, increasingly unobtrusive physiological sensors are emerging (e.g. remote eye trackers and low-cost EEG) that often ship with built-in metrics and classifiers for various user states (e.g. attention, workload and emotional states). Hence, physiological measures may also prove useful when user state must be assessed in a multidimensional manner, i.e. considering various user states and the interactions between them to achieve more holistic diagnostics for adaptive systems design (cf. [4]).

Experimental studies that evaluated physiological measures as candidates for user state assessment have mainly focused on the validity and sensitivity of such metrics. However, their stability[1] and reliability has rarely been assessed. Faulstich et al. [5] analyzed the temporal stability of several physiological measures in relation to physical and mental stressors. The retest was conducted two weeks after the first assessment. They found that most physiological measures provided moderate to high test-retest correlations (mostly ranging between $r = 0.4$ and $r = 0.8$) when absolute baseline and test values were considered. However, difference scores between baseline and test conditions were not found to have adequate reliability.

Tomarken [6] states in a review of psychophysiological research that available findings on the temporal stability of psychophysiological measures indicate substantial variability across measures. He remarks that some measures demonstrated high levels of stability while others yielded disappointing results. This variability can also be observed in the study of Faulstich et al. [5].

It has to be considered, however, that outcomes concerning the validity and stability/reliability of assessment methods are usually reported at a group level. As adaptive systems need to diagnose critical user states of every user individually, group level findings may be of limited value. From this perspective, it is critical that an assessment method can provide a valid and stable user state diagnosis at an individual level.

In an experimental study ($N = 12$) we evaluated the efficacy of two classifiers from the consumer-oriented Emotiv EPOC EEG sensor and two metrics derived from a Tobii X120 remote eyetracker for multidimensional user state assessment [7]. Group-level results showed the Engagement metric of the Emotiv EEG to be sensitive to variations in workload while the validity of Emotiv’s Frustration metric to diagnose the emotional state of frustration could not be confirmed in our experimental paradigm. Results for the eye tracking metrics supported the sensitivity of pupil dilation and fixation duration to workload variations. At an individual level, however, we found outcomes of all four metrics to be highly user-specific, indicating that certain metrics were strong predictors of performance for some subjects but not for others.

[1] In accordance with Tomarken [6], we refer to the term stability for test-retest analyses across sessions and reliability for test-retest-analyses within sessions.

To investigate whether these results could be replicated, a retest was conducted one year later involving ten participants that had also participated in the first test. The following sections describe method and results of this test-retest analysis and draw implications for future research.

2 Method

2.1 Participants and Experimental Task

10 participants (8 m/2f) aged 19–39 (M = 31.4; SD = 6.8) performed a simplified simulated flight control task. Participants were not familiar with air traffic control or air space monitoring. The primary task was to command air tracks to ensure they abide to air safety rules (i.e., stay within their assigned air spaces and keep sufficient distance from each other and external non-controllable traffic passing the air space). Imminent rule violations were indicated by alarms and visual warning icons. A secondary task was used to simulate the mental demands of listening to radio communications and calculation of headings. It consisted of basic arithmetic problems (e.g. 260 + 70–8) that were presented in English via headphones every 45 s. Users entered their responses into a text field. A scoring system was implemented that rewarded error-free operation and correct answers to math problems by adding points to the score. Point deductions occurred for rule violations (air track passing the boundary or near collision of two tracks) and for wrong or no answers to math problems. The current score as well as point increases and deductions were visible during the test in order to raise the motivation of participants Fig. 1.

2.2 Experimental Design

The experimental tests were conducted in a within-subjects design that modulated user state by systematically manipulating two environmental conditions known to affect or impair workload and frustration of users: To modulate workload, the number of air spaces to be monitored were varied between two (low task load) and five (high task load; cf. [8]) between conditions. All air spaces were the same size and contained two controllable air tracks. Concerning the modulation of frustration, literature suggests that frustration can be induced by ignoring operator commands (e.g. [9, 10]). Thus, the second factor addressed the cooperation of controllable air tracks. In the high cooperation condition, air tracks always followed instructions. For low cooperation, 30 % of operator commands were ignored. A third factor (presence versus absence of loud noise) was omitted in the retest, as it was not related to significant changes in user state in the first test. This factor was originally varied between two separate sessions (one with and one without noise). To ensure

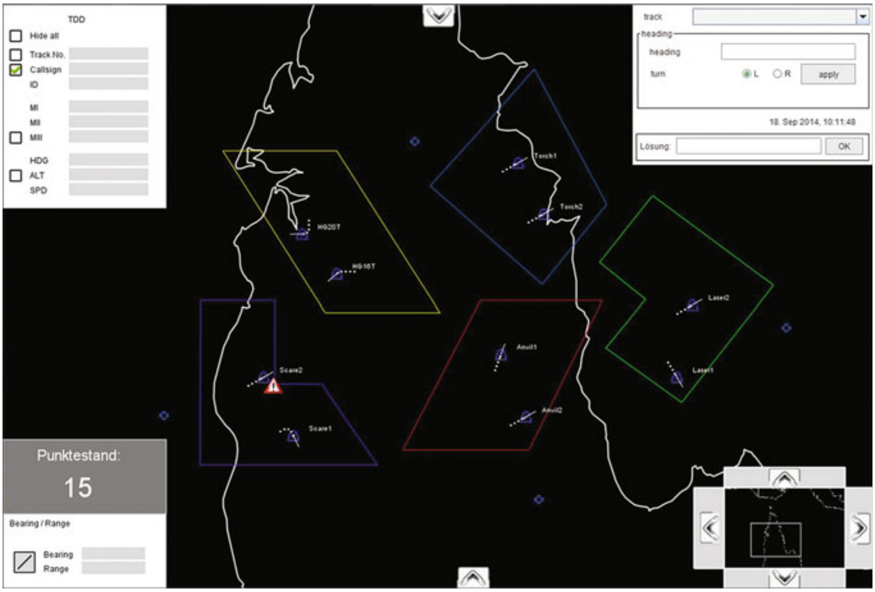


Fig. 1 User interface of the experimental task

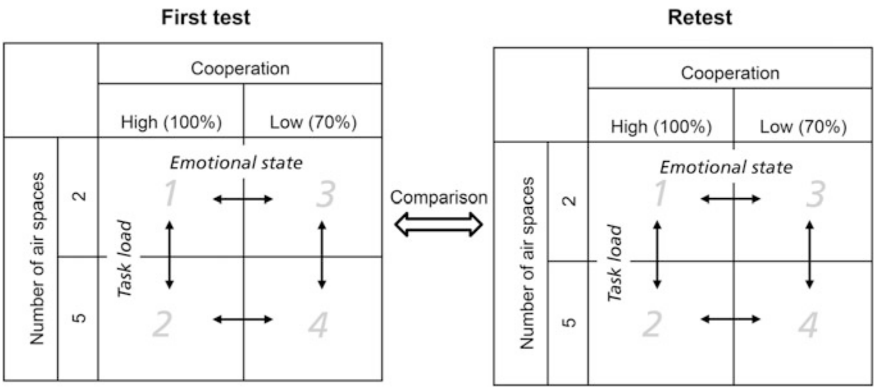


Fig. 2 Experimental design

comparability only those sessions that were identical with the retest were considered in the test-retest analysis.

A systematic combination of both experimental factors results in four experimental conditions illustrated in Fig. 2. Latin Square rotation was used to counterbalance learning or order effects within each test. In the retest, participants completed the experimental conditions in the same order as they did in the first test.

2.3 Assessment Methods

For physiological assessments the Emotiv EPOC headset and the Tobii X120 remote eye tracker were employed in both tests. The Emotiv EPOC device is a low-cost wireless neuroheadset that comprises 14 felt-based sensors plus 2 references that are located on the scalp according to the international 10–20 system. Emotiv's proprietary analysis software provides four distinct affective state classifiers—Engagement, Excitement (short-term and long-term), Frustration and Meditation—that can be obtained in real-time through the Software Development Kit (SDK). In our study, we focused on the Engagement and Frustration metrics. According to the manufacturer's descriptions, Engagement may prove useful for the assessment of workload and attention. And as its name suggests the metric of Frustration is supposed to reflect the emotional state of frustration. In the following, these metrics are referred to as Emotiv Engagement and Emotiv Frustration.

Concerning the metrics derived from the Tobii X120, we chose to focus on fixation duration and pupil dilation. Literature indicates that fixation duration is sensitive to variations in task load. However, the direction of this relationship seems to be highly task dependent. Some researchers (e.g. [11–13]) found fixation duration to increase with higher (mostly cognitive) task load while others (e.g. [14, 15]) observed a negative relationship between fixation duration and visual task complexity. In congruence with the latter, fixation duration in our first experiment was significantly lower in the conditions of high (mostly visual) task load.

Regarding pupil dilation, literature indicates an increase in pupil dilation as a function of cognitive load (e.g. [16–19]). A review by Kramer [20] found pupillary changes to be sensitive to perceptual, cognitive, and response-related processing demands in a variety of tasks. In our experiment we also observed a significant increase in pupil dilation in conditions of high task load.

Albeit not reported in this paper, user state was also assessed by self-rating methods in order to validate the physiological assessments. NASA-TLX [21] was used for subjective workload assessment with the subscale of frustration for frustration evaluation, and the Self-Assessment Manikin (SAM; [22]) for a rating of the emotional state. Additionally the total score achieved in each condition was used as a measure of performance.

2.4 Set-up and Procedure

Figure 3 shows the experimental set-up comprised of a 24-in. monitor to display the user interface, keyboard, and mouse as input devices, and a Tobii X120 eye tracker placed under the monitor. Participants wore the Emotiv EPOC headset and in-ear earphones. Physiological sensors were fitted and calibrated first, followed by an individual characteristics questionnaire, a baseline recording of EEG metrics, and a

Fig. 3 Experimental set-up

training scenario. Subsequently, four 10-min task runs were performed (one for each experimental condition), followed by NASA-TLX and SAM questionnaires.

2.5 Data Preparation

Data preparation included aggregation and normalization of values of the Emotiv and Tobii logfiles. As findings of Faulstich et al. [5] indicate, difference scores (task score minus baseline score) are not a good option to correct for individual differences in response magnitude. Hence, based on recommendations of Tomarken [6] we used z-standardization as an alternative way to normalize measures.

Test-retest correlations were calculated by using the z-standardized mean values of Emotiv Engagement, Emotiv Frustration, fixation duration, and pupil dilation which were obtained from every participant for each of the four experimental conditions. Due to poor signal quality of the Emotiv metrics in some test runs, four participants were excluded from the test-retest analysis. Hence, results on Emotiv metrics are based on a sample of N = 6. Concerning the eye tracking metrics, only one participant had to be excluded due to poor tracking quality, leading to a sample of N = 9.

3 Results

Test-retest correlations of the four physiological measures (Emotiv Engagement, Emotiv Frustration, fixation frequency and pupil dilation) were analyzed at group and individual levels. Additionally, we compared correlations between the physiological measures and performance in both tests to investigate whether the efficacy of physiological metrics to indicate changes in performance remains stable over time.

3.1 Test-Retest Stability at the Group Level

Table 1 shows the test-retest correlations calculated across all experimental conditions and for each experimental condition separately. For pupil dilation, fixation duration, and Emotiv Frustration, the test-retest correlations are above r = 0.6 when all conditions are included.

These outcomes are in a similar range as the test-retest correlations that Faulstich et al. [5] obtained in their study. Emotiv Engagement, however, did not sufficiently correlate with the original results; but when considering the test-retest correlations for each experimental condition, the Emotiv Engagement metric reaches adequate stability in at least three of four experimental conditions (r > 0.5). The other metrics show similar outcomes. Overall, results indicate that test-retest correlations differ quite strongly between experimental conditions (e.g. correlations for Emotiv Frustration are ranging from r = 0 to 0.9).

Table 1 Test-retest correlations of EEG- and eye tracking metrics at a group level

	Emotiv engagement	Emotiv frustration	Fixation duration	Pupil dilation
All conditions	0.36	0.63**	0.63**	0.64**
2 Areas and high cooperation	0.59	0.93**	0.67	0.16
2 Areas and low cooperation	−0.30	−0.03	0.75*	0.67
5 Areas and high cooperation	0.84*	0.44	0.59	0.57
5 Areas and low cooperation	0.86*	0.71	0.12	0.86**

*significant ($p < 0.05$)
**highly significant ($p < 0.01$)

3.2 Test-Retest Stability at the Individual Level

Test-retest correlations at an individual level were based on each participant's mean values obtained from the four experimental conditions in test and retest. Figure 4 is composed of four bar charts illustrating the individual test-retest correlations for each physiological measure.

As Fig. 4 shows, most correlations are positive and at a moderate to high level. However, results also indicate that the stability varies between individuals. Interestingly, individual stability is not consistent across measures: participants that show a high test-retest correlation on one measure do not necessarily show high correlations on the other measures. As an example, VP09 shows the highest test-retest correlation for fixation duration but a negative test-retest correlation for pupil dilation.

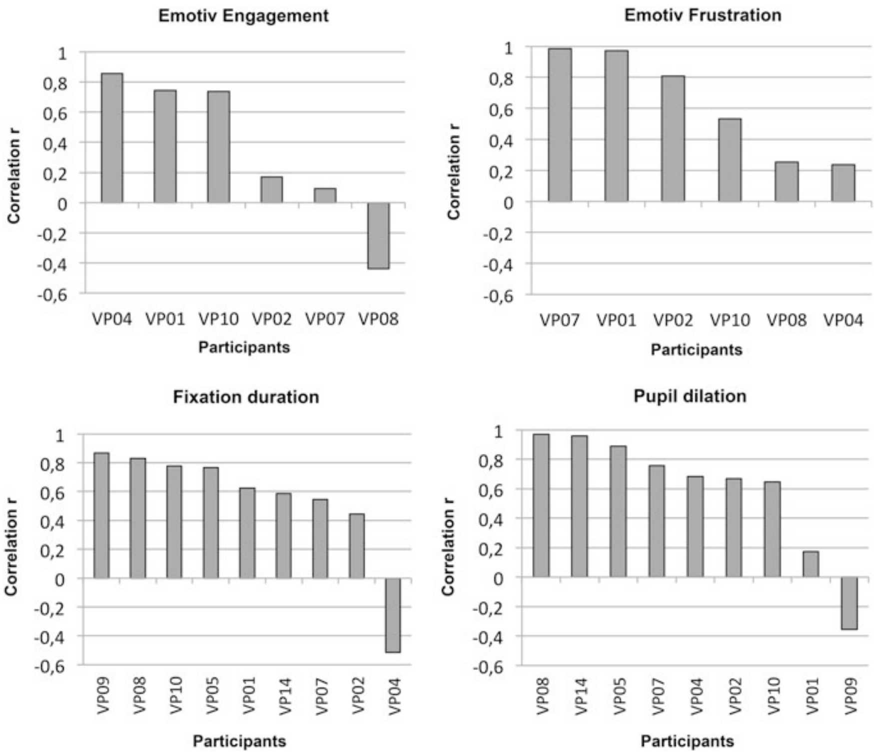


Fig. 4 Test-retest correlations of each participant for emotiv engagement, emotiv frustration, fixation duration, and pupil dilation. Order of participants is based on individual correlation value (from high to low)

3.3 Correlation with Performance in Test and Retest

Table 2 provides correlations between each of the four physiological measures and performance at the group level. Although correlations are rather low correlations obtained for Emotiv Frustration, fixation duration and pupil dilation are similar in range between test and retest, hence supporting the temporal stability of the measures' diagnostic value. Emotiv Engagement, however, significantly correlated with performance in the first experiment but shows only poor correlation with performance in the retest.

The direction of all correlations (with the exception of the correlation obtained for Emotiv Engagement in the retest) is in congruence with our assumptions: Emotiv Engagement (first experiment), Emotiv Frustration, and pupil dilation correlate negatively with performance, indicating that performance (expectedly) decreases with higher levels of workload and frustration. For fixation duration, a slightly positive correlation with performance was found. This positive relationship was also expected as our first experiment revealed that fixation duration was lower in conditions of high task load and high task load conditions were in turn associated with lower performance.

Correlations between physiological measures and performance were also calculated for each participant. These individual correlations are depicted as bar charts in Fig. 5. To allow for easy comparison, test and retest correlations are plotted side-by-side for each participant.

Outcomes indicate that individual correlations with performance not only differ between individuals but they also differ within individuals from test to retest: Metrics that were highly correlated with performance in the first test (e.g. fixation duration for VP02) showed rather weak correlations in the retest. Only for Emotiv Frustration some individual correlations seem to be rather stable between test and retest (VP01, VP04, VP08 and VP07—even though the correlations of VP07 are not in the expected direction).

Table 2 Correlations of EEG- and eye tracking metrics with the performance score in test and retest (group level)

	Emotiv engagement	Emotiv frustration	Fixation duration	Pupil dilation
First experiment	−0.49*	−0.27	0.17	−0.28
Retest	0.18	−0.35	0.23	−0.35*

*significant ($p < 0.05$)

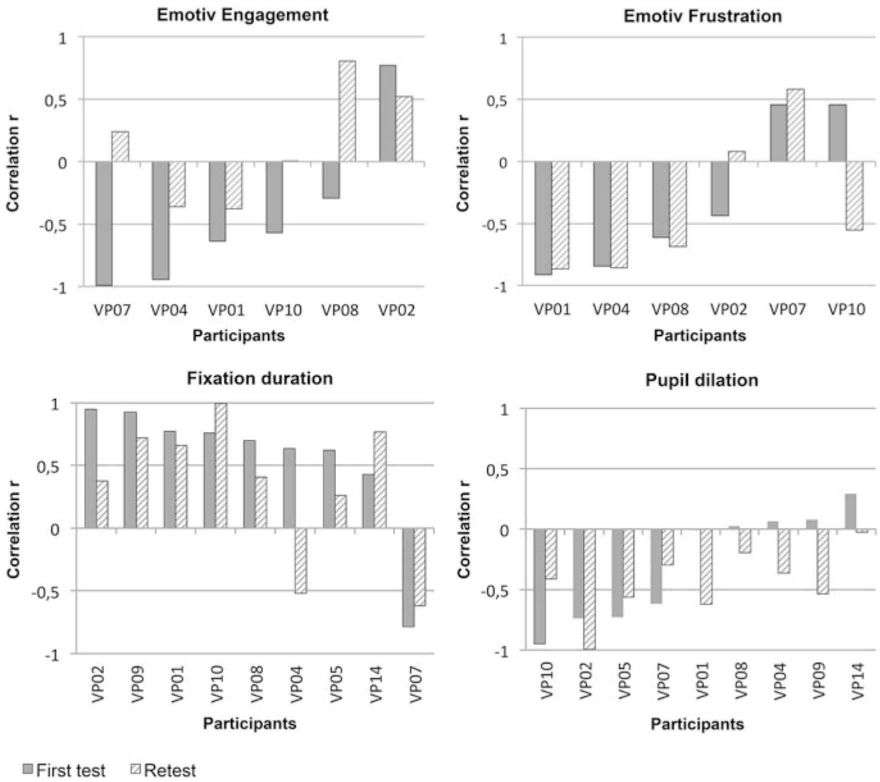


Fig. 5 Individual correlations of EEG and eye tracking metrics with performance in test and retest. Order of participants is based on individual correlation value from the first experiment: From high to low or reverse (not expected direction)

4 Discussion

Our analysis of the test-retest stability of Emotiv Engagement, Emotiv Frustration, fixation duration, and pupil dilation revealed moderate stability at a group level, in line with findings of Faulstich et al. [5]. At an individual level, many test-retest correlations were also moderate to high but differed substantially between individuals ranging from high positive correlations ($r > 0.8$) to even negative correlations.

Tomarken [6] also points out a substantial variability of test-retest stability; however, not between individuals but across physiological measures. He argues that physiological outcomes are often influenced by a large number of confounding “nuisance variables” that can vary across sessions (e.g. attention level of participants, inconsistency in placements of electrodes). These factors may also differ across participants, leading to variations in test-retest stability at an individual level.

Another potential source of variability in the individual test-retest correlations is our applied experimental task paradigm in which participants are free to control air tracks as deemed appropriate. Accordingly, the same scenario (with the same experimental conditions) may unfold and be handled very differently from one test to another. Thus, identical experimental conditions may involve different amounts of demanding or critical situations.

Finally, analysis of correlations between the physiological measures and performance indicates that outcomes differ between test and retest not only between but also within individuals. This implies that the diagnostic value of these physiological measures to indicate user state and performance decrements is not only user-specific but also time-specific (which may in part be attributed to confounding variables that are hardly controllable in applied contexts).

5 Implications for Future Research

Our findings demonstrate that group level findings, while commonly reported in the context of user state assessment, are not indicative of a metric's utility for real-time user state diagnosis, as this metric may only work for some users in a given context. Additional analyses at an individual level provide valuable insights and have important implications for applications that require user-specific state diagnosis, such as adaptive systems design.

In this article, we showed that the efficacy of physiological measures to diagnose user state and performance decrements can strongly vary between individuals. Furthermore, test-retest results at the individual level show a temporal instability of metrics, indicating that, even for individually customized user state indicators, the diagnostic value may vary from one session to the next. Hence, it may be necessary to periodically readjust diagnostic components (e.g. in adaptive human-machine systems based on physiological user state diagnoses) to maintain sensitive and valid diagnoses.

For future research in this area, we feel a need to point out the importance of the temporal stability of physiological measures. This aspect is hardly considered in evaluations but is critical for deciding whether or not to include a metric in a diagnosis framework. We would therefore like to encourage researchers to include and publish temporal stability analyses in their physiological research.

References

1. Dirican, C., Göktürk, M.: Psychophysiological measures of human cognitive states applied in human computer interaction. Procedia Comput. Sci. **3**, 1361–1367 (2011)
2. Kruse, A.A., Schmorrow, D.D.: Session Overview: foundations of augmented cognition. In Proceedings of the 11th International Conference on Human-Computer Interaction (Augmented Cognition International) (2005)

3. Berka, C., Levendowski, D.J., Davis, G., Whitmoyer, M., Hale, K., Fuchs, S.: Objective measures of situational awareness using neurophysiology technology. In: Schmorrow, D.D., Stanney, K.M., Reeves, L.M. (eds.) Foundations of Augmented Cognition, pp. 145–154. Strategic Analysis Inc, Arlington, VA (2006)
4. Schwarz, J., Fuchs, S., Flemisch, F.: Towards a more holistic view on user state assessment in adaptive human-computer interaction. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC 2014), pp. 1228–1234 (2014)
5. Faulstich, M.E., Williamson, D.A., McKenzie, S.J., Duchmann, E.G., Hutchinson, K.M., Blouin, D.C.: Temporal stability of psychophysiological responding: a comparative analysis of mental and physical stressors. Int. J. Neurosci. **30**(1–2), 65–72 (1986)
6. Tomarken, A.J.: A psychometric perspective on psychophysiological measures. Psychol. Assess. **7**(3), 387–395 (1995)
7. Schwarz, J., Fuchs, S.: Efficacy of metrics from a low-cost EEG sensor for multi-dimensional user state assessment. In: Wienrich, C., Zander, T. Gramann, K. (eds.) Trends in Neuroergonomics: Proceedings of the 11. Berlin Workshop Human-Machine Systems, pp. 20–21. Universitätsverlag der TU Berlin, Berlin (2015)
8. Neerincx, M.A.: Cognitive task load design: model, methods and examples. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Design, pp. 283–305. Lawrence Erlbaum Associates, Mahwah, NJ (2003)
9. Reuderink, B., Nijholt, A., Poel, M.: Affective pacman: a frustrating game for brain-computer interface experiments. In: Third International Conference on Intelligent Technologies for Interactive Entertainment, INTETAIN 2009, pp. 221—227. Amsterdam, The Netherlands (2009)
10. Diener, H., Oertel, K.: Experimental approach to affective interaction in games. In: Pan, Z., et al. (eds.) Edutainment 2006, LNCS, vol. 3942, pp. 507–518. Springer, Heidelberg (2006)
11. Pomplun, M., Reingold, E.M., Shen, J.: Investigating the visual span in comparative search: the effects of task difficulty and divided attention. Cognition **81**(2), B57–B67 (2001)
12. King, L.A.: Visual navigation patterns and cognitive load. In: Foundations of augmented cognition. Neuroergonomics and operational neuroscience, pp. 254–259. Springer, Berlin (2009)
13. Meghanathan, R. N., van Leeuwen, C., Nikolaev, A. R.: Fixation duration surpasses pupil size as a measure of memory load in free viewing. Front. Hum. Neurosci. **8**, 1063 http://doi.org/10.3389/fnhum.2014.01063 (2014)
14. Duchowski, A.T.: A breadth-first survey of eye tracking applications. Behav. Res. Methods Instrum. Comput. **34**, 455–471 (2002)
15. De Rivecourt, M., Kuperus, M.N., Post, W.J., Mulder, L.J.M.: Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. Ergonomics **51**(9), 1295–1319 (2008)
16. Richer, F., Beatty, J.: Pupillary dilations in movement preparation and execution. Psychophysiology **22**(2), 204–207 (1985)
17. Van Orden, K.F., Limbert, W., Makeig, S., Jung, T.P.: Eye activity correlates of workload during a visuospatial memory task. Hum. Factors **43**(1), 111–121 (2001)
18. Iqbal, S.T., Zheng, X.S., Bailey, B.P.: Task-evoked pupillary response to mental workload in human computer interaction. In: Ext. Abstracts CHI 2004, pp. 1477–1480. ACM Press (2004)
19. Jiang, X., Zheng, B., Bednarik, R., Atkins, M.S.: Pupil responses to continuous aiming movements. Int. J. Hum Comput. Stud. **83**, 1–11 (2015)
20. Kramer, A.F.: Physiological metrics of mental workload: a review of recent progress. In: Damos, D. (ed) Multiple Task Performance, pp. 279—328. Taylor and Francis, London (1991)
21. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. NorthHolland, Amsterdam (1988)
22. Bradley, M.M.: Lang. P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry **25**(1), 49–59 (1994)

Transcranial Direct Current Stimulation (tDCS) Versus Caffeine to Sustain Wakefulness at Night When Dosing at Start-of-Shift

Lindsey McIntire, R. Andy McKinley, Justin Nelson and Chuck Goodyear

Abstract Shift work is necessary in many industries such as healthcare, trucking, defense, and aviation. It is well documented that during the night shift, workers experience the lowest levels of performance and alertness (Czeisler et al. in Science 210:1264–1267, 1980; Akerstedt and Gillberg in Sleep 4:159–169, 1981 [1, 2]). Research has shown caffeine can enhance alertness and performance during overnight work (Muehlbach and Walsh in Sleep 18(1):22–29, 1995 [4]). However, benefits of caffeine decline over time (Miller et al. in Fatigue and its Effect on Performance in Military Environments (Report No. 0704–0188), 2007 [5]). McIntire et al. (Brain Stimul. 7(4):499–507, 2014 [6]) found a promising alternative for use during sleep deprivation called transcranial direct current stimulation (tDCS). tDCS sustained performance throughout the sleep deprivation vigil and for a longer amount of time when compared to caffeine. Three groups of participants received either active tDCS and placebo gum at the start of their shift (1800), caffeine gum with sham tDCS, or sham tDCS with placebo gum. Participants completed 13 sessions of tasks and questionnaires while remaining awake for 36 h. Our results show tDCS could be a possible fatigue countermeasure.

Keywords Fatigue · Sleep-deprivation · Transcranial direct current stimulation · Caffeine · Shift work

L. McIntire (✉) · J. Nelson · C. Goodyear
Infoscitex, Inc., 4027 Colonel Glenn Highway, Dayton, OH 45431, USA
e-mail: lindsey.mcintire.ctr@us.af.mil

J. Nelson
e-mail: justin.nelson.15.ctr@us.af.mil

C. Goodyear
e-mail: charles.goodyear.1.ctr@us.af.mil

R. Andy McKinley
711th HPW/RHCP, Wright-Patterson Air Force Base, Dayton, OH 45433, USA
e-mail: richard.mckinley.2@us.af.mil

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_14

1 Introduction

Shift work is prevalent and necessary in many industries such as healthcare, trucking, manufacturing, defense, and aviation. It is well documented that during the night shift, workers experience the lowest levels of performance and alertness [1, 2]. This is not only due to circadian factors but also because most night shift workers spend a longer time awake before their shift ends (20–22 h) compared to other shifts [3]. A common countermeasure used to enhance performance and alertness is caffeine. Research has shown that caffeine can enhance alertness and performance during overnight work compared to no intervention [4] as well as provide performance benefits through times of circadian variation [7, 8]. However, the benefits of caffeine decline over time and from chronic use [5]. Also, while caffeine may increase the ability to stay awake, it does not always aid in making good decisions [9]. Therefore, it is necessary to investigate other forms of fatigue countermeasures for shift workers.

A form of non-invasive brain stimulation called transcranial direct current stimulation (tDCS) might be a viable fatigue countermeasure. Although it was originally used to treat clinical disorders such as stroke, major depressive disorder, Parkinson's disease, etc., rapid expansion of research over the past decade has been showing that tDCS can also be effective in enhancing healthy human cognitive performance (see [10, 11] for reviews). In fact, our own research has indicated that tDCS could be a promising alternative to caffeine for use during sleep deprivation because it was found to mitigate the deleterious performance effects of sustained wakefulness for up to 6 h post-stimulation whereas caffeine only provided 2 h of benefit. Participants who received tDCS also reported feeling more energetic, less fatigued, and had a more overall positive mood rating than people who received caffeine. Consequently, we believe tDCS could be beneficial to the performance and mood of night shift workers.

2 Methods

2.1 Equipment

tDCS Stimulator. The MagStim DC stimulator (Magstim Company Limited; Whitland, UK) was used to provide the tDCS stimulation. This battery-powered device was controlled with a microprocessor to ensure constant current at up to 5000 μA. The device automatically shuts off if the impedance becomes greater than 50 kΩ to prevent electric shocks or burns.

tDCS Electrodes. The electrodes included an array of 5 electroencephalographic (EEG) electrodes arranged in a circular pattern. Each electrode had an inner diameter of 1.6 cm yielding a contact area of 2.01 cm^2 for each electrode. At 2 mA of supplied current, there was an average current density of $0.199 \frac{mA}{cm^2}$.

StayAlert® Gum (MarketRight, Inc., Plano, IL). Gum was the delivery mechanism used to administer 200 mg of caffeine to participants in the caffeine group. The placebo gum was also StayAlert® gum which looked and tasted the same as the caffeinated gum.

2.2 Subjects

Thirty active-duty military participants from Wright-Patterson Air Force Base completed this study. There were 22 male and 8 female participants with an average age of 26.6 ± 5. Participants were compensated for their time but where disqualified if they met any of the exclusion criteria described in McKinley et al. [12]. Thirty-two participants enrolled in the study but two were dismissed because they met one or more of the study exclusion criteria. The 30 remaining participants were randomly assigned into one of three groups with 10 individuals in each.

2.3 Performance Tasks

Mackworth Clock Test (Vigilance Task). The vigilance task was developed according to the description of the task used by Kilpeläinen et al. [13]. The task was an adopted version of the Mackworth clock test with parameters adopted from Teikari [14] and run on a standard desktop computer. The participant was presented a visual display with 16 hole-like black circles arranged in a clock-like figure against a black background. Each circle changed from black to red for 0.525 s in turn, with each cycle lasting 3 s. The red light moved in a clockwise pattern by one step, which was considered the normal stimulus appearance. The light moving twice the usual distance (i.e., skipping a circle) was considered a critical signal and the participant was required to respond to this signal by pressing the spacebar as fast as possible on the keyboard with his preferred index finger. The response was defined as a correct hit when it occured less than 8 s after the target signal and a false alarm if the reaction occurred outside this time range (+0.1–8.0 s). Undetected targets were defined as misses.

Psychomotor Vigilance Task (PVT). The PVT -192 (Ambulatory Monitoring, Inc.; Ardsley, NY) was a 8″ × 4/5″ × 2/4″ handheld, battery-operated computerized test presentation that recorded visual reaction times. The visual stimulus was presented on a small LCD that presented a number counted up by milliseconds. The stimulus was presented for up to 1 min, allowing the participant to respond by using a button press with the thumb. The interstimulus interval varied randomly from 2 to 12 s.

2.4 Subjective Questionnaire

Profile of Mood States (POMS). The Profile of Mood States (POMS) was a 65-item questionnaire that measures mood using 6 categories: tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment. Participants rated their feelings about each item (example items: Tense, Vigor, Fatigue) on a scale of 0–4, with 0 being "not at all" and 4 being "extremely." The overall scores for each of the six categories were totaled, resulting in six factor scores. Each of the factor scores, except for the vigor-activity score, was added together; next the vigor-activity score was subtracted from this total to produce a composite mood disturbance score. We also analyzed each factor score independently.

2.5 Procedures

Participants experienced 36 h of continuous wakefulness for this experiment. Participants were randomly assigned to one of three experimental groups (n = 10 for each group). Group 1 received anodal tDCS at 2 mA for 30 min at 1800 with placebo gum at 1700, group 2 received sham tDCS at 1800 with 200 mg of caffeinated gum at 1700, group 3 received placebo gum at 1700 and sham brain stimulation at 1800. After consenting to participate in the study, participants filled out the medical screening questionnaire. Two days prior to their scheduled experimental trial, participants were given an activity wrist monitor and instructed that their daily schedules should include a minimum of 7 h of sleep per night between the hours of 2300 and 0600. Also during this time participants received training on all performance tasks utilized in the study.

On the day of their experimental trial, participants were required to awaken at 0600 and perform their daily activities as normal. They were instructed to not consume any caffeine or central nervous system (CNS)-altering medications/substances on the experimental test day. Each participant arrived at the test facility at 1630 h. Upon arrival, they had their WAM activity data analyzed to ensure that proper sleep cycles were maintained. Starting at 1700 h, participants were given either active or placebo chewing gum depending on the condition they were in. At 1800, depending on the condition the participant was in, the brain stimulation began and the participants completed one session of the sustained attention task (30 min), one session of the PVT task (10 min), and filled out the POMS questionnaire. Participants were then provided a short break of approximately 45 min where they could talk, watch TV, walk, read, or play video games. The second session began at 2000 h and was exactly the same as the first session. These procedures were repeated every two hours. The final testing session took place at 1800 h on the second day (36 h continuous wakefulness).

2.6 Analysis

For each subject, the change from session 1 (1800) was determined for all dependent variables. Pairwise comparisons of the 3 groups at each session were made using 2-tailed 2-sample t-tests and effect size from Cohen's d. No error adjustment was made for the number of tests. At each session, the minimum p-value for the 3 paired comparisons was used in plots and identified as either $0 < p$ 0.01 or $0.01 < p \leq 0.05$. For each analysis, group (Early tDCS, Early Caffeine, Control) was a between factor and session a within factor for sessions 2–13. Proc mixed in SAS 9.2 was used to perform the analyses.

3 Results

An analysis of variance (ANOVA) revealed a significant main effect of "session" for the Mackworth Clock test metric of accuracy ($F(11, 297) = 23.9$, $P < 0.001$) (see Fig. 1 and Table 1). The post hoc t-tests showed that the tDCS group performed significantly better than the sham group at 2200 and again at 1400 the next day, which was Session 11. There was no statistical difference found between the caffeine and control groups as well as the caffeine and tDCS groups. However, at 2200 the caffeine and control groups were close to statistical significance with a Cohen's d of 0.90. A value above 0.8 indicates a possible large effect that might be detected with a larger n-size. At 2400, the tDCS group was not significantly better than the control group but it was also close with a Cohen's d of 0.95, we believe that an effect might have been detected with a larger n-size.

An ANOVA also uncovered a main effect of "session" for mean of log reaction times ($F(11, 291) = 18.23$, $P < 0.001$) on the Mackworth task (Fig. 1 and Table 1). At 2200, the post hoc t-tests showed that tDCS and caffeine groups performed significantly better than the control group. The session at 2000 was not significantly different between the tDCS and control group but the comparison revealed a Cohen's d of 0.92 which indicates it may have reached significance with a larger n-

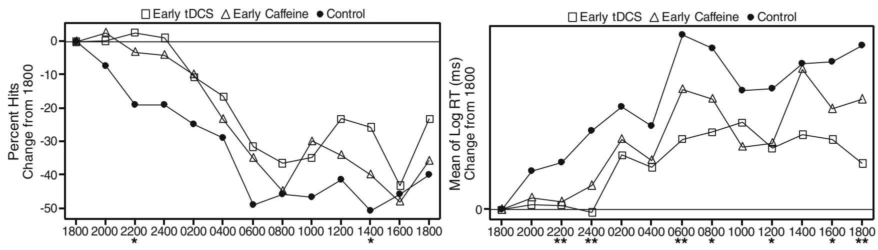


Fig. 1 Mackworth clock test percent hits as a change from 1800 (*left*) and mean of the logged reaction time as a change from 1800 (*right*)

Table 1 *t*-tests of significant main effects for the Mackworth and PVT performance tasks

Task	Dependent variable	Session	Time	Group 1	Group 2	Mean 1	SEM 1	Mean 2	SEM 2	DF	t	P
Mackworth	Percent hits	3	2200	Control	Caffeine	−18.939	6.559	−3.167	4.984	18	−1.91	0.0716
Mackworth	Percent Hits	3	2200	Control	tDCS	−18.939	6.559	2.727	5.011	18	−2.62	0.0172
Mackworth	Percent Hits	4	241100	Control	tDCS	−18.939	8.194	1.061	5.518	18	−2.02	0.058
Mackworth	Percent Hits	11	1400	Control	tDCS	−50.606	5.868	−25.606	8.773	18	−2.37	0.0292
Mackworth	Mean Log RT	2	2000	Control	tDCS	0.136	0.041	0.018	0.046	18	1.94	0.0679
Mackworth	Mean Log RT	3	2200	Control	Caffeine	0.168	0.029	0.028	0.016	18	4.25	0.0005
Mackworth	Mean Log RT	3	2200	Control	tDCS	0.168	0.029	0.016	0.038	18	3.17	0.0053
Mackworth	Mean Log RT	4	2400	Control	Caffeine	0.281	0.042	0.085	0.052	18	2.92	0.0091
Mackworth	Mean Log RT	4	2400	Control	tDCS	0.281	0.042	−0.01	0.056	18	4.15	0.0006
Mackworth	Mean Log RT	7	600	Control	Caffeine	0.621	0.069	0.426	0.06	18	2.15	0.0455
Mackworth	Mean Log RT	7	600	Control	tDCS	0.621	0.069	0.253	0.092	17	3.24	0.0048
Mackworth	Mean Log RT	8	800	Control	tDCS	0.57	0.092	0.275	0.081	18	2.42	0.0263
Mackworth	Mean Log RT	10	1200	Control	tDCS	0.431	0.063	0.217	0.074	18	2.19	0.0419
Mackworth	Mean Log RT	12	1600	Control	tDCS	0.523	0.096	0.252	0.064	17	2.29	0.0348
Mackworth	Mean Log RT	13	1800	Control	tDCS	0.582	0.096	0.167	0.063	17	3.55	0.0025
Mackworth	Mean Log RT	13	1800	Caffeine	tDCS	0.39	0.097	0.167	0.063	17	1.9	0.0751
PVT	Number of Lapses	6	400	Control	tDCS	14.1	3.501	5	1.445	12	2.4	0.0334
PVT	Number of Lapses	7	600	Control	tDCS	23.6	4.241	10.8	3.289	18	2.39	0.0283
PVT	Number of Lapses	7	600	Caffeine	tDCS	27.7	6.391	10.8	3.289	18	2.35	0.0303
PVT	Number of Lapses	10	1200	Control	tDCS	24.7	4.382	12.1	3.328	18	2.29	0.0343
PVT	Number of Lapses	10	1200	Caffeine	tDCS	31.4	5.207	12.1	3.328	18	3.12	0.0059
PVT	Number of Lapses	11	1400	Caffeine	tDCS	29	6.482	12.1	2.718	12.1	2.4	0.0332
PVT	Number of Lapses	13	1800	Control	tDCS	11.5	1.99	6.5	1.176	18	2.16	0.0443

(continued)

Table 1 (continued)

Task	Dependent variable	Session	Time	Group 1	Group 2	Mean 1	SEM 1	Mean 2	SEM 2	DF	t	P
PVT	Mean Reciprocal RT	7	600	Control	tDCS	−0.00145	0.000194	−0.00102	0.000109	18	−1.95	0.0674
PVT	Mean Reciprocal RT	7	600	Caffeine	tDCS	−0.00164	0.000217	−0.00102	0.000109	18	−2.54	0.0205
PVT	Mean Reciprocal RT	8	800	Caffeine	tDCS	−0.00147	0.000149	−0.00106	0.000095	18	−2.35	0.0302
PVT	Mean Reciprocal RT	10	1200	Caffeine	tDCS	−0.00168	0.000186	−0.00102	0.000127	18	−2.96	0.0083
PVT	Mean Reciprocal RT	11	1400	Caffeine	tDCS	−0.00164	0.000166	−0.00107	0.0001	18	−2.97	0.0082
PVT	Mean Reciprocal RT	12	1600	Caffeine	tDCS	−0.00146	0.000197	−0.00102	0.00009	12.6	−2.04	0.0632

size. There was a significant effect at 2400 (or Session 4). Both the tDCS and caffeine groups had lower mean reaction times compared to the control group. Again at 0600 or session 7 the tDCS and caffeine groups had faster (better) reaction times than the control group. Although, at the next session (Session 8, 0800) reaction time in the tDCS group was found to be significantly faster than control, there was no statistical difference between the caffeine and control group or caffeine and tDCS. At 1200 and 1600, the tDCS group performed statistically different from the control group. The means revealed the tDCS group had faster (better) reaction times at both of these sessions. There was no difference between the other groups at these two time points. During the last session at 1800 the tDCS group performed statistically better than the control group. While there was no significant difference between the tDCS group and caffeine group, the Cohen's *d* was 0.92 indicating a possible effect if there had been a larger *n*-size. Taken together, the evidence suggests that the effects of tDCS delivered at 1800 the previous day may last 24 h post-stimulation, whereas, caffeine's effects disappear after 12 h. In fact, the tDCS group had 10–44 % better reaction times than the control group.

A separate ANOVA revealed a significant main effect of "group" for the metric of lapses in the PVT task ($F(2, 27) = 4.02$, $P = 0.030$). The post hoc *t*-tests revealed that tDCS had fewer lapses than the control group at sessions 6, 7, 10, and 13. The tDCS group also had fewer lapses than the caffeine group at sessions 7, 10, and 11. There were no statistical differences found between the caffeine and control group for any sessions. In fact, the tDCS group performed between 10 and 22 % better than the other groups between 0400 and 1800 because they were making fewer lapses (Fig. 2 and Table 1).

The ANOVA also showed a significant main effect of "session" for the metric of mean reciprocal reaction time in the PVT task ($F(11, 297) = 61.86$, $P < 0.001$). Post hoc *t*-tests showed that the tDCS group had faster (better) reaction times than the caffeine group at sessions 7, 8, 10, and 11. There was no difference between the control and caffeine group at any time point. At session 7 (0600), the test between the tDCS group and control group was approaching significance with a Cohen's *d* of 0.92. It is likely this would have achieved significance with a larger *n*-size.

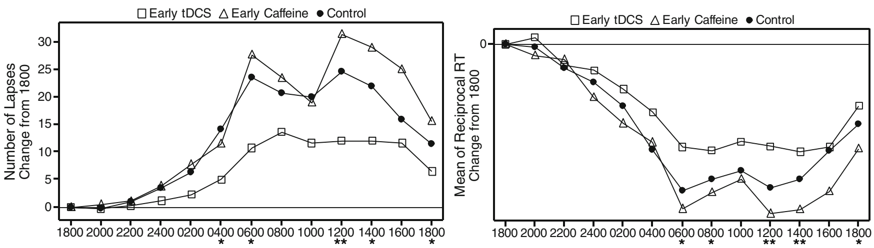


Fig. 2 Number of lapses on the PVT as a change from 1800 (*left*) and mean of the reciprocal reaction time on the PVT as a change from 1800 (*right*)

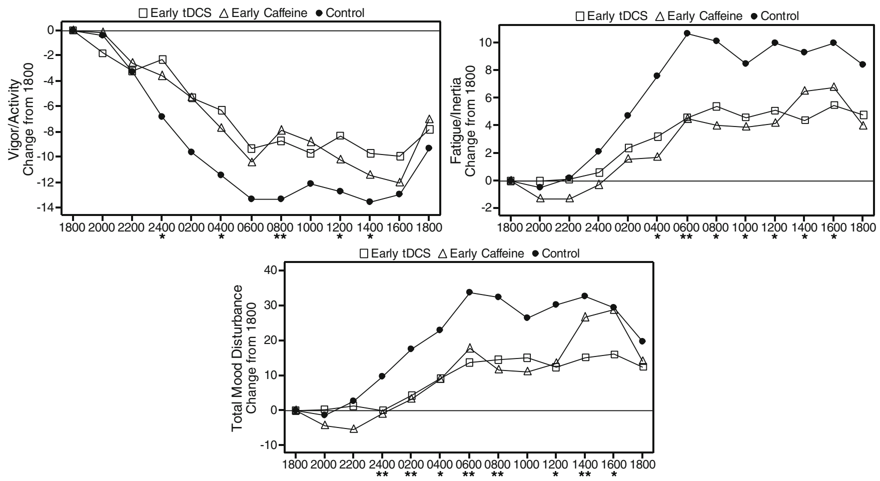


Fig. 3 POMS questionnaire as a change from 1800 on the factor score of vigor/activity (*top left*), fatigue/inertia (*top right*), and the total mood disturbance score (*bottom*)

Also at session 12 (1600) the Cohen's *d* was 0.96 in the test between the tDCS and caffeine group so again a larger *n*-size may have revealed a significant difference here (Fig. 2 and Table 1).

Using an ANOVA, we found a significant main effect of "session" on the Vigor/Activity factor score in the POMS questionnaire ($F(11, 297) = 36.14$, $P < 0.001$). Post hoc *t*-tests revealed that the tDCS group reported feeling more vigorous than the control group at sessions 4, 6, 8, 10, and 11. The caffeine group reported significantly more vigor than the control group only at session 8. There was no difference found between the tDCS and caffeine groups (Fig. 3 and Table 2). We also found a significant main effect for the "group and session" ($F(22, 297) = 1.76$, $P = 0.02$) interaction on the POMS questionnaire for the Fatigue/Inertia factor score. The post hoc *t*-tests indicated that the tDCS group reported less fatigue than the control group at sessions 7, 10, 11, and 12. While session 6 and 8 did not reveal a significant effect, the Cohen's *d* was 0.89 and 0.86 respectively. Therefore, a larger *n*-size might show a significant difference. The caffeine group reported significantly less fatigue than the control group at sessions 6, 7, 8, 9, and 10. There was no statistical difference between the caffeine and tDCS group (Fig. 3 and Table 2). Finally, on the POMS questionnaire we found a significant main effect of "session" on the calculated total mood disturbance ($F(11, 297) = 18.75$, $P < 0.001$). The post hoc *t*-tests showed that tDCS had less of a disturbance to overall mood than the control group at sessions 4–12, which was up to 22 h post-stimulation. Caffeine only showed a benefit over control at session 8. There was no statistical difference between the caffeine and tDCS groups (Fig. 3 and Table 2). On average, the tDCS group felt 19 % better than the control group.

Table 2 *t*-tests of significant main effects for the POMS questionnnaire

Task	Dependent Variable	Session	Time	Group 1	Group 2	Mean 1	SEM 1	Mean 2	SEM 2	DF	t	P
POMS	Vigor/Activity	4	2400	Control	tDCS	−6.8	1.02	−2.3	1.557	18	−2.42	0.0264
POMS	Vigor/Activity	6	400	Control	tDCS	-11.4	1.087	−6.3	1.578	18	−2.66	0.0159
POMS	Vigor/Activity	8	800	Control	Caffeine	−13.3	0.731	−7.9	2.34	10.7	−2.2	0.0504
POMS	Vigor/Activity	8	800	Control	tDCS	−13.3	0.731	−8.7	1.274	18	−3.13	0.0058
POMS	Vigor/Activity	10	1200	Control	tDCS	−12.7	0.932	−8.3	1.513	18	−2.48	0.0234
POMS	Vigor/Activity	11	1400	Control	tDCS	−13.5	0.749	−9.7	1.469	18	−2.31	0.0333
POMS	Fatigue/Inertia	6	400	Control	Caffeine	7.6	2.177	1.7	0.989	12.6	2.47	0.0288
POMS	Fatigue/Inertia	6	400	Control	tDCS	7.6	2.177	3.2	0.854	11.7	1.88	0.0849
POMS	Fatigue/Inertia	7	600	Control	Caffeine	10.7	1.613	4.5	1.31	18	2.98	0.008
POMS	Fatigue/Inertia	7	600	Control	tDCS	10.7	1.613	4.6	1.046	18	3.17	0.0053
POMS	Fatigue/Inertia	8	800	Control	Caffeine	10.1	2.089	4	1.599	18	2.32	0.0324
POMS	Fatigue/Inertia	8	800	Control	tDCS	10.1	2.089	5.4	1.485	18	1.83	0.0833
POMS	Fatigue/Inertia	9	1000	Control	Caffeine	8.5	1.759	3.9	1.286	18	2.11	0.049
POMS	Fatigue/Inertia	10	1200	Control	Caffeine	10	1.88	4.2	1.744	18	2.26	0.0363
POMS	Fatigue/Inertia	10	1200	Control	tDCS	10	1.88	5.1	1.059	18	2.27	0.0356
POMS	Fatigue/Inertia	11	1400	Control	tDCS	9.3	1.627	4.4	1.204	18	2.42	0.0262
POMS	Fatigue/Inertia	12	1600	Control	tDCS	10	1.238	5.5	1.157	18	2.66	0.0161
POMS	Total Mood Disturbance	4	2400	Control	tDCS	9.7	1.82	0.1	2.147	18	3.41	0.0031
POMS	Total Mood Disturbance	5	200	Control	tDCS	17.6	2.544	4.3	3.091	18	3.32	0.0038
POMS	Total Mood Disturbance	6	400	Control	tDCS	22.8	4.514	9.2	3.169	18	2.47	0.0239
POMS	Total Mood Disturbance	7	600	Control	tDCS	33.6	5.296	13.7	3.496	18	3.14	0.0057
POMS	Total Mood Disturbance	8	800	Control	Caffeine	32.3	4.527	11.6	8.354	18	2.18	0.0429

(continued)

Table 2 (continued)

Task	Dependent Variable	Session	Time	Group 1	Group 2	Mean 1	SEM 1	Mean 2	SEM 2	DF	t	P
POMS	Total Mood Disturbance	8	800	Control	tDCS	32.3	4.527	14.5	3.67	18	3.05	0.0068
POMS	Total Mood Disturbance	9	1000	Control	tDCS	26.4	4.551	15.1	3.038	18	2.06	0.0536
POMS	Total Mood Disturbance	10	1200	Control	tDCS	30.2	6.048	12.4	3.078	18	2.62	0.0172
POMS	Total Mood Disturbance	11	1400	Control	tDCS	32.5	4.324	15.2	3.844	18	2.99	0.0078
POMS	Total Mood Disturbance	12	1600	Control	tDCS	29.5	4.641	16.1	3.013	18	2.42	0.0262

4 Discussion

This effort investigated the effects of anodal tDCS applied to the dorsolateral prefrontal cortex on attention, reaction time, and mood during an extended-duty night shift that equated to 36 h of total sleep loss. These effects were compared to the most widely available and commonly used intervention to combat fatigue: caffeine. The results suggest that tDCS had a beneficial and long-lasting effect on vigilance, reaction time, and many aspects of mood that are negatively influenced by fatigue stress.

One of the notable effects was a significant improvement in target detection accuracy in the Mackworth Vigilance Test. Vigilance refers to the ability to detect randomly, and infrequently occurring targets among non-target cutter for sustained periods of time [15–17]. Over the period of watch, the ability to detect targets declines in a linear fashion. This phenomenon is known as the "vigilance decrement" [18] and two competing theories have dominated the literature to describe the underlying cause. "Arousal theory" contends that the decline in vigilance is caused by a decline in cortical arousal or energy over time [19, 20]. Colloquially, the operator loses interest in the task or becomes bored and attention suffers. Over the years, arousal theory has been unable to consistently establish a causal link between arousal and the vigilance decrement [20]. Further, we have found that arousal theory was not able to fully explain the vigilance decrement in our own studies [6, 21, 22]. Given that the PVT is a simple reaction time task that serves as a test of overall arousal levels coupled with the fact that tDCS led significantly faster reaction times, it is tempting to conclude that the effects on vigilance accuracy (i.e. preservation of the vigilance decrement) were due to increased arousal. However, caffeine has been shown to be a robust modulator of arousal (e.g. 23), but there was no statistical difference between the caffeine and control groups in the current experiment. Hence caffeine did not have a meaningful effect on the vigilance decrement, leaving doubt as to the effect of arousal on the vigilance decrement. Increased arousal is commonly accompanied by increased operator response bias [23]. However, we did not find a significant difference in false alarms among the groups (i.e. no change in response bias), indicating that arousal changes alone are not sufficient to describe the changes in the vigilance decrement found in the tDCS group. In fact, several studies have dismissed the argument that tDCS effects area caused by modulating arousal alone [24–27].

An alternative concept to explain the vigilance decrement is known as "resource theory." Resource theory posits that cognitive performance depends on utilization of some "cognitive resources" that are limited in supply. Once they have been exhausted, cognitive processing and function begin to suffer [28–30]. While the experiment described herein did not examine cognitive resources (such as metabolite resources) directly, tDCS has been shown to increase oxygen saturation and blood flow in the stimulated regions [21, 31]. Coupled with the lack any difference in false alarms (i.e. no change in operator response bias), we argue that the effects of tDCS on the vigilance decrement are caused in part by changes in available metabolic resources. However, more research is needed to confirm this hypothesis. It should be noted that these results are supported by Falcone et al. [32] who showed that tDCS only modulates perceptual sensitivity and not operator response bias. This was confirmed with our past findings showing an improvement in detection rate without a corresponding increase in errors [6, 21]. In our previous study, the effect of tDCS was found to last 6 h post-stimulation, although this 6-h segment occurred at the end of the data collection period [6]. Hence, it was not possible to determine if this effect remained past the 6-h point. This issue has now been addressed with the new data from the current experiment. The data confirmed that was no change in target detection accuracy for the tDCS group for 6 h post-tDCS and the effects dissipated thereafter.

While changes in arousal may not be sufficient to explain the effects of tDCS on the vigilance decrement, there is plenty of evidence that suggests that our tDCS paradigm also modulates arousal. Because the PVT is a sensitive test of arousal and arousal is highly influenced by fatigue, the PVT is common tool to assess the impairment induced by sleep deprivation [33, 34]. Examining the results from the PVT, the data revealed that the tDCS group had significantly faster response times and fewer lapses when compared to the control or caffeine groups. In fact, these effects lasted approximately 24 h post tDCS (i.e. for the duration of the data collection period). Lapses are defined as a failure to respond to a stimulus or a reaction time exceeding 500 ms. In fact, the tDCS group averaged 15 % fewer lapses than the other two groups throughout the entire 13 sessions. It is theorized that lapses on the PVT is a result of perceptual, processing, or executive failures in the central nervous system and sleep deprivation increases the tendency for these systems to fail [35]. The data supports this theory as caffeine had no effect on lapses and many studies have shown that tDCS over the LDLPFC influences various executive functions such as: multi-tasking [36], working memory [37], implicit learning [38], and vigilance [6, 21] in both healthy and clinical populations. It is possible that tDCS sustained processing power in executive functioning during this extended wakefulness period causing fewer lapses. In addition, the tDCS and caffeine groups demonstrated significantly faster response times in the Mackworth Clock test when compared to the control group. However, the duration of effect was at least twice as long for the tDCS group (24 h). It is possible that effects on vigilance target detection are caused by changes in available cognitive resources, while the effects on reaction time are related to changes in arousal. In either case, the evidence suggests that exogenous stimulation modulates both.

The performance results were coupled with differences in the subjective questionnaire scores. Fatigue stress has a profound effect on mood and a meta-analysis revealed that mood was more affected by sleep deprivation than cognitive and motor performance [39]. Further, caffeine can lead to increases in depression and confusion [40] and any positive effect that it has on mood during sleep deprivation is short-lived [5, 6]. The analyses of the POMS questionnaire revealed that the tDCS group had a higher significantly higher score on the Vigor/Activity factor than the control group and this effect lasted up to 20 h post-stimulation. For the Fatigue/Inertia factor score both tDCS and caffeine groups indicated that they felt significantly less fatigued than the control group. However, this effect only lasted for 6 h longer for the tDCS group. Finally, decline in the Total Mood Disturbance score was significantly attenuated in the tDCS group when compared to the control group. This effect also remained for 22 h post-stimulation. Taken together, the evidence suggests that tDCS has a positive effect on mood that lasted longer than the caffeinated intervention. Specifically, the tDCS group reported feeling more vigorous, less fatigued, and their composite mood scores were superior throughout the entire sleep deprivation period. Therefore, tDCS not only led to superior performance in vigilance and reaction time tests, but participants also *felt* better than their counterparts in the other treatment groups.

These results reported herein suggest that tDCS may be an effective intervention to reduce the deleterious effects of cognitive fatigue. Applications are wide and varied from military and civilian applications such as air traffic control, image analysis, and nuclear reactor monitoring, to clinical/therapeutic applications to possible reduce chronic fatigue in patients suffering from a number of neurological ailments. However, more research is needed to confirm these findings outside our own laboratory.

References

1. Czeisler, C.A., Weitzman, E.D., Moore-Ede, M.C., Zimmerman, J.C., Knauer, R.S.: Human sleep: its duration and organization depend on its circadian phase. Science **210**, 1264–1267 (1980)
2. Akerstedt, T., Gillberg, M.: The circadian variation of experimentally displaced sleep. Sleep **4**, 159–169 (1981)
3. Fröber, J.E., Karlsson, C.G., Levi, L., Lidberg, L.: Psychobiological circadian rhythms during a 72-Hour Vigil. Försvarsmedicin. **II**, 192–201 (1975)
4. Muehlbach, M.J., Walsh, J.K.: The effects of caffeine on simulated night-shift work and subsequent daytime sleep. Sleep **18**(1), 22–29 (1995)
5. Miller, N.L., Matsangas, P., Shattuck, L.G.: Fatigue and its Effect on Performance in Military Environments (Report No. 0704–0188). Naval Postgraduate School Operations Research Department, Monterey (2007)
6. McIntire, L.K., McKinley, R.A., Nelson, J.M., Goodyear, C.: A comparison of the effects of transcranial direct current stimulation and caffeine on vigilance and cognitive performance during extended wakefulness. Brain Stimul. **7**(4), 499–507 (2014)
7. McLellan, T.M., Kamimori, G.H., Voss, D.M., Bell, D.G., Cole, K.G., Johnson, D.: Caffeine maintains vigilance and improves run times during night operations for special forces. Aviat. Space Environ. Med. **76**, 647–654 (2005)
8. Tharion, W.J., Shukitt-Hale, B., Lieberman, H.R.: Caffeine effects on marksmanship during high-stress military training with 72 hour sleep deprivation. Aviat. Space Environ. Med. **74**(4), 309–314 (2003)
9. Kilgore, W.D.S., Balkin, T.J., Wesensten, N.J.: Impaired decision making following 49 h of sleep deprivation. J. Sleep Res. **15**(1), 7–13 (2006)
10. Nitsche, M.A., Boggio, P.S., Fregni, F., Pascual-Leone, A.: Treatment of depression with transcranil direct current stimulation (tDCS): a review. Exp. Neurol. **219**, 14–19 (2009)
11. McKinley, R.A., Nelson, J., Bridges, N., Walters, C.: Modulating the brain at work using noninvasive transcranial stimulation. NeuroImage **59**(1), 129–137 (2012)

12. McKinley, R.A., Weisend, M.P., McIntire, L.K., Bridges, N., Walters, C.: Acceleration of image analyst training with transcranial direct current stimulation. Behav. Neurosci. **127**(6), 936–946 (2013)
13. Kilpeläinen, A.A., Huttunen, K.H., Lohi, J.J., Lyytinen, H.: Effect of caffeine on vigilance and cognitive performance during extended wakefulnes. Int J Aviat Psychol. **20**(2), 144–159 (2010)
14. Teikari, V.: Vigilanssi-ilmiön Mittaamisesta Ja Selitysmahdollisuuksista [On Measurement And Explanation Of Vigilance] (Jyväskylä Studies In Education, Psychology And Social Research No. 35). University of Jyväskylä, Jyväskylä (1977)
15. Davies, D.R., Parasuraman, R.: The Psychology of Vigilance. Academic Press, London (1982)
16. Parasuraman, R. (ed.): The Attentive Brain. The MIT Press, Cambridge (1998)
17. Warm, J.S., Parasuraman, R., Matthews, G.: Vigilance requires hard mental work and is stressful. Hum. Factors **50**, 433–441 (2008)
18. Mackworth, J.F.: Performance decrement in vigilance, threshold, and high-speed perceptual motor tasks. Can. J. Psychol. **18**, 209–223 (1964)
19. Frankmann, J.P., Adams, J.A.: Theories of vigilance. Psychol. Bull. **59**, 257–272 (1962)
20. Parasuraman, R.: Sustained attention: a multifactorial approach. In: Posner, M.I., Marin, O.S. (eds.) Attention and Performance XI, pp. 493–511. Erlbaum Associates, New Jersey (1985)
21. Nelson, J.T., McKinley, R.A., Golob, E.J., Warm, J.S., Parasuramen, R.: Modulating the prefrontal cortex during sustained attention with transcranial direct current stimulation. NeuroImage **85**, 909–917 (2014)
22. Nelson, J.M., McKinley, R.A., McIntire, L.K., Goodyear, C., Walters, C.: Augmenting visual search performance with transcranial direct current stimulation. Mil. Psychol. **27**(6), 335–347 (2015)
23. Posner, M.I.: Chronometric Explorations of Mind. Erlbaum, Hillsdale, NJ (1978)
24. Boros, K., Poreisz, C., Munchau, A., Paulus, W., Nitsche, M.A.: Premotor transcranial direct current stimulation (tDCS) affects primary motor excitability in humans. Eur. J. Neurosci. **27** (5), 1292–1300 (2008)
25. Antal, A., Nitsche, M.A., Kruse, W., Kincses, T.Z., Hoffmann, K., Paulus, W.: Facilitation of visuo-motor learning by transcranial direct current stimulation of the motor and extrastriate visual areas in humans. Eur. J. Neurosci. **19**, 2888–2892 (2004)
26. Flöel, A., Rösser, N., Michka, O., Knecht, S., Breitenstein, C.: Noninvasive brain stimulation improves language learning. J. Cogn. Neurosci. **20**(8), 1415–1422 (2008)
27. Nitsche, M.A., Roth, A., Kuo, M.F., Fischer, A.K., Liebetanz, D., Lang, N., Tergau, F., Paulus, W.: Timing-dependent modulation of associative plasticity by general network excitability in the human motor cortex. J. Neurosci. **27**(14), 3807–3812 (2007)
28. Kahneman, D.: Attention and Effort. Prentice Hall, Englewood Cliffs, NJ (1973)
29. Matthews, G., Davies, D.R., Westerman, S.J., Stammers, R.B.: Vigilance And Sustained Attention. Human Performance: Cognition, Stress, and Individual Differences. Psychology Press, Philadelphia, PA, pp. 107–124 (2000)
30. Wickens, C.D.: Processing resources in attention. In: Parasuraman, R., Davies, D.R. (eds.) Varieties of Attention, pp. 63–102. Academic Press, Orlando (1984)
31. Merzagora, A.C., Foffani, G., Panyavin, L., Mordillo-Mateos, L., Aguilar, J., Onaral, B., Oliviero, A.: Prefrontal hemodynamic changes produced by anodal direct current stimulation. NeuroImage **49**(3), 2304–2310 (2010)
32. Falcone, B., Coffman, B.A., Clark, V.P. Parasuraman, R.: Transcranial direct current stimulation augments perceptual sensitivity and 24-hour retention in a complex threat detection task. PLoS One (2012)
33. Dorrian, J., Rogers, N.L., Dinges, D.F.: Psychomotor Vigilance Performance: Neurocognitve Assay Sensitive to Sleep Loss. Marcel Dekker, New York (2004)
34. Lamond, N., Dawson, D., Roach, G.D.: Fatigue assessment in the field: validation of a hand-held electronic psychomotor vigilance task. Aviat. Space Environ. Med. **76**, 486–489 (2005)

35. Lim, J., Dinges, D.F.: Sleep deprivation and vigilant attention. Ann. NY Acad. Sci. **1129**, 305–322 (2008)
36. Nelson, J., McKinley, R.A., Phillips, C., McIntire, L., Goodyear, C., Camden, A., Monforton, L.: the effects of noninvasive brains stimulation on multitasking throughput capacity. IEEE Trans. Hum. Mach. Syst. (Submitted)
37. Fregni, F., Boggio, P.S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., Marcolin, M.A., Rigonatti, S.P., Silva, M.T.A., Paulus, W., Pascual-Leone, A.: Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. Exp. Brain Res. **166**, 23–30 (2005)
38. Kincses, T.Z., Antal, A., Nitsche, M.A., Bártfai, O., Paulus, W.: Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the human. Neuropsychologia **42**(1), 113–117 (2004)
39. Pilcher, J.J., Huffcutt, A.J.: Effects of sleep deprivation on performance: a meta-analysis. J. Sleep Res. Sleep Med. **19**(4), 318–326 (1996)
40. Lieberman, H.R., Tharion, W.J., Shukitt-Hale, B., Speckman, K.L., Tulley, R: Effects of caffeine, sleep loss, and stress on cognitive performance and mood during U.S. navy SEAL training. Psychophar. **164**, 250–261 (2002)

The Effects of Transcranial Direct Current Stimulation (tDCS) on Training During a Complex Procedural Task

R. Andy McKinley, Lindsey McIntire, Jeremy Nelson, Justin Nelson and Charles Goodyear

Abstract There is a growing body of literature suggesting transcranial direct current stimulation (tDCS) is effective in accelerating certain types of memory including working memory, language learning, and object recognition. Recent studies have provided new evidence that non-declarative memories such as motor skill acquisition may be enhanced through direct stimulation of motor cortex. Additionally, Galea and Celnik (J Neurophysiol 102:294–301, [10]) showed that inhibition of the prefrontal cortex following motor training led to enhanced procedural memory consolidation. This effort examined the effects of excitatory transcranial direct current stimulation (tDCS) over the primary motor cortex on memory acquisition and inhibitory tDCS over dorsolateral prefrontal cortex in memory consolidation. Thirty-six Air Force members volunteered to participate. They were divided into four groups: anodal tDCS over motor cortex, cathodal stimulation over dorsolateral prefrontal cortex (DLPFC), both anodal tDCS over motor cortex and cathodal tDCS over DLPFC, or sham tDCS over motor cortex and DLFPC. All participants received their stimulation condition while training on a procedural task that required them to identify incoming aircraft as friend or foe. Twenty-four hours after the training session, participants returned to the lab for retention testing. When comparing the day 1 (training) scores to the day 2 (test) score, the results showed the cathodal tDCS group performed 2× better than sham and all real tDCS groups exhibited scores

R.A. McKinley (✉)
711th HPW/RHCP, Wright-Patterson Air Force Base, Dayton, OH 45433, USA
e-mail: richard.mckinley.2@us.af.mil

L. McIntire · J. Nelson · C. Goodyear
Infoscitex, Inc, 4027 Colonel Glenn Highway, Dayton, OH 45431, USA
e-mail: lindsey.mcintire.ctr@us.af.mil

J. Nelson
e-mail: justin.nelson.15.ctr@us.af.mil

C. Goodyear
e-mail: charles.goodyear.1.ctr@us.af.mil

J. Nelson
AFMOA/HCE, Port San Antonio, San Antonio, TX, USA
e-mail: jeremy.nelson.15.ctr@us.af.mil

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_15

significantly higher than sham. The evidence suggests that inhibition of the prefrontal cortex leads to the greatest improvement in performance. We hypothesize that this is a result of a shift in dominance of the declarative memory system to the non-declarative procedural system, which improves consolidation of the procedural memories gained during training.

Keywords Noninvasive brain stimulation · Learning · Transcranial direct current stimulation · Non-declarative · Procedural

1 Introduction

In the past few years many studies have shown that tDCS can enhance performance on a variety of tasks. Transcranial DC stimulation studies have shown enhancements in language learning [1] and language performance [2], working memory [3, 4], declarative memory [5], and spatial tactile acuity [6]. Other studies have also shown that tDCS can increase task performance in many clinical populations (e.g., [7, 8]).

Typically, the formation of memory is described as having three stages: acquisition, consolidation, and recall. Further, multiple memory systems have been proposed, each dealing with particular aspects of memory (for a review, see [9]). For purposes of the proposed study, we are using the declarative/non-declarative dichotomy. Declarative memory refers to remembering history, personal events, and other facts of personal experience such as whether the door was re-locked after leaving for work. Non-declarative memory encompasses memories that are much less accessible by conscious recollection and are difficult to verbalize. Learned habits and skills, classical conditioning, and priming (memory driven by external stimuli not a decision) are all examples of non-declarative memory. The neuroanatomical bases for the various long-term memory systems differ, and even within non-declarative memory, there are several sub-types. Procedural (skills and habits) are thought to derive mainly from the striatum, whereas priming and perceptual learning or knowledge about facts and events (declarative memory) are thought to derive from the neocortex and medial temporal lobe, respectively.

Recent work has suggested that tDCS may be well-suited for the enhancement of motor skill acquisition [10, 11]. These results are seen when stimulation increasing the plasticity in the primary motor cortex is used. Plasticity refers to the brains ability to change over time by developing new neural connections (synapses) and reorganizing brain networks during learning. Further, research suggests that due to the competitive nature of the brain's multiple memory systems, disruption of one system may lead to an enhancement in the consolidation of another. This has been repeatedly established in animal studies (e.g., [12]) and was recently suggested as possible noninvasively in humans [10]. In their study, Galea and colleagues demonstrated that disruption of the dorsolateral prefrontal cortex using inhibitory transcranial magnetic stimulation (TMS) directly after motor skill acquisition led to enhanced procedural memory consolidation and subsequent recall (there was no motor cortex stimulation in this study).

Previous research investigating the efficacy of tDCS to improve skill learning has only been performed with basic motor functions (i.e., patterns of finger tapping). As a result, it is not known whether similar stimulation procedures will result in accelerated retention/consolidation when applied to a more complex task. In addition, it is not known whether depression of competing cortical circuits with cathodal tDCS will result in enhanced procedural memory in humans. The present study was aimed at establishing the efficacy of brain stimulation for acquisition of procedural skills. The results will also inform future work aimed at determining optimal training schedules (using brain stimulation) for maintaining high-performance skills.

2 Methods

2.1 Design

This study utilized a 1 × 4 between-subjects experimental design with the factor being "stimulation type." The five levels include: (1) Active (anodal) tDCS over left hemisphere primary motor cortex (hand region) during skill acquisition; (2) Active (cathodal) tDCS over right hemisphere dorsolateral prefrontal cortex skill consolidation; (3) Active (anodal) tDCS over left hemisphere primary motor cortex during skill acquisition AND active (cathodal) tDCS over right dorsolateral prefrontal cortex during skill consolidation; and (4) Sham tDCS during skill acquisition and skill consolidation (mimicking condition 3 but with no actual stimulation). The active tDCS electrode array was placed on the subjects' scalp while the inactive array was placed on their contralateral upper arm.

2.2 Subjects

A total of 32 active duty, right-handed Air Force military members Wright Patterson Air Force Base (26 male, 6 female) volunteered to participate in the study. Participants were dismissed if they had evidence of any of the following: a neurological diagnosis, a psychological diagnosis, psychological hospitalization, hospitalization for surgery/illness within 6 months of participation, taking of psychotropic medications, a shot in the left arm within one week of participation (i.e. flu, allergy, pain), non-removable metal or tattoos around the head, uncorrectable vision impairments, pregnant or could become pregnant, smoking, treatment for drug/alcohol within 6 months of participation, head injury within 30 days of participation, or history of any of the following: learning difficulty, frequent headaches, attention deficit, severe head injury, seizures, fainting, migraine headaches, high blood pressure, diabetes, or heart disease. Participants were randomly assigned to one of 4 groups (n = 8 each). Group 1 received anodal tDCS over their left motor cortex (M1) during training, group 2 received cathodal stimulation over the dorsolateral prefrontal cortex (DLPFC) during immediately following training, group 3 received both anodal tDCS over M1 during training and cathodal tDCS over DLPFC immediately following training, and group 4 received sham tDCS at both sites. All participants received a souvenir coin for their participation. Those that participated in an off-duty status also received $20/h as compensation for their time.

2.3 Apparatus

Transcranial Direct Current Stimulation. Transcranial direct current stimulation was delivered with a MagStim (Whiteland, Wales, UK) DC stimulator. In place of the standard wet sponge electrodes that are delivered with the unit, custom silver-silver chloride electroencephalography (EEG) electrodes were utilized (see Fig. 1). Both the anode and cathode each consisted of five of these EEG electrodes arranged in a circular pattern, as described in [13]. Electrodes were secured using medical bandages, and connectivity was ensured using highly conductive gel (SignaGel, Parker Laboratories, Fairfield, NJ, USA).

Warship Commander Task. Participants were required to perform the "Warship Commander" cognitive task during testing in their assigned treatment condition. Only the "Airspace Monitoring" portion of the task was utilized under setting "T—High Yellow—12 aircraft per wave." This task was selected due to the fact that it includes a set of very specific procedures the subjects must learn and follow. By learning the procedures, their scores should improve and the number of mistakes (incorrect button presses) should decline. Prior to their training session, subjects were briefed on the task instructions, proper procedures, and associated controls. They were permitted to ask questions for clarification after the briefing and

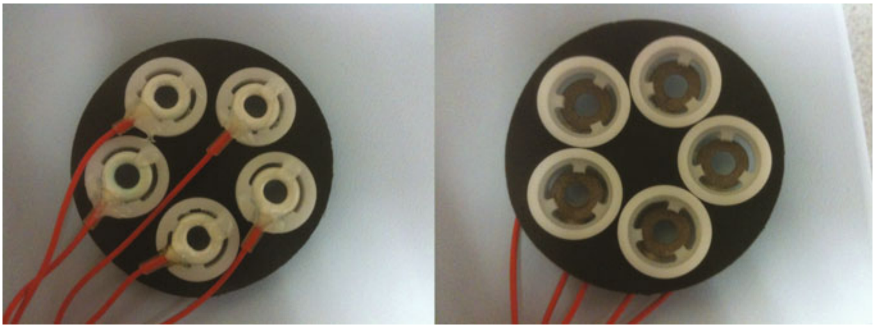


Fig. 1 EEG-based tDCS electrodes

during their training. The Airspace monitoring portion of the Warship Commander task includes a plan view (top-down) image of a sea environment. A two-dimensional display of an aquatic surface, as viewed from directly above was presented to the subjects on a standard 17 in. computer monitor with a viewing distance of 32 in. Near the bottom edge of the screen is a grey naval vessel that represents the subject's "ownship." Figure 2 provides a screenshot of the task.

Upon task execution, waves of white aircraft begin entering the scene. By using the point-and-click method with the supplied mouse, targets can be selected and then subsequently identified using the "Identification, Friend or Foe" (IFF) virtual button located at the bottom of the display. Once the IFF button is depressed, the selected aircraft changes from white to one of three colors: red, yellow, or blue. Red denotes an enemy aircraft, yellow indicates the aircraft type is unknown, and blue designates friendly aircraft. Yellow (unknown) aircraft required further analysis to determine the aircraft type. The participant was required to select the aircraft of interest with the mouse and then depress the highlighted number to the left of the communications window that indicates the aircraft's ID. Two seconds after selecting the aircraft, a statement appeared in the communications window that indicated whether the aircraft was assumed to be hostile or friendly. Hostile, yellow aircraft require a warning before they can be engaged. This is completed by selecting the aircraft to be warned and then depressing the "Warn" button located to the right of the IFF button within the interface. The aircraft are given 3 s to turn away from the participant's "ownship" at the bottom of the screen. If still advancing toward the bottom of the screen after the 3 s have elapsed, the participant was

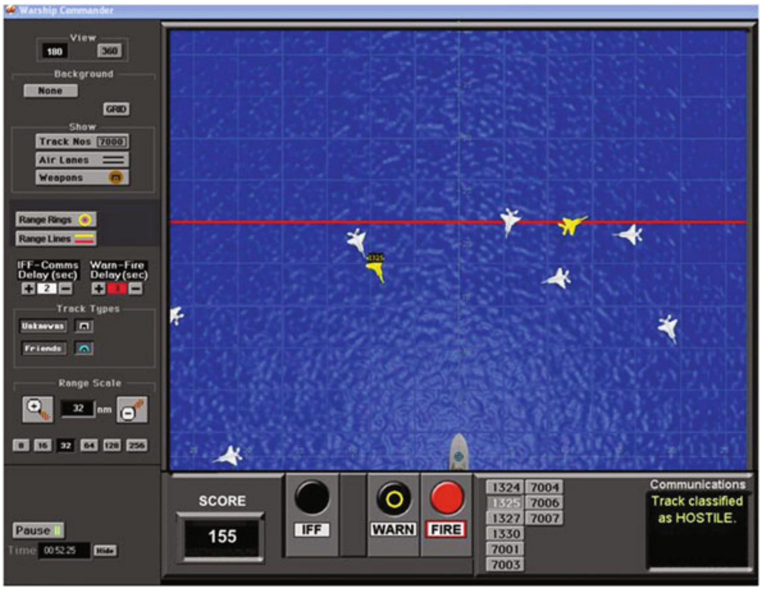


Fig. 2 Warship commander airspace monitoring task

required to fire upon the target by depressing the “Fire” button. Likewise, any red (enemy) aircraft were to be fired upon immediately. Blue or yellow friendly aircraft are to be ignored. A score is presented to the left of the IFF button near the bottom of the screen. Points were awarded automatically for target identifications, proper warnings, and destruction of enemy/hostile targets. Incorrect button presses, scores, and event times were recorded by the data logger within the software package.

2.4 Procedures

All experimental procedures used in the protocol were reviewed by the Wright-Patterson Air Force Base Institutional Review Board (IRB). Subjects first provided written informed consent to participate before being registered into the study or filling out Initial Screening Questionnaires. These questionnaires gathered information regarding their fitness for participation and background. If the participant failed to meet the inclusion criteria, they were excluded from participation and no identifiable data was retained. Those that did meet the inclusion criteria completed a neuropsychological test including a shortened Wechsler Abbreviated Scale of Intelligence and a NEO-Five Factor Inventory (NEO-FFI) personality test administered by a trained research associate. The behavioral battery consists of measures of individual differences comprising intellectual, personality, working memory, speed of processing and problem solving ability. This testing lasted approximately one hour. We then applied the tDCS electrodes to the scalp and upper arm according to their stimulation condition. Afterwards, subjects were given a short break followed by a short movie clip describing the Warship Commander task instructions and tDCS procedures. Next, subjects were permitted to ask questions for clarification of the task rules and procedures.

Afterwards, they received stimulation according to their group assignment as they perform the Warship Commander task for 20 min. Immediately after the 20 min session, those assigned in groups that receive stimulation during skill consolidation received either active or sham prefrontal stimulation for 10 min. Each subject then returned to the lab twenty-four hours after the beginning of the initial practice trial and was tested again on the Warship Commander task. In this session, they performed the task without the tDCS treatment.

3 Results

The primary variable of interest was the score achieved in each session. Due to individual differences in skill, initial scores on the task varied greatly. In an effort to normalize the data to examine the relative improvement in performance rather than the raw scores, we calculated performance as a percentage change from baseline. Here, we determined the mean score in the first session of the training day and set

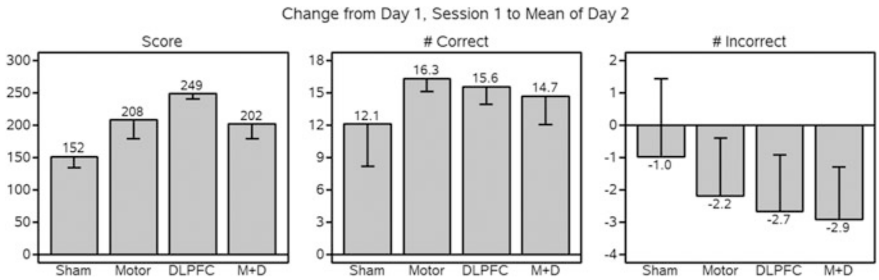


Fig. 3 Warship commander test score (*left panel*), number of correct button presses (*center panel*), and number of incorrect button presses (*right panel*) as a change from day 1 session 1 to the mean across sessions on day 2

this as the "baseline" for each individual participant. The means reported within this document are the means of the percent changes in performance from this baseline.

A one-way analysis of variance (ANOVA) showed a significant effect of stimulation type ($F(3, 36) = 6.34$, $P = 0.0015$) on score. Post hoc two-tailed two-sample t-tests revealed that the score for the DLPFC tDCS group was significantly greater than the score for the sham tDCS group ($p = 0.0002$) and the combined M1 and DLPFC tDCS group ($p = 0.0053$). The comparison between the sham tDCS and M1 tDCS group was approaching significance ($p = 0.0524$). The test had a Cohen's d of 1.02 indicating a large effects size. Hence, this test may have achieved significance with a slightly larger sample size. None of the other pairwise comparisons reached statistical significance.

The ANOVA failed to find a main effect of group on the number of correct button presses ($F(3, 36) = 2.30$, $P = 0.0935$) or the number of incorrect button presses ($F(3, 36) = 0.53$, $P = 0.6622$). Additionally, there were no significant group*session interactions for any of the dependent variables examined. Figure 3 shows the means from each group across the three dependent variables.

4 Discussion

This experiment evaluated the efficacy of two different tDCS paradigms in isolation and in combination to influence non-declarative memory processes. To assess the change in performance, the total score, number of correct and number of incorrect responses were compared across the treatment groups. Examining the data from the test day (i.e. day 2), the analyses did not reveal any significant differences among the 4 sessions. As a result, they were collapsed into one average score/number. The change in performance from the first training session to the average performance during the test on the second day was then examined. The results showed there was a significant effect of group on score, but not accuracy (i.e. number of correct and number of incorrect responses). Pair-wise comparisons of the means were conducted using paired t-tests. Only the group receiving cathodal tDCS over dorsolateral prefrontal cortex (DLPFC) during the memory consolidation period had a significantly higher score than sham. Examining the group means, the participants receiving tDCS over DLPFC improved their score by more than 60 % over those that received sham. Hence, the data suggest that inhibition of the DLPFC using cathodal tDCS led to a large increase in procedural learning. While it did not achieve statistical significance, the analysis suggested that anodal tDCS over primary motor cortex applied during training trended toward significant improvements in score as well. However, the effect size was smaller according to the Cohen's d's for these comparisons (i.e. 2.26 for sham vs. DLPFC and 1.02 for sham vs. M1). It is likely that this comparison would have achieved significance with a larger sample size.

Following the work of Galea and Celnik [10], it was hypothesized that disruption or inhibition of brain areas or networks involved in cognitive processes that compete with procedural memory would paradoxically facilitate procedural memory encoding and consolidation. Specifically, they found disruption of DLPFC to be effective in improving procedural memory when applied immediately following training (i.e. during the memory consolidation period). To test this idea using tDCS, the traditional model of cathodal tDCS generating cortical inhibition effects and anodal tDCS producing cortical excitation effects was leveraged using previous findings in the literature [14–16]. Assuming inhibition effects of cathodal tDCS in the motor cortex previously reported [14] extend to the prefrontal cortex, the evidence reported herein supports original evidence from Galea and Celnik [10]. It is hypothesized that inhibiting prefrontal cortex results in a shift in dominance of the declarative memory system to the non-declarative procedural system. This theoretically frees up cognitive resources for the non-declarative memory processes such as acquisition and consolidation of the procedural memories gained during training. However, it should be noted that when applied to prefrontal cortex, cathodal tDCS does not always produce the opposite cognitive effects of anodal tDCS. A meta-analysis conducted by Jacobson et al. [17] showed that when applied over non-motor areas, cathodal tDCS rarely results in inhibition effects as measured by cognitive task performance. The authors concluded that because higher order cognitive functions use highly complex brain networks, other brain regions may help compensate for the reduction in activity in the area affected by tDCS thus reducing any negative consequence of the cathodal tDCS. Coupling the behavioral results with neurophysiology would help elucidate such effects and should be considered for future studies.

It also appears that there may be a weaker effect of anodal tDCS on procedural learning when applied during training over the motor cortex. The literature suggests that anodal tDCS over primary motor cortex modulates cortical excitability in a way that facilitates acquisition and consolidation of motor memory (e.g. [18–20]). While the effect of anodal tDCS during training did not lead to a statistically significant improvement in score when compared to sham, the Cohen's d revealed a large effect size and the t-test was approaching significance. The effect may have been weakened by the fact that the cognitive test was not purely motor learning. Instead, the procedural task selected (Warship Commander) was complex and involved both a motor memory component in the form of the sequence of button presses *and* an executive decision making component (i.e. is the aircraft a friend or foe?). The effect was further weakened by a relatively small sample size. Nevertheless, a larger study should be conducted to confirm whether such an effect exists. It has long been theorized that tDCS modulates plasticity, mimicking the natural memory formation processes such as long-term potentiation (LTP). Recent evidence confirms that 30 min of tDCS yields strong and repeatable LTP effects in the hippocampus [21]. In our study, it is possible that tDCS induced LTP effects that aided in the memory formation process (i.e. acquisition) and the consolidation process immediately post-tDCS.

Many experiments have tested the effects of tDCS on learning and cognitive behavior using simplistic, psychological, laboratory-based tests. One goal of this study was to determine if effects found in these simplistic tasks (e.g. finger tapping exercises) extend to more complex tasks that are more relevant to real-world activities. The behavioral results support that tDCS influences procedural learning in a more complex task environment with an increased number of rules and situations. Recently, the efficacy of tDCS to enhance procedural learning in a complex simulated aircraft landing task was investigated [22]. While the authors did not find any significant effects of tDCS on landing performance metrics, there was an effect of tDCS in reduction of variance that was coupled with changes in midline frontal theta-band power. They concluded that tDCS improved learned skill consistency and this may have applications to accelerating training to acquire complex skills. Taken together, these results are an important initial step towards advancing our understanding of the practicality of tDCS. However, additional research needs to be conducted to examine the effects in tasks of increasing real-world fidelity.

5 Conclusion

Similar to the effects discovered by Galea and Celnik [10], the data suggests that cathodal (inhibitory) tDCS applied over DLPFC immediately following training has beneficial effects on procedural learning. It is theorized the declarative and no-declarative memory processes compete for cognitive resources within the brain. We conclude that cathodal tDCS likely inhibited activity in DLPFC which shifted cognitive resources from declarative memory processes to non-declarative memory processes, thereby facilitating procedural memory consolidation. Additional research that combines neurophysiology with the behavioral outcomes should be conducted to confirm that DLPFC is inhibited with this tDCS paradigm.

References

1. Floel, A., Rosser, N., Michka, O., Knecht, S., Breitenstein, C.: Noninvasive brain stimulation improves language learning. J. Cogn. Neurosci. **20**(8), 1415–1422 (2008)
2. Sparing, R., Dafotakis, M., Meister, I.G., Thirugnanasambandam, N., Fink, G.R.: Enhancing language performance with non-invasive brain stimulation—a transcranial direct current stimulation study in healthy humans. Neuropsychologia **46**(1), 261–268 (2008)
3. Ohn, S.H., Park, C.I., Yoo, W.K., Ko, M.H., Choi, K.P., Kim, G.M., Lee, Y.T., Kim, Y.H.: Time-dependent effect of transcranial direct current stimulation on the enhancement of working memory. Neuroreport **19**(1), 43–47 (2008)
4. Fregni, F., Boggio, P.S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., Marcolin, M.A., Rigonatti, S.P., Silva, M.T., Paulus, W., Pascual-Leone, A.: Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. Exp. Brain Res. **166**(1), 23–30 (2005)
5. Marshall, L., Molle, M., Hallschmid, M., Born, J.: Transcranial direct current stimulation during sleep improves declarative memory. J. Neurosci. **24**(44), 9985–9992. (2004) (Erratum in: J Neurosci. 25(2), 1 p following 531 2005)
6. Ragert, P., Vandermeeren, Y., Camus, M., Cohen, L.G.: Improvement of spatial tactile acuity by transcranial direct current stimulation. Clin. Neurophysiol. **119**(4), 805–811 (2008)
7. Monti, A., Cogiamanian, F., Marceglia, S., Ferrucci, R., Mameli, F., Mrakic-Sposta, S., Vergari, M., Zago, S., Priori, A.: Improved naming after transcranial direct current stimulation in aphasia. J. Neurol. Neurosurg. Psychiatry **79**(4), 451–453 (2008)
8. Boggio, P.S., Ferrucci, R., Rigonatti, S.P., Covre, P., Nitsche, M., Pascual-Leone, A., Fregni, F.: Effects of transcranial direct current stimulation on the enhancement of working memory. NeuroReport **19**(1), 43–47 (2008)
9. Squire, L.R.: Memory systems of the brain: a brief history and current perspective. Neurobiol. Learn. Mem. **82**(3), 171–177 (2004)
10. Galea, J.M., Celnik, P.: Brain polarization enhances the formation and retention of motor memories. J. Neurophysiol. **102**, 294–301 (2009)
11. Hunter, T., Sacco, P., Nitsche, M.A., Turner, D.L.: Modulation of internal model formation during force field induced motor learning by anodal transcranial direct current stimulation (tDCS) of primary motor cortex. J. Physiol. **12**, 2949–2961 (2009)
12. Eichenbaum, H., Cohen, N.J.: From Conditioning to Conscious Recollection: Memory Systems of the Brain. Oxford University Press, Oxford (2004)
13. McKinley, R.A., Weisend, M.P., McIntire, L.K., Bridges, N., Walters, C.: Acceleration of image analyst training with transcranial direct current stimulation. Behav. Neurosci. **127**(6), 936–946 (2013)
14. Nitsche, M.A., Paulus, W.: Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J. Physiol. **527**(3), 633–639 (2000)
15. Nitsche, M.A., Paulus, W.: Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology **57**, 899–901 (2001)
16. Purpura, D.P., McMurtry, J.G.: Intracellular activities and evoked potential changes during polarization of motor cortex. J. Neurophysiol. **28**, 166–185 (1965)
17. Jacobson, L., Koslowsky, M., Lavidor, M.: tDCS polarity effects in motor and cognitive domains: a meta-analytical review. Exp. Brain Res. **216**(1), 1–10 (2012)
18. Nitsche, M.A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., Tergau, F.: Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. J. Cogn. Neurosci. **15**(4), 619–626 (2003)
19. Antal, A., Nitsche, M.A., Kincses, T.Z., Kruse, W., Hoffmann, K.P., Paulus, W.: Facilitation of visuo-motor learning by transcranial direct current stimulation of the motor and extrastriate visual areas in humans. Eur. J. Neurosci. **19**(10), 2888–2892 (2004)

20. Stagg, C.J., Jayaram, G., Pastor, D., Kincses, Z.T., Matthews, P.M., Johansen-Berg, H.: Polarity and timing-dependent effects of transcranial direct current stimulation in explicit motor learning. Neuropsychologia **49**(5), 800–804 (2011)
21. Rohan, J., Carhuantanta, K.A., McInturf, S.M., Miklasevich, M.K., Jankord, R.: Modulating hippocampal plasticity with in vivo brain stimulation. J. Neurosci. **35**(37), 12824–12832 (2015)
22. Choe, J., Coffman, B.A., Bergstedt, D.T., Ziegler, M.D., Phillips, M.E.: Transcranial direct current stimulation modulates neuronal activity and learning in pilot training. Front. Hum. Neurosci. **9**, 10–34 (2016)

Eliciting Sustained Mental Effort Using the Toulouse N-Back Task: Prefrontal Cortex and Pupillary Responses

Mickaël Causse, Vsevolod Peysakhovich and Kevin Mandrick

Abstract In safety-critical environments such as piloting or air-traffic control, the study of mental overload is crucial to further reduce accident rates. However, researchers face the complexity of inducing an important amount of mental effort in laboratory conditions. Therefore, we designed a novel paradigm, named "Toulouse N-back Task" (TNT), combining the classical n-back task with a mathematical processing to replicate the multidimensional sustained high mental workload (MW) existing in many complex occupations. Instead of memorizing and comparing unique items, as in classical n-back task, participants have to memorize and to compare the results of mathematics operations. Twenty participants were tested with the TNT under three load factors (n = 0, 1, or 2) with functional Near-InfraRed Spectroscopy (fNIRS) and pupillary measurements. The results revealed that higher difficulty degraded the cognitive performance together with increased prefrontal oxygenation and an increase in pupil diameter. Hence, hemodynamic responses and pupil diameter were sensitive to different levels of TNT's difficulty. This paradigm could serve as a viable alternative to the classical n-back task and enable the progressive increase of the difficulty, for example, to test "high performer" individuals.

Keywords Neuroergonomics · Mental workload · Toulouse N-back task · Prefrontal cortex · Near-infrared spectroscopy · Pupillometry

1 Introduction

In the neuroscientific literature, one of the most used experimental tasks for the study of working memory (WM) is the n-back task [see 1]. It involves monitoring, updating, and manipulation of information and is assumed to heavily tax several

M. Causse (✉) · V. Peysakhovich · K. Mandrick
ISAE-SUPAERO (Institut Supérieur de l'Aéronautique et de l'Espace), Toulouse, France
e-mail: mickael.causse@isae.fr

M. Causse
Université Laval, Québec, Canada

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_16

key processes within WM. The parametric modulation of task difficulty via the load factor n ∈ {0, 1, 2, 3,…} without modifying the visual input or motor responses made it widely used as a tool to manipulate mental workload in applied studies [e.g., 2–6]. Moreover, cognitive performance during this task is sensitive to several conditions such as stress [7, 8] or fatigue [9]. Implicitly, one considers the WM load during n-back paradigm as a key component and a reasonable approximation of mental workload, which is likely true to some extent. However, when a study requires increasing the information processing demand, which is high in many safety-critical occupations (e.g., aircraft piloting), the n value can be too high (e.g., 3-back) and participants often disengage from the task. Therefore, we designed a novel paradigm, called "Toulouse N-back Task" (TNT), combining the classical n-back task with a mathematical processing to replicate the multidimensional sustained high mental WL existing in many safety-critical occupations (e.g., pilots). The task enables the progressive increase of the difficulty to reach a high mental load, for example, to test "high performer" individuals. Instead of memorizing and comparing unique items, as in classical task, the participants have to memorize and to compare the results of mathematics operations, computed beforehand. Mathematical problems are either subtractions or additions of 2-digits numbers, multiples of 5 for ease (e.g., 15 + 40; 90 − 35, etc.). The task also allows studying both tonic (during a block) and phasic (during mathematical processing) components of mental activity. This advantage of different time scale studies makes the task suitable for both low latency (such as phasic pupil response) and high latency signals (such as brain oxygenation or tonic pupil response). In this study, 20 participants were submitted to the TNT. Half of them were tested with pupillometry measurements and the other half were equipped with functional Near-InfraRed Spectroscopy (fNIRS). We measured the pupil diameter changes and concentration changes in oxygenated (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb) over the prefrontal cortex (PFC). They performed several blocks at three load factors (n = 0, 1, or 2) by indicating the presence/absence of a target by a button press. In the 0-back condition, an arbitrary value of "50" was defined as the target. During the 1-back (2-back) condition, the target was a result that matched the result presented one trial (two trials, respectively) before.

2 Materials and Methods

2.1 Participants

Twenty healthy volunteers participated in the experiment. Pupillometric recordings were performed on 10 participants in study 1 (2 women, age 26.2 ± 4.2 years) and fNIRS recordings were performed on 10 other participants in Study 2 (2 women, age 25 ± 5.1 years). Participants were split into two separate studies as simultaneous near-infrared spectroscopy and pupillometry measurement is complex due to wavelength overlap in the infrared band of both techniques. All were students at Ecole Nationale de l'Aviation Civile (ENAC) and Institut Supérieur de l'Aéronautique et de l'Espace (ISAE) in Toulouse, France. None reported neither affective or anxiety disorder, nor any neurological or cardiovascular disease. None was under medication that might affect the brain or autonomic functions. All volunteers reported normal auditory acuity and normal or corrected-to-normal vision. All participants gave written informed consent in accordance with local ethical board committee.

2.2 Toulouse N-Back Task

The novel mental arithmetic N-back task called Toulouse N-back Task (TNT) developed for this study combines a classical n-back task with mental arithmetic operations (Fig. 1). WM load varied between conditions with the n level. In each trial, volunteers were required to compute the result and compare it with either a fixed number (0-back), or the result obtained 1 (1-back) or 2 (2-back) trials before. Mathematical operations were either additions or subtractions of numbers always multiple of 5 (e.g., 15 + 40, 90 − 35). In addition to the three difficulty levels, rest period screens with "00 + 00" operations were presented. Volunteers did not give any response during this condition.

Operations were displayed in the center of a gray background. Participants were given a 2-button Cedrus response pad (RB-740, Cedrus Corporation, San Pedro, CA) and were asked to press either a green button if the result matched the target number or a red button if not. Participants had to give their responses as quickly as possible. The task was implemented in Matlab (MathWorks) using the Psychophysics Toolbox (Psychtoolbox 3, [10]).

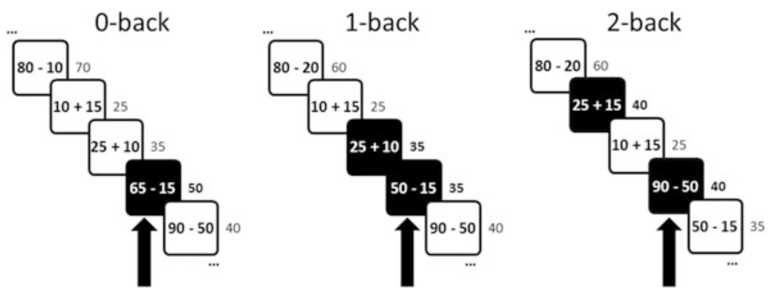


Fig. 1 Toulouse N-back task. An example of trials for the TNT, in which subjects computed mental calculations combined with a N-back task. The 0-, 1-, and 2-back task blocks lasted 36 s, interleaved with 18 s rest periods. Subjects responded to targets and non-targets by pressing one of two different buttons. During the 0-back (*left part* of the figure), a simple comparison of the current result with "50" was required. During the 1-back (*middle part* of the figure) and the 2-back (*right part*) tasks, the current result had to be compared with the one presented 1 or 2 trials before, respectively. *Black arrows* indicate target trials

The experiment included a total of 30 blocks, with 10 blocks for each of the three difficulty levels, presented in a pseudo-randomized order. A block lasted 36 s, and they were interleaved with 18 s rest periods. Stimuli were presented for 2 s with an inter-stimulus interval of 1 s. Each block contained 12 stimuli, with 4 targets in random positions. Following the training period with the TNT, participants rated the perceived difficulty using a DP15 scale [11]. The DP15 scale consists in a 15-point category scale, with 7 labels, from 2 (extremely easy) to 14 (extremely difficult), symmetrically placed around a central label 8—somewhat difficult).

2.3 Behavioral Measurements

Performance in the TNT was measured for each participant. TNT accuracy was calculated using the d-prime measure, computed as follows: Z (% hit rate)−Z (% false alarm rate). Additionally, response times were computed.

2.4 Study 1: Pupillometry Measurements

Participants were seated at approximately 70 cm from a 22″ computer screen (1680 × 1250). Ambient luminance was of 10 lx. During the whole experiment, participants' gaze position and pupil diameter were tracked using a remote SMI RED500 eye-tracker (SensoMotoric Instruments GmbH, Germany) at a sampling rate of 120 Hz. This device allows tracking the pupil despite small head movements. Before each run, participants performed a 5-point calibration procedure validated with 4 additional fixation points. The data acquisition routine used iViewX SDK to communicate with Matlab software. Identified blinks and short periods of signal loss were linearly interpolated. Then the signal was filtered with a two-pass 9-point filter (low-pass with a cutoff frequency of 5.9 Hz). Pupil diameter analyses were performed upon the median pupil diameter value for each block.

2.5 Study 2: Functional Near Infrared Spectrometry (fNIRS) Measurements

To illuminate the forehead, a CW fNIRS 16-channel headband model 100 fNIRS system (fNIRS Devices LLC, Photomac MD; http://www.fnirdevices.com) was used to obtain raw light intensity by specific dual wavelengths of 730 and 850 nm. Data were acquired at a sampling frequency of 2 Hz.

At the beginning of the experiment, participants were fitted with the fNIRS headband. The baseline was taken while they relaxed with eyes closed. fNIRS-PFC activity was recorded through the entire experiment. COBI Studio software (Drexel University) was used for data acquisition and visualization. The version 4.0 of fnirSoft software package was used for filtering, converting and analyzing data [12]. First, the raw optical density signals were converted to concentration changes of oxy-Hb and deoxy-Hb using the modified Beer-lambert law. At this point, few channels with no signal were rejected automatically and individually (for one subject channel 5 and 11; for one subject channel 11; for one subject channels 5 and 15). fNIRS data were then bandpass filtered using a FIR filter of order 20 and cutoff frequencies of 0.01 and 0.1 Hz. No detrending was applied.

To dissociate effects of N-back task difficulty (0-back vs. 1-back vs. 2-back) in each region of interests (ROI), we extracted the fNIRS response from each block for the left, the center, and the right PFC. Signals were normalized towards zero by subtracting the current signal with the first data point. Then, signals were averaged on all trials for each condition. For fNIRS data analysis, we compared concentration changes in oxy-Hb signals calculating the slope index as suggested by Mandrick et al. [13] during the 36 s of each block. This index reflects the magnitude of the PFC oxygenation response for each ROI and task difficulty.

2.6 Data Analysis

Results were analyzed using Statistica software (StatSoft). Group data were reported as means ± SD in the text and means + one SD in the figures. Normality and homoscedasticity of data were assessed using Kolmogorov-Smirnov. Data were analyzed using repeated-measures analyses of variance (ANOVA) for normally distributed variables (i.e., pupillary and fNIRS measurements, d’, response times). In the case of significant main effect or interaction, significant differences between conditions were identified using Tukey’s HSD post hoc tests. A significance level of $p < 0.05$ was used for all comparisons. Effect sizes are reported using partial eta-squared (η_p^2).

3 Results

3.1 Behavioral Data

Participants demonstrated lower d′ scores with increasing task difficulty ($F(2, 38) = 72.95$, $p < 0.001$). Post hoc comparison revealed significant differences between 0-back and 1-back and between 1-back and 2-back ($p < 0.001$ in both comparisons). Similar results were observed for the response times, which

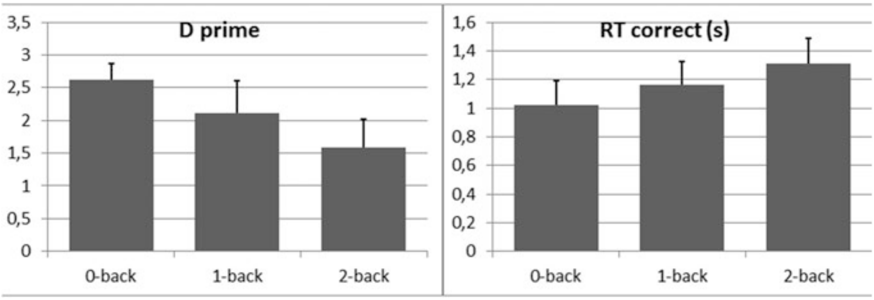


Fig. 2 Behavioral performance. N-back task scores for each level of difficulty (0-, 1-, and 2-back). *Error bars* indicate SD. *Left* d prime calculated as z (hit rate) − z (false alarm rate). *Right* reaction times. n = 20

increased with mental workload ($F(2, 38) = 38.94$, $p < 0.001$), see Fig. 2. Post hoc comparison also revealed significant differences between 0-back and 1-back and between 1-back and 2-back ($p < 0.001$ in both comparisons).

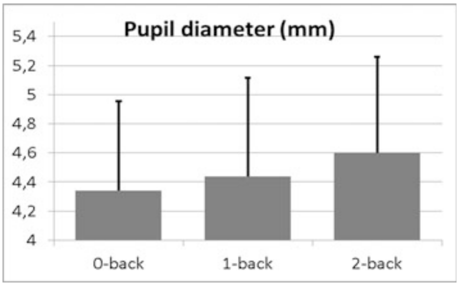
3.2 Pupillometric Results

Participants demonstrated larger pupillary responses across level of task difficulty ($F(1, 9) = 14.70$, $p < 0.001$). Post hoc comparison revealed significant differences between 0-back and 2-back and between 1-back and 2-back ($p < 0.001$ and $p < 0.05$ respectively), see Fig. 3.

3.3 fNIRS Results

We found a significant interaction between brain region and level of difficulty on slope index for changes in oxy-Hb concentration ($F(4, 36) = 6.71$, $p < 0.001$). Post hoc comparison revealed that difficulty did not significantly impact the center region of the PFC whereas significant differences were found on the right PFC with

Fig. 3 Mean pupil diameter according to the TNT levels of difficulty (0-, 1-, and 2-back). *Error bars* indicate SD. n = 10



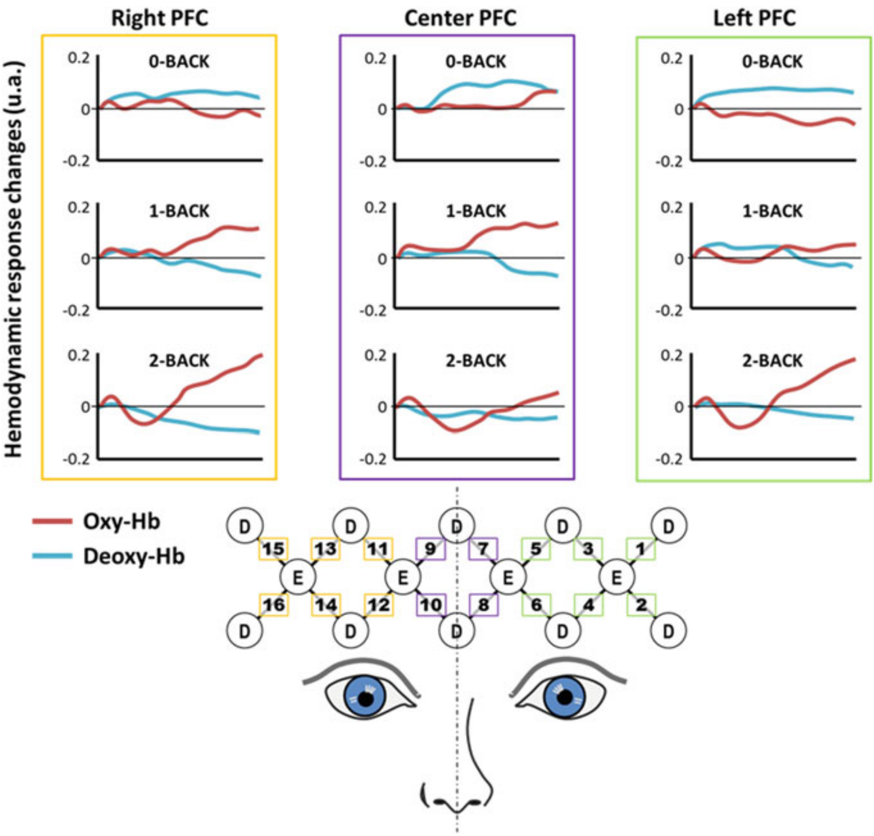


Fig. 4 *Up* grand average of all 10 subjects in the three load of the TNT for the three ROIs. *Red* and *blue lines* indicate the changes of oxy-Hb and deoxy-Hb respectively. The *square with the number* represent the 16 channels and the color of the square indicates the AOI in which the channel was included. Emitters are marked as E, while D indicates detectors. Areas underlying the channel are approximately over Brodmann's areas 10 and 46

a significant difference between 0-back and 2-back ($p < 0.001$). Moreover, oxy-Hb concentration changes, in the 2-back condition, localized in right and left PFC were higher than in the center PFC ($p < 0.001$ in both comparisons) see Fig. 4.

4 Discussion

In safety-critical environments such as piloting or air-traffic control, the study of mental overload is crucial to further reduce accident rates. However, researchers face the complexity of inducing an important amount of mental effort in the laboratory conditions. In this paper, we present a novel paradigm called Toulouse

N-back Task (TNT) that combines the classical n-back task with a mathematical processing to replicate the multidimensional sustained high mental workload (MW) existing in many safety-critical occupations.

Twenty participants were tested with the TNT under three load factors (n = 0, 1, or 2) with fNIRS and pupillometric measurements to validate the effectiveness of the protocol. The results revealed that higher difficulty degraded the cognitive performance together with an increased bilateral PFC oxygenation (demonstrated by larger changes in oxy-Hb concentration) and an increase in pupil diameter. The variations in TNT difficulty successfully impacted behavioral without the need to reach high n-level (e.g., 3-back), that can provoke a disengaging of the task. Consequently, the TNT enables the progressive increase of the difficulty to generate a high and sustained mental load, for example, to test "high performer" individuals (airline pilots, air traffic controllers). The task also allows studying both tonic (during a block) and phasic (during mathematical processing) components of mental activity. This advantage of different time scale studies makes the task suitable for both tonic (such as brain oxygenation or tonic pupil diameter presented in this work) and phasic signal analysis (such as task-evoked pupillary response). Finally, the study confirmed that prefrontal oxygenation and pupil response are sensitive to variations in mental effort. For example, measurements of mental effort with fNIRS can help developing an effective design for future cockpit concepts.

References

1. Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E.: N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. **25**, 46–59 (2005)
2. Brouwer, A.-M., Hogervorst, M.A., Van Erp, J.B., Heffelaar, T., Zimmerman, P.H., et al.: Estimating workload using EEG spectral power and ERPs in the n-back task. J. Neural Eng. **9**, 045008 (2012)
3. Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., et al.: Optical brain monitoring for operator training and mental workload assessment. Neuroimage **59**, 36–47 (2012)
4. Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., et al.: Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. Front. Hum. Neurosci. **7**, 935–940 (2013)
5. Durantin, G., Gagnon, J.-F., Tremblay, S., Dehais, F.: Using near infrared spectroscopy and heart rate variability to detect mental overload. Behav. Brain Res. **259**, 16–23 (2014)
6. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., et al.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat. Space Environ. Med. **78**, B231–B244 (2007)
7. Schoofs, D., Preuß, D., Wolf, O.: Psychosocial stress induces working memory impairments in an n-back paradigm. Psychoneuroendocrinology **33**, 643–653 (2008)
8. Qin, S., Hermans, E., Marle, H., Luo, J., Fernández, G.: Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. Biol. Psychiatry **66**, 25–32 (2009)

9. Sasahara, I., Fujimura, N., Nozawa, Y., Furuhata, Y., Sato, H.: The effect of histidine on mental fatigue and cognitive performance in subjects with high fatigue and sleep disruption scores. Physiol. Behav. **147**, 238–244 (2015)
10. Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., et al.: What's new in psychtoolbox-3. Perception **36**, 1 (2007)
11. Delignières, D., Famose, J.-P., Genty, J.: Validation d'une échelle de catégories pour la perception de la difficulté. Revue STAPS **34**, 77–88 (1994)
12. Ayaz, H., Izzetoglu, M., Shewokis, P.A., Onaral, B.: Sliding-Window Motion Artifact Rejection for Functional Near-Infrared Spectroscopy, pp. 6567–6570. IEEE, USA (2010)
13. Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., et al.: 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)

An Eye Movement Research on 2D Interactive Game Design

Na Lin, Lei Feng, Tian Kang, Shiyu Fan and Yunhong Zhang

Abstract In order to provide suggestions for further game design, an eye movement experiment was carried out to investigate how the player's visual focus was distributed in shooter game. In the experiment, the 2D video games were taken as material. The participant were required using visual fixation to finish the shooter game by eye controlling, which the fixation times and fixation time were recorded by eye-tracker. The results showed that the central region was the best location for shooting or the main elements of paying attention. The result was consistent with the general visual habits. On the contrary, the elements located in top left and top right regions were hardly captured by players' attention. According to classic composition prototypes, we could consider putting the important elements of 2D interactive video games in the central region, and distributing consist with human visual processing habits which was along with the directions of diffusing or concentrating of tension in triangle, square or circle prototypes.

Keywords Eye movement · 2D interactive video games · Visual focus distribution

1 Introduction

With the development of technology and the diversification needs of entertainment, more and more teams are committed to providing more sophisticated videos and games to their consumers. In the development process of videos or games, in order to reduce the risk of development and enhance the overall quality of videos and games effectively, developers and investors has attached importance to the user

N. Lin · L. Feng · Y. Zhang (✉)
Human Factor and Ergonomics Laboratory,
China National Institute of Standardization, Beijing, China
e-mail: zhangyh@cnis.gov.cn

T. Kang · S. Fan
Wanda Cultural Tourism Planning & Research Institute, Beijing, China

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_17

experience research. With the development of eye movement theory and tracer technique, the eye tracking data can reflect the visual processing characteristic data, which has significant value to explain the psychological mechanism of cognitive processing. The relative theory and technique are now widely used in advertisement, website and video game design.

The study has proved that more than 80 % information is acquired by the visual system, and the vision range of human beings was limited. Objects can be clearly seen only in the central fossa, which was about in 2° visual angle, however, objects can be identified in the range of 30° field near the center of vision, where the optic nerve accept stimulation [1]. Therefore, the human visual system was needed to be able to quickly and accurately focus on important and crucial visual objects in the visual field, it was particularly important when looking at the image or video with complicated information elements, this process was known as visual attention [2].

Wilhelm Wundt pointed out that visual attention has its limitations, accordingly, selective attention theory was put forward. Selective attention helped individuals from information overload in cognitive processing. When it comes to which specific theory the selective attention based on, several theories have research support. Some scholars believe that different type of tasks, requirements and uses' behavioral objectives lead to different attention selection [1]. Ulric Neisser, the father of cognitive psychology, considered that "what the audiences see depends on how their attention distributed". In other words, depends on their expectation and perceptual detection [1]. Therefore, there were researches put forward the top-down and bottom-up models of visual attention. The top-down model proposes to set up visual expectations in advance, and completes the selection of interested regions in images or videos under these expectations, then continues the subsequent processing. The bottom-up model advocates that selective attention was based on data driven. Visual processing system extract some low-level features such as color, brightness and direction before deep processing, and create saliency map for each kind of features [3]. In visual processing, the top-down and bottom-up attention models usually work together.

In terms of the specific viewing mode, audiences have the expectation about the location and orientation of the visual elements when they watch visual compositions, therefore,audiences can view the materials in a relatively fixed pattern when there are certain fundamental mode of composition. The study revealed audiences follow a certain scanning sequence order when they view static images. The top three are: (1) The center of the picture; (2) Element takes the largest area; (3) Element have the maximum contrast with the background, on brightness, color [4], size, direction etc. [1].

Based on these theories, this study used a 2D interactive video game as research material, studied the composition of key elements in the video to verify previous research results, and extract the pattern of distribution in accordance with the pattern of human being`s cognitive processing, in order to provide suggestions for further design.

2 Method

2.1 Experimental Material

The 2D interactive games based on the background of "Journey to the west" were as experimental material, which included 6 interactive game units. 6 interactive game units were taken from the classic chapters of Journey to the west, which the images in the game were well known to the target population. The effective window size of each game unit video was 1440 × 1060 pixels and it lasted about 35 s.

2.2 Subjects

Eighteen volunteers from 6 to 45 years old (9 males and 9 females, mean age = 22.22, standard deviation of age = 10.74) were recruited and paid to participate in the experiment. According to the age, they were divided into 3 groups, including groups of children group (5 volunteers from 6 to 12 years old, 2 boys and 3 girls, M = 9.8 SD = 1.79), adolescent group (4 volunteers from 13 to 18 years old, M = 16.75, SD = 2.50, 2 boys and 2 girls), adult group (9 volunteers from 18 to 45 years old, 5 males and 4 females, M = 31.56, SD = 5.61). All of them had more than 1.0 normal or corrected-to-normal visual acuities and healthy physical conditions, without ophthalmic diseases. They did not have any history of neurological and mental diseases.

2.3 Apparatus

The eye tracker that recorded the eye movement data was Tobii X2-60 by Tobbi Corporation and its sampling rate was 60 Hz. The experimental apparatus was composed of an eye tracker, a 2D display and a host that was installed the Tobii Studio. Experimental host is Dell workstation with Xeon X5690 INTEL, basic frequency 3460 MHz, 24G memory and Quadro NVIDIA 6000 professional graphics hardware. The host were connected to a 19 in. 2D display (AOC IPS display) with a maximum resolution of 1920 × 1080 and 60 Hz refresh rate. And participants continued the experiment by mouse and keyboard, and the eye tracker was placed below the display to collect the eye movement data.

2.4 Experiment Procedure

Experiment was carried out in a quiet and bright environment. After arriving at the laboratory, participants signed the informed consent and completed a general survey about their demographic information. Then, experimenter introduced the process and requirements to participants and adjusted his/her gesture to find the appropriate distance and height to collect data by eye tracker. The viewing distance between participant and the screen was about 60–70 cm. After calibration, 2D interactive game units were presented to participants and their eye movement data was recorded. Participant was required using visual fixation to finish the shooter game by eye controlling and try his/her best to watch the position that he/she want shooting. Participant was informed about the possibility of error. Experimenter observed and recorded participant` behavioral and language feedback during the whole process. Few of participants detected that sometimes appeared inaccurate shooting, but none of them doubted about reality of eye control operation.

2.5 Data Analysis

The eye movement and behavior data were recorded when the player watching the video. Each segment was regarded as a fragment. Researcher created the scene for each fragment and used static picture instead of dynamic video by data processing program in Tobii Studio, investigated visual focus distribution of each fragment by superimposed fixation times on these pictures. We delineated the available region of attention or shooting according to the fixation results. By the initial observation, we delineated the valid attention/shooting region that concluded 90 % fixation counts, which included 1255 $\times$ 855 pixel range. In the valid attention/shooting regions, we divided the region into 9 AOI: 3 parts (up/centerline/down) in vertical direction $\times$3 parts (left/midcourt/right) in horizontal direction. The fixation times and fixation time data of 9 AOI were analyzed by IBM SPSS 20 Statistics software (IBM-SPSS Inc. Chicago, IL). The repeated-measures analysis of variance (rANOVA) was applied to analysis the eye movement data. Greenhouse–Geisser correction was applied to p values associated with multiple df repeated measures comparisons where appropriate.

3 Results

3.1 The Means of Fixation Time and Fixation Times in Game Unit 1

A repeated-measure ANOVA was applied to fixation time data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The simple effect analysis showed that the fixation time of top central axial region was obviously longer than top left region and top right region ($p < 0.05$); that of central region was longer than left region and right region ($p < 0.05$); that of left bottom region was obviously longer than central bottom region ($p < 0.05$). And, the fixation time of left central region and left bottom region were longer than top left region ($p < 0.05$); that of central region was longer than central bottom region and top central region ($p < 0.05$); on right region, right central region and right bottom region were longer than top right region (ps < 0.05) (Table 1).

A repeated-measure ANOVA was applied to fixation times data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The results of simple effect analysis showed that the fixation times distribution of central region was obviously more than top left region and top right region ($p < 0.05$); that of center region was more than left region and right region ($p < 0.05$); that of central bottom region was more obviously than left bottom region and right bottom region ($p < 0.05$). And, the fixation times distribution of central region and left bottom region were obviously more than that of top left region (ps < 0.05); on central region, that of central region was the most, distribution of top central region was secondary and central bottom region was the least ($p < 0.05$); on right region, fixation times of distribution of central region and right bottom region were much more than top right region (ps < 0.05), the fixation times distribution difference between central region and right bottom region was not significant (Table 2).

Table 1 Results of fixation time by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	138.89	2	69.44	12.20	0.000
Left/midcourt/right	76.28	2	38.14	10.55	0.000
Up/centerline/down × left/midcourt/right	139.18	2.05	67.95	11.43	0.000

Table 2 Results of fixation times by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	1893.70	2	946.85	16.98	0.000
Left/midcourt/right	958.16	2	479.08	11.93	0.000
Up/centerline/down × left/midcourt/right	1833.02	1.91	960.65	12.56	0.000

3.2 *The Means of Fixation Time and Fixation Times in Game Unit 2*

A repeated-measure ANOVA was applied to fixation time data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The simple effect analysis showed that up/centerline/down directions: on top region, fixation time of top central axial region was obviously longer than that of top left region and top right region; on central region, fixation time of central axial region was longer than left central region and right central region; on bottom region, fixation time of central bottom region was obviously longer than left bottom region and right bottom region (ps < 0.05). And on left region, fixation time of left central region was longer than top left region and left bottom region; on central region, fixation time of central axial region was longer than top central region and central bottom region; on right region, right center region was longer than top right region and right bottom region (ps < 0.05) (Table 3).

A repeated-measure ANOVA was applied to fixation times data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The results of simple effect analysis showed that up/centerline/down directions: on top region, fixation times distribution of center region was the most, top left region was secondary, top right region was the least; on center region, fixation times distribution of center region was more than left region and right region; on bottom region, fixation times distribution of center bottom region was more obviously than left bottom region and right bottom region (ps < 0.05). Left/midcourt/right directions: on left region, fixation times distribution of center region was the most, distribution of top left region was secondary and left bottom region was the least; on center region, fixation times distribution of center region

Table 3 Results of fixation time by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	434.38	1.38	313.79	45.51	0.000
Left/midcourt/right	251.73	1.09	230.21	46.10	0.000
Up/centerline/down × left/midcourt/right	274.24	1.35	203.04	30.30	0.000

Table 4 Results of fixation times by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	7816.80	2	3908.40	47.80	0.000
Left/midcourt/right	4728.97	1.25	3797.56	69.54	0.000
Up/centerline/down × left/midcourt/right	4121.82	1.71	2404.93	35.16	0.000

was more than distribution of top center region and center bottom region; on right region, fixation times of distribution of center region was much more than right bottom region and top right region (ps < 0.05) (Table 4).

3.3 *The Means of Fixation Time and Fixation Times in Game Unit 3*

A repeated-measure ANOVA was applied to fixation time data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The simple effect analysis showed that fixation time of top left region was obviously longer than top right region; that of center region was longer than left center region and right center region; that of left center region was longer than left bottom region and top left region; that of center region was longer than center bottom region and top center region; that of right center region was longer than right bottom region and top right region (ps < 0.05) (Table 5).

A repeated-measure ANOVA was applied to fixation times data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The results of simple effect analysis showed that up/centerline/down directions: on center region, fixation times distribution of center region was more than left center region and right center region. Left/midcourt/right directions: on left region, fixation times distribution of center region was more than distribution of top left region and left bottom region; on center region, fixation times distribution of center region was more than distribution of top center region and center bottom region; on right region, fixation times of distribution of center region was much more than right bottom region and top right region (ps < 0.05) (Table 6).

Table 5 Results of fixation time by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	581.18	1.23	471.87	35.33	0.000
Left/midcourt/right	209.89	1.22	171.97	30.17	0.000

Table 6 Results of fixation times by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	8318.74	1.34	6225.23	36.83	0.000
Left/midcourt/right	1794.52	2	897.26	27.91	0.000
Up/centerline/down × left/midcourt/right	2782.65	4	695.66	28.52	0.000

3.4 The Means of Fixation Time and Fixation Times in Game Unit 4

A repeated-measure ANOVA was applied to fixation time data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The simple effect analysis showed that up/centerline/down directions: on top region, fixation time of top central axial region was obviously longer than top left region and top right region; on central region, fixation time of central axial region was obviously longer than left central region and right central region; on bottom region, fixation time of central bottom region was obviously longer than left bottom region and right bottom region. Left/midcourt/right directions: on left region, fixation time of left central region was longer than top left region and left bottom region; on central region, fixation time of central axial region was longer than top central region and central bottom region; on right region, right center region was longer than top right region and right bottom region (ps < 0.05) (Table 7).

A repeated-measure ANOVA was applied to fixation times data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The results of simple effect analysis showed that up/centerline/down directions: on top region, central region was obviously more than top left region and top right region; on central region, fixation times distribution of center region was obviously more than left center region and right center region; on bottom region, fixation times distribution of central bottom region was more obviously than left bottom region and right bottom region. Left/midcourt/right directions: on left region, fixation times distribution of central region was obviously more than

Table 7 Results of fixation time by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	309.73	2	154.86	26.29	0.000
Left/midcourt/right	345.68	1.25	275.74	47.67	0.000
Up/centerline/down × left/midcourt/right	190.91	2.00	95.35	15.08	0.000

Table 8 Results of fixation times by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	5130.79	2	2565.39	30.26	0.000
Left/midcourt/right	5175.32	1.33	3887.62	47.16	0.000
Up/centerline/down × left/midcourt/right	2451.28	2.19	1117.03	15.02	0.000

distribution of top left region and left bottom region; on central region, fixation times distribution of central region was more than distribution of top central region and central bottom region; on right region, fixation times of distribution of central region was much more than and right bottom region and top right region (Table 8).

3.5 *The Means of Fixation Time and Fixation Times in Game Unit 5*

A repeated-measure ANOVA was applied to fixation time data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) ($ps < 0.001$), and there was a significant interaction ($p < 0.001$). The simple effect analysis showed that up/centerline/down directions: on top region, fixation time of top central axial region was the longest, up right region was the secondary and top left region was the shortest; on central region, fixation time of central axial region was the longest, left central region was the secondary and right central region was the shortest; on bottom region, fixation time of central bottom region was obviously longer than left bottom region and right bottom region. Left/midcourt/right directions: on left region, fixation time of left central region was the longest, top left region was the secondary and left bottom region was the shortest; on central region, fixation time of central axial region was the longest, top central region was the secondary and central bottom region was the shortest; on right region, right center region was longer than top right region and right bottom region ($ps < 0.05$) (Table 9).

A repeated-measure ANOVA was applied to fixation times data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) ($ps < 0.001$), and there was a significant interaction ($p < 0.001$). The results of simple effect analysis showed that, on top region,

Table 9 Results of fixation time by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	331.96	1.21	274.75	36.23	0.000
Left/midcourt/right	455.37	1.05	432.02	75.19	0.000
Up/centerline/down × left/midcourt/right	386.53	1.28	301.05	27.70	0.000

Table 10 Results of fixation times by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	5382.36	2	3458.32	50.98	0.000
Left/midcourt/right	7561.25	1.12	6774.42	80.37	0.000
Up/centerline/down × left/midcourt/right	5541.81	1.85	2995.86	42.94	0.000

fixation times distribution of central region was the most, top left region was the secondary and top right region was the least; on center region, fixation times distribution of center region was the most, left center region was the secondary and right region was the least; on bottom region, fixation times distribution of central bottom region was more obviously than left bottom region and right bottom region. And, on left region, fixation times distribution of central region was the most, distribution of top left region was the secondary and left bottom region was the least; on central region, fixation times distribution of central region was the most, distribution of top central region was secondary and central bottom region is the least; on right region, fixation times of distribution of central region was much more than distribution of top right region and right bottom region ($ps < 0.05$) (Table 10).

3.6 *The Means of Fixation Time and Fixation Times in Game Unit 6*

A repeated-measure ANOVA was applied to fixation time data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) ($ps < 0.001$), and there was a significant interaction ($p < 0.001$). The simple effect analysis showed that up/centerline/down directions: on central region, fixation time of central axial region was the longest, left central region was the secondary and right central region was the shortest; on bottom region, fixation time of central bottom region was obviously longer than left bottom region. And, on left region, fixation time of left central region was the longest, left bottom region was the secondary and top left region was the shortest; on central region, fixation time of central axial region was the longest, central bottom region was the secondary and top central region was the shortest; on right region, right center region was the longest, right bottom region was the secondary and top right region was the least ($ps < 0.05$) (Table 11).

Table 11 Results of fixation time by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	173.24	2	86.62	30.62	0.000
Left/midcourt/right	35.32	2	17.66	11.39	0.000
Up/centerline/down × left/midcourt/right	46.83	1.63	28.75	11.95	0.001

Table 12 Results of fixation times by repeated measure variance analysis

	SS	df	MS	F value	Sig. (two-tailed)
Up/centerline/down	3363.54	2	1681.77	29.66	0.000
Left/midcourt/right	705.98	2	352.99	14.74	0.000
Up/centerline/down × left/midcourt/right	685.96	1.96	349.91	13.22	0.000

A repeated-measure ANOVA was applied to fixation times data of the different conditions, a significant main effect of vertical direction (up/centerline/down) and horizontal (left/midcourt/right) (ps < 0.001), and there was a significant interaction ($p < 0.001$). The results of simple effect analysis showed that, on top region, fixation times distribution of central region was more than top left region and top right region; on center region, fixation times distribution of center region was the most, left center region was the secondary and right region was the least; on bottom region, fixation times distribution of central bottom region was more obviously than left bottom region and right bottom region. And, on left region, fixation times distribution of central region was more than distribution of top left region and left bottom region; on central region, fixation times distribution of central region was the most, distribution of central bottom region was secondary and top central region is the least; on right region, fixation times of distribution of central region was the most, distribution of right bottom region was secondary and top right region was the least (Table 12).

4 Discussion

There was a study proved that participants' consciousness can cause by the composition design principles of balance, unity, dynamic and comparison, the purpose of design principles was to build visual aesthetic feeling and variation, so that they could absorb the audiences' attention. Balance principle involved in visual weight of every region. The purpose of uniformity principles was to let participants organizing different visual elements in visual layout that unified main expression which could make a higher information expression efficiency to reach the purpose of visual guidance. Dynamic, which was using lines and shapes in layout to guide routine of participants' sight moving, thus to produce effects of dynamic and changes. Visual element designs could for the principles of guiding eyesight moving directions and expressing information in order. Comparison was to strengthen opposite relation of special visual elements to highlight the dynamic effect and let it to be visual emphasis that was the final purpose. Three basic original shapes in the comparison was circle, square and triangle, which included visual tension that inter-traction of internal and external tension to balance the strengthen. Figure 1 as follows: Due to the strengthen of inward or outward tension, participants felt visual distance and formed a kind of harmonious movements. In the view

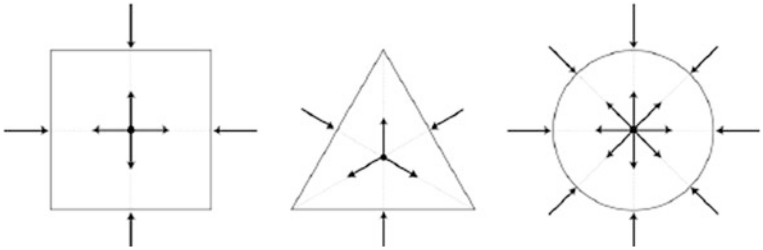


Fig. 1 Visual tension for basic prototypes

of Fig. 1, no matter tension of inward or outward, the tension was concentrated to the center or diverged to outside, these two kinds of tension located in the center of modelling [1].

Summarizing the statistics result (Fig. 2), following is the tendency:

- Game unit 1: fixation times and fixation time concentrated, present "凸" type distribution (not including center axial region). Players generally thought it was difficult for target tracking, and fixation times and fixation time distribution was scattered. This may due to the main target movement speed for the sight of players' tracking smoothly, study revealed that, in most cases, rising head was necessary when watching angle beyond 12°, Catching targets with saccade at the speed of targets' moving rate beyond 30°/s [5]. In addition, the secondary targets were not inconformity for design purpose. Location of elements distributed was too low.
- Game unit 2: Fixation times and fixation time distribution presented by cross shape, but the fixation times of top left region was obviously more than the top right regions, it showed the fixation times distribution tended to left slightly. In this unit, there are important targets in the top left region. According to distribution difference of fixation times and fixation time, it was unsuccessful to set important note or attack on the top left region, these elements can hardly capture

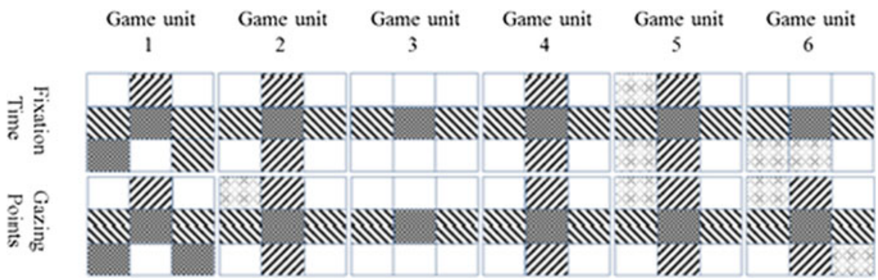


Fig. 2 Distribution diagram of fixation time and fixation times about rules explanation and interaction games. *Top right to left bottom diagonals* stand for vertical comparison results, *top left to right bottom diagonals* stand for horizontal comparison results. Deep density slash (*darker*) show longer fixation time or more fixation times

players' attention, then effective attack is unable formed, even though a small number of players can find targets.

- Game unit 3: Fixation times and fixation time focused on center region, distribution as—type. Players didn't have deep imagination, it may because the layout of the interactive game was too simple and easy.
- Game unit 4: Fixation times and fixation time distribution presented by cross shape. Studies suggested that purpose of visual scanned was searching understanding by "proper order" to link up images, so establish visual orderliness of participants to reach more efficiently and understanding the meaning quickly [1]. This unit accorded with the design elements and had a good feedback from players.
- Game unit 5: Fixation times and fixation time arranged in cross shape. In terms of fixation time, on top left region and left bottom region were much more than the top right region and right bottom region. Fixation times distribution on the top left region distribution much more than top region, distribution on left region was presented heavy. According to the balance principle, there were some optimization space for the attention/attacked elements.
- Game unit 6: Fixation time distribution mainly concentrated in the central region, fixation time on the central bottom region and left region spent more time than right region. Distribution of fixation times presented by cross shape, distribution of right bottom region was much more than left bottom region. Fixation times and fixation time distribution was inconsistent. In the unit, players' feedback that the discrimination of target and false one was difficult and could cause the mistake easily. Inconsistent tendency of fixation times and fixation time distribution might reflect the difficulties, more fixation times distribution reflected players' attention distribution, and more fixation time reflected player's shooting purpose.

5 Conclusion

In conclusion, central was the best area for position/attack elements. This result was in accordance with the general visual habits and previous conclusion obtained from researches using print ads as experimental materials. According to classic composition prototypes, they were triangle, square or circle prototypes (see Fig. 1). The "+" type distribution of the fixation times and fixation time in this study were similar to square prototype; "⊤" or "⊥" type distribution of the fixation times and fixation time were similar to triangle prototype; "凸" type distribution could be optimized to square or circle prototype. Therefore, we could consider centering the elements on the central region in 2D interactive video games, distributing along with the directions of diffusing or concentrating of tension in triangle, square or circle prototypes.

Acknowledgments The authors would like to gratefully acknowledge the support from the National Key Technology R&D Program of the Ministry of Science and Technology (2014BAK01B03), China National Institute of Standardization through the "special funds for the basic R&D undertakings by welfare research institutions" (522016Y-4483).

References

1. Wang, J., Zhang, J.: Visual Guide in the Research on the Application of Plane Design. D. Qiqihar University, China (2013)
2. Ye, G., Du, Q.: ROI Extraction for Stereoscopic Video Based. D. ZheJiang University, Hangzhou (2013)
3. Chen, Y., Zhou, J., Zheng, L.: The Research of Interested Objects Detection Based on Visual Attention. D. TangMeng University (2014)
4. Han, Yuchang: Observe the sequence of eye movements in different shapes and colors. J. Psychol. Sci. **20**(40–43), 95–96 (1997)
5. Yan, B., Yuan, X.: The Methods of Eye Movement Research on Web Usability Test. D. Beijing University of Posts and Telecommunications, Beijing (2011)

Part IV
Theoretical Advances in Cognitive Engineering and Neuroergonomics

Cognitive and Emotional-Motivational Implications in the Job Design of Digitized Production Drilling in Deep Mines

Mohammed-Aminu Sanda

Abstract The aim of this paper was to understand the influence of cognitive and emotional-motivational aspects of task complexity on workers performances in high-technology driven drilling activity in a deep mine. Data was collected by observing and video recording miners' engaged in two separate production drilling activities, using two Boomers simultaneously. Based on the analysis, it is found that the workers encounter cognitive challenges in their ability to process information marked on rock surfaces for the positioning of the boomers, resulting in added complexity to their drilling tasks. The workers' were also found to have issues with the quality of their designed job environment, and which emotional-motivational challenge also added to their tasks complexity. It is concluded that by understanding the emerging cognitive and emotional-motivational aspects of task complexities, future design processes of a friendly and performance enhancing work environments and technologies could evolve for efficient and effective human work.

Keywords Cognitive complexity · Emotional-motivational complexity · Job design · Production drilling · Deep mine

1 Introduction

In the contemporary world of job design, the automation of the work system is viewed as a key enabler for increasing productivity per worker. The expectation here is to make the future work environments of deep mines become a substantially different place than what prevails today. In this vein, much progress has been made in modeling intelligent production systems of the future mine for them to become

M.-A. Sanda (✉)
University of Ghana Business School, P. O. Box LG 78,
Legon, Accra, Ghana
e-mail: masanda@ug.edu.gh; mohami@ltu.se

M.-A. Sanda
Luleå University of Technology, 97187 Luleå, Sweden

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_18

enablers for learning and collaboration across organizational borders. With such intelligent system and work organization in place, the mining industry is to be seen as a place where cooperation, skills development and gender equality are key components [1]. By this perspective of the prevalence of an enabling environment, both the youth and women might view mining activity as an attractive job and the work environment as safe, and this could go a long way in breaking the unequal gender balance that exists in most mines today [1]. Even if the idea of a holistic perspective on production systems is commonplace in most research areas of today, there is a true challenge in multidisciplinary research that reconnect the research fields and their theories, methods, ideas and results [1]. This observation reinforces the prevailing problem of developing a holistic work organization model with the human at the centre to guide the future integration of the technical, organizational and human systems of work design. It also reinforces the challenge of developing deep-specialized knowledge in areas associated with each of these systems (i.e. technical, organizational and human) that could contribute towards the efficiencies and effectiveness of employees in future deep mines.

Starting from the mid 1980s, a series of technological advances have been made in the world of mining, and whose applications have contributed greatly in changing the working conditions and requirements of workers in deep mine. But using the argument of [2] as a point of departure, it could be argued that though the technological advances in the mining industry may have positively affected workers' tasks and the efficiency of their work performances, such advances may not have reduced the cumulative effect of complexity inherent in the workers job design. In job designs, task complexity is a major factor that is known to create a challenge for operator performance. In this respect therefore, identification of the different complexities that are associated with workers job design in technologically advanced deep mines must be seen as a pre-requisite for enabling an enhance employees performance. By implication, a digitized production drilling job design in deep mines that highly integrates the work organization and human organizational activity components that supports high productivity as well as good working and social conditions could be designed by understanding the relationship between the behavioral and physical domains of the workers during the task performance. Based on the realistic notion that one cannot perform a complex motor task without significant mental effort and concentration, and that the relationship between these different components of task (i.e. motor and cognitive) is critical in evaluating the complexity associated, not only with the cognitive and motor aspects of high-technology driven drilling, but also with its emotional-motivational components, it becomes imperative that the complexity challenges in the job design of a digitized production drilling activity are understood. In a study on drilling activity in deep mining, [3] found that the subject-oriented activity of the drilling task entailed social collaboration between the operators' physical activity (i.e. production drilling operations) and mental activity (i.e. the simultaneous observation of the production drilling work and listening to communication models). According to [3], the operators (miners) were apparently happy with the communication models they were using in their work environment. The purpose of this study therefore, is to identify and understand the cognitive and emotional-motivational aspects of task complexity in the conduct of a high-technology driven drilling activity, and the relevance of such learning towards mediating task complexity in the future design of an efficient drilling job in deep mines, and by implication, make the human-aspect of the drilling job design very efficient.

2 Literature Review

The complexity of task has generally been looked at from the perspectives of the design process [4], technical systems [5], and behavioral standpoint [6]. The aspect of task complexity that contemporarily requires attention is that which describes the relationship between the behavioral and physical domains workers during task performance [7]. Viewing complexity as the converse of simplicity, it can be argued that lower complexity will result in greater simplicity. Based on this logic, the concept of task complexity is seen to be of critical importance for usability evaluation [7]. Different aspects of task complexity have been outlined by different authors. According to [8], task complexity in a multiple-choice task must be measured by the number of alternatives available for performing the task. In a sense, the number of production rules or condition-action pairs in a task that the task-undertaker has to learn must serve as the major criterion of task complexity [9]. It is suggested by [10] that for an inspection task, the number of different fault types must be used as a measure of task complexity. Thus, from the perspective of [7], task complexity may depend on the quantity of task elements and the specification of the interactions of different task elements. By implication, task complexity can be defined in terms of the number of static and dynamic components of the task and interaction among these components [7]. In this regard, the degree to which a task is viewed as unpredictable, and also the uncertainty associated with that unpredictability are important factors that influence task complexity. It is explained by [7] that uncertainty in a task depends on both objective and subjective characteristics of the task. According to [7], 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 [7]. Thus, the specificity of an operator's memory workload, as [7] noted, is one important source of task complexity. This perspective is supported by the argument of [11] that a situation where an operator has to deal with more than three intermediate data on dynamic objects in memory should be subjectively evaluated as a very difficult situation which produces errors, while the duration for which information is kept in memory should also be considered a critical factor.

According to [7], the specificity of the extraction of information from long-term memory also influences any evaluation of task complexity, and the degree of familiarity of retrieving such information is an important factor. It is explained by [7] that if the information that is retrieved in the performance of task by an operator has similarities with information considered task-irrelevant, then the complexity of the task is increased. The sensory-perceptual characteristics of signals also influence the complexity of task performance [7]. The complexity increases if the operator's perceived signals are in the threshold area, with the threshold data representing the extreme performance which the individual is capable of, meaning that he/she must exert the maximum effort in order to perform successfully [7]. Therefore, the complexity of a task will increase when an operator is required to alter the stereotypical actions of the designed task. It is noted by [7] 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.

According to [7], the level of concentration of attention is also a critically important characteristic of task complexity, in the sense that the more an operator concentrates on a task, the more complex will be the task. Underlined by this logic, task complexity is the basic characteristic determining demands on the cognitive components of activity during task performance [7]. Therefore, the concept of complexity can be applied to the motor aspect of task, or more specially, to the mental regulation of movements [7]. In the real world of work, as [7] explain, a complex motor task cannot be performed by an operator without significant cognitive (mental) effort and concentration. As such, the relationship existing between these different components of a task is critical when evaluating the complexity associated not only with the cognitive and motor aspects of the activity, but also with the emotional-motivational components [7]. This is because, emotional tension and motivational force increase as task complexity increases. Therefore, it is important to distinguish between those cognitive aspects of complexity that depend on the specificity of information processing and those emotional-motivational aspects of complexity that reflect the energetic aspects of activity. These two aspects of complexity are interdependent and influence each other [7].

3 Methodological Issues

As it has been outlined by [7], 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 [12–13]. It is argued by [7] that task complexity cannot be successfully evaluated by such methods, principally because they employ incommensurable units of measure. According to [7], the quantitative method of task complexity evaluation 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 [7]. Thus, task complexity can be evaluated both experimentally and theoretically [7]. 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 [7]. 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 [7]. But as [7] pointed out, the motivational aspects of an activity are usually ignored in the design process. These functional blocks of self-regulation which are connected with the evaluation of task difficulty and significance, play a central role in integrating the cognitive and motivational aspects of activity [7]. Thus arguing from the perspectives of [7], it is deduced that the fundamental notions of task complexity, difficulty and significance in the technology-driven drilling activity in deep mines, and the concentration of attention on the conduct of the drilling 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.

3.1 Data Collection Procedure

Data was collected in an underground mine by observing, video recording and interviewing miners' engaged in two separate production bolting activities using two Boomers simultaneously. A Boomer is a highly automated and computer-based programmable robotic arm attached to the front of highly reinforced tractors that is computer guided by an operator to carry out rock drilling operations in a mining activity.

3.2 Data Analysis Procedure

Using the systemic analytical approach, the cognitive aspect of complexity that depended on the specificity of information processing in the bolting activity, and those emotional-motivational aspects of complexity that reflected the energetic aspects of the bolting activity were evaluated.

4 Results Analyses

4.1 Cognitive Aspect of Complexity in the Drilling Activity

Analysis of the cognitive aspect of complexity that depended on the specificity of information processing in the bolting activity showed that the operators are challenged in their ability to focus on the two boomers that function simultaneously, as highlighted in the following comment by a miner.

The technology with 2 boomers is excessive for one operator to handle. This is because of the difficulty in focusing in the ability to observe the simultaneous functioning of all the 4 boomers.

The operators were also found to experience challenges in their ability to clearly process information marked on the rock surfaces (drilling spots) due to parts of their operational views being blocked by metal guards provided as reinforcement in the tractor cabin design, and which spots they were expected to use as guide for the start of the drilling operation using the robotic drilling tool (boomers). A bid to process such information correctly requires that an operator engages in the act of avoiding the metal guards' blockage by stretching the neck numerously in order to get clear view of the drilling spots on the walls, but which action results in operators' developing neck pains, as highlighted in the following comment by a miner.

The provision of the metal guards as a protective measure is good. Yet, due to operator's view-blockage by the metal guards of the Technology's glass screen, the operator is forced to engage in the stretching of the neck in order to get a clearer view of the marked drilling spots on walls. This has resulted in operators suffering pains in their necks.

Additionally, an operator had to continuously shift the upper body while in a sitting posture in order to help widen his/her operational view, resulting in the development of musculoskeletal pains in the back. The consequence of these challenges appeared to have influenced the operators' notion that the programmed boomers do not always responds rightly to the programmed commands, and as such requires an operator to rely on his experience in guiding the boomers to operate optimally, as highlighted in the following comment by a miner.

The act of shifting our upper bodies to widen our view of operation, while in sitting posture has been giving us musculoskeletal pains in the back. We always have to engage in such act of shifting the body, because the technology does not always get it right, and as such the operator, from his acquired experience, mostly have to find ways to guide the technology for optimum performance.

The operators also wished for improvement in the work station design in the tractor cabins. They viewed as comforting the listening of music despite the challenge of having to concentrate on their objective tasks and also trying to listen to information transmission from their communication gadgets. In this regard, there was a sense of these operators socializing with the technologies objects with which they work. An example in this case is the use of pen drives loaded with music and connected to the music player fitted in the cabins of the tractors. Operators listened to the music in their tractor cabins while they work.

There are consequences when you work down here, and you need not think about it so much. So I listen to music inside the cabin all the time. I always carry along with me a pen drive loaded with music which I play when working. It is nice to work and listen to your favourite songs.

Though such brain-activity was perceived by the interviewees as overloaded cognitive activity, they were able to accommodate it. In this regard, they socialize with a technology-oriented object.

It was observed from the activity of the technology operators that, despite the incorporation of human factor elements in the design of the technology in its design, the operators were confronted with elements of cognitive complexities in their task undertaking through the interface of the technology and the task design. These was largely accredited by the operators to the reality that the designers of the machines used in the drilling task were not interested in seeking the opinions of the operators (who practically use the machines) on, not only the functionalities of the machines, but also their human-related friendliness, in terms of reducing the level of cognitive complexity expectation of operators' in the drilling task undertaking. These observations are highlighted in the following comment by a miner.

Those who design the machines normally come to our workplace when the firm wants to buy a new machine. They normally do not ask us about the machines, because we all do things in different ways, and there is no way to satisfy everybody…It will be nice if manufacturers of the machines bring their young designers to join us at work, so that they observe how we interact with the machines and also talk to us. This will enable them come out with the best designs.

The above observation therefore, brings to the fore, the issue of human factors in both the drilling work design and the design of the technology to be used in the conduct of the work.

4.2 Emotional-Motivational Aspect of Complexity in the Drilling Activity

Analysis of the emotional-motivational aspect of complexity that reflected the energetic aspects of the bolting activity showed that the operators were not satisfied with the quality of work life component of the bolting job design. There was the feeling among the operators that though the provision of new technology has made it easier for them to work longer hours without the machines needing repairs, it also minimized the level of social interaction that the frequent breakdown of machines used to enhance among team of operators.

Though it was difficult in the olden days, one can always turn around and see a friend by nearby. This because when you are repairing your broken-down machine, your friends will be there to provide you support. In the course of the repair, you can turn round and discuss ice hockey or share a joke with your friends. Today, the work design has changed due to the influence of the new machines we use. Because these machines do no easily breakdown, an operator is always down here working alone. It always makes us feel like the new work design due to the technological change has taken something away from us.

Even though the operators tend to make themselves happy in the work environment by taking away the loneliness of working alone one kilometre below the earth through socialization with the technology they use in their task undertakings, coming to work in darkness during the winter season and going home in darkness, without seeing light for many months was viewed as quite depressing by the operators. A sense of this observation is reflected in the flowing exchanges between the researcher and an operator engaged in the drilling task.

Researcher How do you see the worklife? In the winter, you come here in darkness, go underground to work, and finally go home in darkness?
Operator It is depressing. You get more depressed when you start
Researcher Do you think there is a way to reduce the depression?
Operator Yeah! Management should provide us with a sauna

The effect of the emotional-motivational issue raised by the operator in the observation above is compounded by the added observation that the operators were faced with the issue of non-water placement compartment in their high technology machine and also quality water provision. As such, they are not able to shower when they finish work due to the poor quality of water in the underground workplace. A sense of this observation is reflected in the flowing exchanges between the researcher and an operator engaged in the drilling task.

Researcher I have looked carefully inside the machines, especially the new ones and I realized that there is no compartment for you to put your bottle of water. So everything is for the machine!
Operator Yeah, you have a point. There is no place for water in the machine. You get thirsty all the time
Researcher I think it will be nice to provide you with bottled water
Operator Yes. You see, a test of the water quality both here and up have shown that the water is not fit for humans. That is why we have the water box in the cafeteria now. So if I want to drink water, I stop work and take a car to the cafeteria and then come back. Quite a task!
Researcher So like this types of machines you are using, they can provide a small place where you can bring good water and put there!
Operator Yeah, yeah! And we have itching in the hair all the time because the water is not so good

Based on the conversation above, it is obvious that the operators are of the view that their quality of work life will be enhanced by improving the work station design in the machine cabins. They viewed that the provision of a compartment in the tractor cabins for the storage of drinking water will be welcomed idea, since this will alleviate the thirst situation that the operators most often encounter due to the heat condition underground.

5 Discussion

The results analyses have shown that in the production drilling activity, the principal actor is the individual (operator) employee who is simultaneously is confronted with the challenges of handling the consequences of both the cognitive and motivational-emotional aspects of the drilling task's complexity. For the production drilling activity, one activity type is the operator's engagement with the physical task through the manipulation of digitized computer technology to programme robotic work tasks. 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 listening to background music, information transmission from the mines control centers, and information transmission from/to colleagues approaching or leaving the individual's activity location inside the deep mine. 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. Therefore, the increase in the cognitive aspect of complexity in the operators' could be related to the need for them to alter the stereotypical actions of the designed tasks in the drilling activity. This is 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 [7]. The implication from the results analyses is that the operators' objective and subjective activities are interdependent and shaped based on the mechanisms of self-regulation. The cognitive aspects of complexity in their task were found to depend on the number of task elements and the specification of interactions of the different task elements. 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. Thus, the complexity in the drilling task undertaking by the operators was defined by the number of static and dynamic components of the task and interaction among these components [7]. In this regard, by using the argumentation of [7], the degree to which the drilling task could be viewed as unpredictable, and the uncertainty that could be associated with such unpredictability must be seen as important influencing factors entailed in drilling task complexity. Thus arguing from the perspectives of [9] the number of production rules in the drilling activity task that the operators have to learn must serve as the major criterion of determining the level of the cognitive and motivational-emotional aspects of the drilling task's complexity. This understanding will help the designers of both the "task" and the "technology for carrying out the task" to optimize operators' physical activity by successfully integrating both the cognitive and motivational-emotional aspects complexity in the drilling activity.

As it was highlighted by the result, the operators make use of generic skills, such as flexibility, technical intelligence, perceptive ability, technical sensibility, a sense of responsibility, trustworthiness, and independence [14] to enhance the cognitive aspect of the complexity associated with the interaction between object-oriented and subject-oriented activities. Going by the arguments of [7], the uncertainty in the operators' drilling activities is dependent on both the objective and subjective characteristics of the tasks. Thus, the specificity of the operator's memory workload [7], from the perspectives of the cognitive aspects of complexity entailed in the drilling activity, is an important source of task complexity that needs to be understood and factored in the drilling task design. This postulation is supported by [11] argument that a situation where an operator has to deal with more than three intermediate data on dynamic objects in memory should be subjectively evaluated as a very difficult situation which produces errors, while the duration for which information is kept in memory should also be considered a critical factor. The results analyses have shown the production drilling activity entailed several operator-oriented motives (both conscious and unconscious) with varying priorities assigned to each influencing factor. Arguing along the line of [7], the totality of these motives could be viewed as influencing the operators' emotions and motivations in their conduct of the production drilling activities. Thus designers of the task environment for the drilling activity and the technologies used by the operators in the task undertakings 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 automated and digitized work environment for the production drilling activity. The work environment must also have good work and social conditions that enhance the workers 'emotional-motivational orientation. This model of uncovering

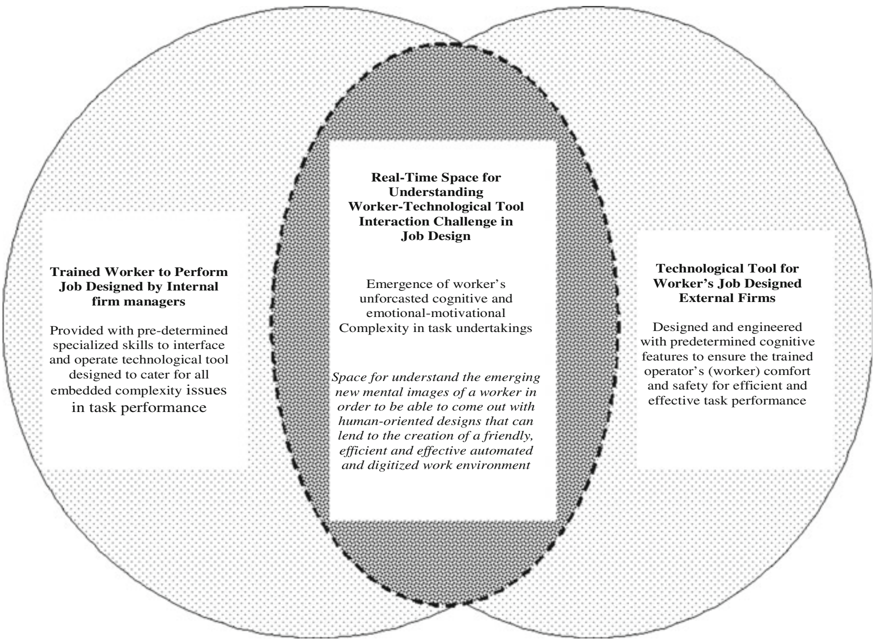


Fig. 1 Model for understanding a worker's emerging cognitive and emotional-motivational aspect of task complexity

and subsequent understanding of a worker's margining cognitive and emotional-motivational aspect of task complexity that could be used to enhance worker-technology-work environment design in production drilling activity in deep mines is visualized in Fig. 1.

6 Conclusion

The study as shown that, the concept of complexity can be applied to the motor aspect of task, or more specially, to the mental regulation of operators movements in the performance of production drilling activity in deep mines. The findings have also shown that a complex motor task in the drilling activity cannot be performed by an operator without significant cognitive (mental) effort and concentration. It is also established in the study that the relationship that exists between the different components of the drilling activity performed by the operators is critical when the complexity associated with both the cognitive aspect and the emotional-motivational aspects of the production drilling activity is being evaluated. This is because, as these complexities increase, the emotional tension and motivational force associated with the operators' work also increase. It is therefore, concluded that in the work environment and technology design processes of the production drilling activity, it is important to identify and distinguish the cognitive aspects of complexity that depend on the specificity of information processing, and the emotional-motivational aspects of complexity that reflect the energetic aspects of the production drilling activity. This will enable the designers understand the performance enhancing strategies used by workers to mediate the cognitive difficulties and the emotional-motivational challenges associated with in the production drilling activity, and which understanding can be used to reduce the task complexity by integrating them in the design of a friendly and efficient work system for the drilling job in deep mines.

References

1. Sanda, M.A., Johansson, J., Johansson, B.: Conceptualization of the attractive work environment and organizational activity for humans in future deep mines. WASET Int. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng. **5**(2), 256–262 (2011)
2. Moore, S.M., Bauer, E.R., Steiner, L.J.: Prevalence and Cost of Cumulative Injuries over Two Decades of Technological Advances: A Look at Underground Coal Mining in the U.S. http://www.cdc.gov/NIOSH/mining/pubs/pdfs/pacoc.pdf (2011)
3. Sanda, M.A., Johansson, J., Johansson, B., Abrahamsson, L.: Understanding social collaboration between actors and technology in an automated and digitized deep mining environment. Ergonomics **54**(10), 90–916 (2011)
4. Braha, D., Maimon, O.: The measurement of a design structural and functional complexity. IEEE Trans. Syst. Manage. **28**, 527–535 (1998)

5. Vicente, K.J.: Cognitive Work Analysis: Toward Safe Productive and Healthy Computer-Based Work. Lawrence Erlbaum Associates Publishers, Mahwah, NJ (1999)
6. Meister, D.: The History of the Human Factors and Ergonomics. Lawrence Erlbaum Associates Publishers, Mahwah, NJ (1999)
7. Bedny, G.Z., Karwowski, W.: A Systemic-Structural Theory of Activity: Applications to Human Performance and Work Design. Taylor and Francis, Boca Raton, FL (2007)
8. Payne, W.: Task complexity and contingent processing in decision-making: an information search and protocol analysis. Org. Behav. Hum. Perform. **16**, 366–387 (1997)
9. Kieras, D., Polson, P.G.: An approach to the formal analysis of user complexity. Int. J. Man-Mach. Stud. **22**, 365–394 (1985)
10. Gallwey, T.J., Drury, G.G.: Task complexity in visual inspection. Hum. Factor **28**, 596–606 (1986)
11. Zarakovsky, G.M., Magazannik, V.D.: Psychological criteria of complex decision-making processes. In: Lomov, B.F., Venda, V.F. (eds.) Methodology of Engineering Psychology of Work Management, pp. 63—79. Science, Moscow (1981)
12. 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)
13. Venda, V.F.: Engineering Psychology and Synthesis of Informational Sources. Manufacturing Publishers, Moscow (1975)
14. 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. Rel. **48**(5), 657–676 (2006)

Measuring Operator Emotion Objectively at a Complex Final Assembly Station

Sandra Mattsson, Dan Li, Åsa Fast-Berglund and Liang Gong

Abstract To meet future challenges of production systems, especially in high-wage countries with high technological complexity in factories, it is important to focus on human operators. The perceived operator view is an important aspect, but takes time. This paper will discuss the physiological measurements used in four commercial and semi-commercial devices in terms of *usability in industry, meaning of measurement data* and *relation to intuition and flow*. Results indicate that three physiological measurements can be used in combination to measure well-being to some extent, but that subjective data needs to be incorporated to support the individual perspective (due to that the meaning of data is subjective). By using these devices, physiological data can be measured and evaluated in real-time, which increases the possibility of studying operator emotion (or memory constructs). More studies are needed to evaluate how cognitive processes and measurement data are connected.

Keywords Operator emotion · Assembly · Measurement · Real-time data · Production complexity

S. Mattsson · D. Li (✉) · Å. Fast-Berglund · L. Gong
Department of Product and Production Development,
Chalmers University of Technology, Hörsalsvägen 7A,
41296 Gothenburg, Sweden
e-mail: dan.li@chalmers.se

S. Mattsson
e-mail: sandra.mattsson@chalmers.se

Å. Fast-Berglund
e-mail: asa.fasth@chalmers.se

L. Gong
e-mail: liang.gong@chalmers.se

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_19

1 Introduction

The customization of mass production demands flexible production systems to accommodate for the increased number of product variants in final assembly [1, 2]. This increase of product variants in final assembly increases the perceived work complexity by assembly operators [3, 4]. In such complex final assembly workplaces it has become increasingly important to study the interaction between humans and automation [5] and how the human work can be supported [6]. When working in final assembly, operators should therefore mainly use cognitive processes connected to intuition [7], skill-based and rule-based behavior [8] and flow [9]. Recent technological advancement driven by Industry 4.0 has enabled objective measurement tools to emerge [10] and therefore four commercial and semi-commercial devices studying operator emotion in real-time was tested (measuring physiological data) [11].

From a social sustainability perspective, the cognitive demands and work life satisfaction affect performance and quality of work [12, 13]. By studying emotions related to a task it is possible to study stress, frustration and boredom with the goal to reduce the number of human errors in production [14]. If relevant measurements for the well-being of operators in assembly are known, it might be easier to identify the suitability of specific changes in a production system from a social sustainability perspective.

The aim of this paper is to discuss how operator emotion can be measured in complex assembly using objective physiological measurement devices. Physiological measurements are discussed in terms in terms of: *usability in industry, meaning of measurement data, relation to intuition and flow* and *measuring well-being*.

2 Life Quality and Well-Being

Life quality assessments have been seen as combination of cognitive evaluation and affect, i.e. positive and negative feelings [15, 16] and/or other representation of a person's mood and emotions [17]. Cognitive evaluation is the person's intellectual view of their well-being and it is usually represented in terms of satisfaction measures. Schwarz et al. saw that when making judgments of how happy and satisfied people are with their lives in general they rely on their affective states (momentary). However, if they were unhappy they try to explain their state more (affective state), than those that are in pleasant affective states [16].

A recent study has shown that three psychological needs affect intrinsic motivation, self-regulation and well-being: competence, autonomy and relatedness [18]. When these three are satisfied, they increase self-motivation and mental health but when not satisfied, they instead diminish both motivation and well-being. The study also highlight other aspects connected to needs within motivation and well-being e.g. education, psychotherapy and health care. Increased well-being have recently also been connected to cognitive awareness i.e. measuring the quality of consciousness might be important [19]. Increased awareness of emotions has also been seen as important when learning [20].

Subjective well-being depend on a number of factors, e.g. health outcome, cognitive functioning and cognitive characteristics [21]. When measuring subjective well-being it is important to report emotions close in time and directly connected to actual experience, i.e. not in terms of a remembered utility ("what did you feel about what just happened?") [21]. This is due to that past experiences are often connected to systematic biases (although in general close to the described experience, but not as detailed or correct), e.g. they are not a good sum of the complete experience. Capturing real-time data is one way to minimize memory biases, i.e. people don't remember details of their experiences [17, 21].

2.1 Intuition and Flow in Complex Final Assembly

From a cognitive perspective the tasks carried out in final assembly could be described as fast and automatic, i.e. intuitive behavior [22]. Intuition is defined as recognising patterns already stored in the memory [23] or association [24]. Cognitive processes that are included in this behavior are gathering information, forming characteristics, recognizing elements in a situation, comprehending the situation and assessing the importance of objects in order to reach a goal [25, 26].

A concept that is interesting in this context is *flow* where the operator performs at an optimum [27]. Studies point towards that work-related flow (work motivation, enjoyment and absorption) increase personal and organizational resources [28], a positive mood [29, 30] and that flow is connected to situational characteristics [29].

Flow has also been related to life quality. In order for the quality of life to increase through job, that two strategies are needed [27]:

- Jobs should be redesigned so that they resemble flow activities, e.g. rules, clear goals, feedback and control.
- The person needs training to recognize the opportunities to perfect their skill and to set goals that they can reach.

3 Physiological Measurements in Devices

Four devices were chosen to measure operator emotion. The selection of devices was based on their possible application in industry applications (complex production). The aim was also to choose devices that measure different types of physiological data. The four devices are seen in Fig. 1.

Fig. 1 The four devices #1–4 (*top* to *bottom*) and visualizations of their outputs

The devices are described in terms of their physiological measurements in Table 1.

Changes in emotion, motivation, habits and attitude have been successfully investigated by studying the changes in the sympathetic branch of the Autonomic Nervous System (ANS) [31, 32]. ANS signals could be due to reactions to the situation (noise in background, people walking by) and not to the task itself

Table 1 Four devices and their physiological measurements

Device	Physiological measurements
#1. Arousal bracelet (Empatica)	BVP = blood volume pulse HRV = heart rate variability GSR = galvanic skin response (arousal) TMP = skin temperature
#2. Breathing activity (Spire)	Measures breathing activity in the body by abdominals and lungs movement
#3. Activity bracelet (Sony Smartband 2)	HRV = heart rate variability
#4. Brain activity (EPOC+)	EEG = electroencephalogram

(so there is also a difference between participants being passive and active during a measurement) [31, 32].

Physiological characteristics have been measured by studying skin conductance, blood volume pulse [33], electroencephalogram, electromyography, respiratory activity and pupil dilation [34]. Measuring ANS has been done by looking at Skin Conductance (SC) which is a measure of the Electro Dermal Activity (EDA) or Galvanic Skin Response (GSR) to measure human arousal, attention and cognitive effort [31]. As the sensors are both cheap and can be measured reliably [31], the method can easily be conducted. EDA does not however measure one exact emotion but instead serves as a general indicator for arousal, attention, habituation, preferences and cognitive effort [31, 32]. Another way of measuring ANS is to measure heart rate variability (HRV) [35, 36]. HRV is linked to high fitness levels and good health while a decreased HRV connects to stress, burnouts and fatigue. It has been suggested as a tool to monitor overtraining [37]. HRV has become a popular indication of e.g. stress since it is easy to measure. HRV can indicate this while other factors e.g. blood pressure stays within normal values [38]. HRV have been connected to anger, anxiety, disgust, embarrassment as well as some positive emotions, e.g. contentment, happiness and joy [39]. HRV is often captured today using photoplethysmography (used in device #1 and #3) by illumination of the skin, measuring changes of light absorption [40]. Another way to study ANS signals is to measure respiratory factors, i.e. breathing activities [39]. In terms of digital interaction, measures of EEG have become interesting in order to study facial expressions and vocal intonations [41]. EEG measures have also been used as a tool to differentiate positive and negative emotions [42] and to differentiate between task difficulty (attentional demands WJ Ray [43] and IQJaušovec [44]). Eye-tracking is another measure that has been used to measure differences in experience levels (in chess players [45]).

4 Evaluation of Physiological Measurements in Devices

The evaluation of physiological measurements was performed according to *usability in industry*, *meaning of measurement data* and *relation to intuition and flow* (Table 2).

HRV, BVP and GSR are easy to read and measure and are therefore usable in an industrial application. Some of the metrics are however difficult to interpret: EEG, GSR and BVP. From a cognitive process perspective, i.e. related to intuition and flow, HRV and GSR are useful in combination with the EEG metrics (studying valence).

Table 2 Evaluation of physiological measurements in the four devices

Physiological measurement	Usability in industry	Meaning of measurement data	Relation to intuition and flow
HRV, heart rate variability (device #1 and #3)	Easy to measure and read	Easy to categorize in device, but could be stronger combined with other data	Stress and fatigue
BVP, blood volume pulse (device #1)	Easy to measure and read	Requires further studies	Stress, arousal
GSR, galvanic scin response/electrodermal activity/scin conductance (device #1)	Easy to measure and read (requires real-time data)	Are individual, requires to some extent subjective interpretation	Arousal, attention and cognitive effort
EEG (device #4)	Difficult to use (requires real-time data)	Difficult to interpret, many EEG curves that relate to one another	Valence, cognitive effort, arousal etc.
Respiratory changes (device #2)	Difficult to interpret and measure	When combined with other data might be very useful	Stress, tenseness, calmness

5 Discussion

The aim of this paper was to discuss how operator emotion could be measured in a complex assembly using objective physiological measurements.

Life quality has been measured by using cognitive evaluations, i.e. intellectual view of well-being and affect [15]. From a complex final assembly perspective life quality can be increased by making jobs more similar to a flow state, e.g. by stating rules, clear goals, providing feedback and control and receiving personal training [27]. In theory, life quality could be measured by combining HRV and GSR data with EEG. Although device #4 measures EEG it is not suitable for industrial implementation and it can also be difficult to read which therefore it is recommended that EEG can be tested at training stations or that another type of measurement is needed.

Well-being can in a simplified way be seen as satisfaction ratings [17] and is connected to for instance health outcome, cognitive functioning, cognitive characteristics [21], competence, autonomy, relatedness [18] and cognitive awareness. In theory, some of these aspects could be measured in real-time by using HRV and GSR signals. However, previous studies with the GSR have shown that the amplitude of the signal depends on the individual [46, 47]. Therefore, additional subjective interpretations are needed. Also, the data is too complex to interpret since the physiological measures are connected to several activities (both cognitive and physical) [31, 32].

5.1 Reflections and Future Work

It is important to justify and further investigate what physiological measurements should be measured in final assembly. It is suggested that GSR and HRV can be measured to capture life quality to some extent. The measurements should be combined with subjective ratings or interviews to make sure that reliable interpretations are done. Except for the already suggested measurements, other interesting physiological measurements that could be added in an evaluation are respiratory activities (measured using another device or technique), eye-monitoring and/or pupil dilation.

When presenting physiological data to the operator there are however many other issues that needs to be considered. For instance, previous research has shown that the design of affective data should include familiarity, leave space for users' to interpret data and help users to make sense of the representations [48]. As an example, design was used to manage stress better by introducing adaptive music, empathetic GPS, calming temperature, corrective headlights and using a reflective dashboard [49]. There is also an issue of personal integrity that is important to consider. Future work therefore include finding devices for respiratory activities, eye-monitoring and pupil dilation as well as studying how physiological measurements should be presented in real-time to an operator working in complex final assembly.

Due to that well-being depend on a lot of factors it is not believed that objective data would replace subjective ratings entirely, but evaluations could be simplified by the introduction of objective measurements. An important aspect of supporting the operators' well-being is also to increase the operators' awareness of their well-being [19], which it is believed that the introduction of objective measurements can accomplish.

6 Conclusions

Factories in the future need to consider being attractive workplaces to develop social sustainable work environment and attract workers. If preemptive actions can be taken to help relieve operators before it is too late, sick leaves may reduce at factories. Measuring objective emotion data can be useful for promoting operators' well-being proactively in the future. Operators can be made aware of their situation and get support according to their needs and well-being.

This study indicates that although well-being include many factors, there is a strength in using objective physiological data. Due to that subjective ratings often cannot occur close in time to when the studied activity is present, objective data offers one way to accomplish that. More research is needed to interpret and find threshold values for individual data in order to state what is good or bad in the complex assembly situation.

Acknowledgments This work has been carried out within the Sustainable Production Initiative and the Production Area of Advance at Chalmers. The support is gratefully acknowledged. The work has been part of the following research projects DIG IN and DYNAMITE, which has been supported by VINNOVA.

References

1. Smith, S., Jiao, R., Chu, C.H.: Editorial: advances in mass customization. J. Intell. Manuf. **24**, 873–876 (2013)
2. Barman, S., Canizares, A.E.: A survey of mass customization in practice. Int. J. Supp. Chain Manag. **4**, 65–72 (2015)
3. Hu, S.J., Zhu, X., Wang, H., Koren, Y.: Product variety and manufacturing complexity in assembly systems and supply chains. CIRP Ann. Manuf. Technol. **57**, 45–48 (2008)
4. Mattsson, S.: What is perceived as complex in final assembly? Gothenburg: Licentiate thesis, Chalmers University of Technology (2013)
5. ElMaraghy, W., ElMaraghy, H., Tomiyama, T., Monostori, L.: Complexity in engineering design and manufacturing. CIRP Ann. Manuf. Techn. **61**, 793–814 (2012)
6. Schlick, C.M., Bützler, J.: Ergonomic design of future production. Occup. Erg. **12**, 71–72 (2015)
7. Kahneman, D.: A perspective on judgment and choice: mapping bounded rationality. Am. Psychol. **58**, 697–720 (2003)
8. Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans. Syst. Man Cyb. **13**, 257–266 (1983)
9. Csíkszentmihályi, M.: Good Business: Leadership, Flow, and the Making of Meaning. Penguin Books, New York (2003)
10. Whitmore, A.A.A., Da Xu, L.: The internet of things—a survey of topics and trends. Inform. Syst. Front. **17**, 261–274 (2015)
11. Mattsson, S., Partini, J., Fast-Berglund, Å.: Evaluating four devices that present operator emotions in real-time. In: Submitted to: 26th CIRP Design Conference. Stockholm (2016)
12. Layer, J.K., Karwowski, W., Furr, A.: The effect of cognitive demands and perceived quality of work life on human performance in manufacturing environments. Int. J. Ind. Ergon. **39**, 413–421 (2009)
13. Avey, J.B., Nimnicht, J.L., Pigeon, N.G.: Two field studies examining the association between positive psychological capital and employee performance. Leadersh. Org. Dev. J. **31**, 384–401 (2010)
14. Hudlicka, E.: To Feel or Not to Feel: The Role of Affect in Human-Computer Interaction. Int. J. Hum.-Comput. St. **59**, 1–32 (2003)
15. Andrews, F.M., Withey, S.B.: Social Indicators of Well-Being—Americans' Perceptions of Life Quality. Springer, New York (1976)
16. Schwarz, N., Clore, G.L.: Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. J. Pers. Soc. Psychol. **45**, 513–523 (1983)
17. Frey, B.S., Stutzer, A.: Happiness and Economics: How the Economy and Institutions Affect Human Well-Being. Princeton University Press, Princeton (2002)
18. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**, 68–78 (2000)
19. Brown, K.W., Ryan, R.M.: The benefits of being present: mindfulness and its role in psychological well-being. J. Pers. Soc. Psychol. **84**, 822–848 (2003)
20. Shen, L., Wang, M., Shen, R.: Affective e-Learning: using "emotional" data to improve learning in pervasive learning environment. Edu. Technol. Soc. **12**, 176–189 (2009)
21. Kahneman, D., Krueger, A.B.: Developments in the measurement of subjective well-being. J. Econ. Perspect. **20**, 3–24 (2006)

22. Mattsson, S., Fast-Berglund, Å., Stahre, J.: Managing production complexity by supporting cognitive processes in final assembly. In: 6th Swedish Production Symposium. Gothenburg (2014)
23. Chase, W.G., Simon, H.A.: Perception in chess. Cognitive Psychol. **4**, 55–81 (1973)
24. Hollnagel, E.: Cognitive ergonomics: it's all in the mind. Ergonomics **40**, 1170–1182 (1997)
25. Rasmussen, J., Pejtersen, A.M., Schmidt, K.: Taxonomy for Cognitive Work Analysis. Risø National Laboratory, Roskilde (1990)
26. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**, 32–64 (1995)
27. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper Perennial Modern Classics, New York (1990)
28. Salanova, M., Bakker, A.B., Llorens, S.: Flow at work: evidence for an upward spiral of personal and organizational resources. J. Happiness Stud. **7**, 1–22 (2006)
29. Fullagar, C.J., Kelloway, E.K.: Flow at work: an experience sampling approach. J. Occup. Organ. Psych. **82**, 595–615 (2009)
30. Eisenberger, R., Jones, J.R., Stinglhamber, F., Shanock, L., Randall, A.T.: Flow experiences at work: for high need achievers alone? J. Organ. Behav. **26**, 755–775 (2005)
31. Figner, B., Murphy, R.O.: Using skin conductance in judgement and decision making research. In: Schulte-Mecklenbeck, A.K.M., Ranyard, R. (eds.) A Handbook of Process and Tracing Methods for Decision Research, pp. 163–184. Psychology Press, New York (2011)
32. Mendes, W.B.: Assessing autonomic nervous system activity. In: Harmon-Jones, E., Beer, J.S. (eds.) Methods in Social Neuroscience, pp. 118–147. Guilford Press, New York (2009)
33. Lo, A.W., Repin, D.V.: The psychophysiology of real-time financial risk processing. J. Cogn. Neurosci. **14**, 323–339 (2002)
34. Mauri, M., Cipresso, P., Balgera, A., Villamira, M., Riva, G.: Why is facebook so successful? Psychophysiological measures describe a core flow state while using facebook. Cyberpsychol. Behav. **14**, 723–731 (2011)
35. Task Force of The European Society of Cardiology and The North American Society for Pacing and Electrophysiology: Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Eur. Heart J. **17**, 354–381 (1996)
36. Stauss, H.M.: Heart rate variability. Am. J. Physiol. Regul. Integr. Comp. Physiol. **285**, R927–R931 (2003)
37. Ithlete: Monitoring Heart Rate Variability (HRV) is so much more valuable than just Monitoring Heart Rate. http://www.myithlete.com/what-is-hrv/
38. Mega: What is HRV. http://www.megaemg.com/knowledge/heart-rate-variability-hrv/
39. Kreibig, S.D.: Autonomic nervous system activity in emotion: a review. Biol. Psychol. **84**, 394–421 (2010)
40. Jonathan, E., Leahy, M.: Investigating a smartphone imaging unit for photoplethysmography. Physiol. Meas. **31**, N79–N83 (2010)
41. Liu, Y., Sourina, O., Nguyen, M.K.: Real-time EEG-based human emotion recognition and visualization. In: International Conference on Cyberworlds (CW), pp. 262–269. Singapore (2010)
42. Nie, D., Wang, X.W., Shi, L.C., Lu, B.L.: EEG-based emotion recognition during watching movies. In: 5th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 667–670. Cancun (2011)
43. Ray, W.J., Cole, H.W.: EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. Science **228**, 750–752 (1985)
44. Jaušovec, N.: Differences in cognitive processes between gifted, intelligent, creative, and average individuals while solving complex problems: an EEG study. Intelligence **28**, 213–237 (2000)
45. Sheridan, H., Reingold, E.M.: Expert vs. novice differences in the detection of relevant information during a chess game: evidence from eye movements. Front. Psychol. **5**, 941 (2014)

46. Franzén, G., Johansson, J., Landström, A., Larsson, K., Li, D., Söderberg, C.: Takttidens, materialfasadens och instruktionernas påverkan på monteringsprocessen med avseende på kvalitet, produktivitet och montörens sinnesstämning. Bachelor thesis, Chalmers University of Technology, Gothenburg (2013)
47. Söderberg, C., Johansson, A., Mattsson, S.: Design of simple guidelines to improve assembly instructions and operator performance. In: 6th Swedish Production Symposium. Gothenburg (2014)
48. Ståhl, A., Höök, K., Kosmack-Vaara, E.: Reflecting on the design process of affective health. In: 4th World Conference on Design Research. Delft (2011)
49. Hernandez, J., McDuff, D., Benavides, X., Amores, J., Maes, P., Picard, R.: AutoEmotive: bringing empathy to the driving experience to manage stress. In: Companion Publication on Designing Interactive Systems, pp. 53–56. Vancouver (2014)

A Comparative Study on 3D/2D Visual Search Performance on Different Visual Display Terminal

Yunhong Zhang, Ruifeng Yu, Lei Feng and Xin Wu

Abstract Visual search tasks are mainly test methods of user performances for electronic visual displays, which was usually used to assess the subjective quality of the visual display terminal. This study investigated the effect of different 3D styles on static 3D/2D visual search performance. A 2 (search environment: 3D versus 2D) $\times$ 2 (display styles: 3D polarization versus 3D switch) within-subject factorial design was used in this experiment. The visual search contents included static 3D/2D visual search performance, which the search target was hexagonal pyramid and the background was pentagonal pyramid. The experiment task was to find out one hexagonal pyramid from many five pyramid in a trial. The experiment was carried out in two 47-inch screen televisions and their matching 3D glasses. As the 47-inch screen televisions, one was a typical 3D polarization television, which used LG display panel; the other was a typical 3D switch television, which used the Samsung's display panel. Sixteen subjects participated in this experiment. The search time and accuracy of each participant were recorded. The difference in search performance between 3D polarization television condition and 3D switch television condition was not significant whereas that between static 3D visual search and 2D visual search condition was significant. Post hoc comparisons found that the search time under the 2D environment was significantly longer than the search time under the 3D environment. Those results revealed that search performance was sensitive to search environment and the performance was not sensitive to 3D display styles. The obtained results could be a reference for deciding the visual search efficiency on different 3D display styles.

Keywords 3D/2D · Visual search performance · Visual display terminal

Y. Zhang (✉) · L. Feng
Human Factor and Ergonomics Laboratory,
China National Institute of Standardization, 10084 Beijing, China
e-mail: zhangyh@cnis.gov.cn

R. Yu · X. Wu
Department of Industrial Engineering, Tsinghua University,
10084 Beijing, China

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_20

1 Introduction

Visual search is very common in everyday life [1]. In visual search tasks such as Medical operation, security checking, military detection, 3D game are required to find certain subjects among background objects. In the past, the targets and background objects or the display were usually 2D, searches had to find targets in 2D searching area. With the development of display manufacturing technology, 3D display production technology is increasingly mature and cheaper, and they could be applied to various fields, which included medical operation, military detection and 3D game etc. But is the visual search performance in 3D display environment better than that of 2D display environment? There is no relevant literature description.

In addition, the mainstream 3D manufacturing technology mainly included switched mode and polarized mode, produced by Samsung Company and LG Company respectively. Moreover, which is better or worse about the two kinds of manufacturing technology for the user's visual search performance? The problem is a hot issue about which the manufacturers and consumers concerned. Researcher had did a comparison experiment on different 3D television displays produced by different 3D manufacturing technology in the dynamic visual search performance on 2D environment, but there was no significant difference of visual search performance and visual fatigue degree [2]. However, is there significant difference of visual search performance and visual fatigue between two kinds of 3D display manufacturing technology in static visual search task? The problem is also the questions concerned and focused on by the researcher.

This study intends to investigate the effect of watching visual displays produced by different 3D manufacturing technology though 2D and 3D visual search task paradigm and subjective survey. The study was expected to provide evidence to evaluate the efficiency of different types of 3D displays and different 2D or 3D visual environment. According to ISO international standard [3] about the user performance test methods for electronic visual displays, we adopted a visual search task paradigm [4] to evaluate the fatigue degree of different displays. We had tested the static visual search task performance, user experience and the visual fatigue degree of two kinds of displays, which provide the reference for the consumers, fans and gamers buying 3D displays. The specific evaluation indicators include the effectiveness, efficiency, comfortable and satisfaction etc. The visual fatigue perception scale developed by James et al. [5] was used to evaluate the eye and mental fatigue degree.

2 Method

The experiment was designed referencing to [6, 7]. The experiment task was to find out one hexagonal pyramid from many five pyramid in a trial. A within-subject factorial design was conducted to investigate the different performance of visual search on the different visual display terminals. A 2 (image types: two dimensional image vs. three dimensional image) × 2 (visual display format: 3D switch TV or 3D polarization TV). In this experiment, the visual search time and search accuracy were measured to assess and compare which condition is better.

2.1 Participants

Sixteen ordinary adults from 18 to 30 years old (8 male and 8 female, mean age = 26.56, standard deviation of age = 4.43) were recruited and paid to participate in the experiment. All of them had more than 1.5 normal or corrected-to-normal visual acuities and healthy physical conditions, without ophthalmic diseases. They did not have any history of neurological and mental diseases.

2.2 Stimulus and Apparatus

The experiment task was to find the hexagonal pyramid among the pentagonal pyramids. There was only one target in each trial and the gap may be at left upper, right upper, left under or right under. The size of the background pyramid and target pyramid were same. Targets and background objects was green(R 255, G 0, B 5), blue (R 0, G 12, B 255) and red(R 30, G 255, B 95) that were balanced in all experiment conditions. Moreover, the background was black(R 255, G 255, B 255). The resolutions of all the display stimulus pictures were 1280 × 768, which the size of the space on the screen occupied by the shape in each picture is same. The order of pictures presentations was completely random. Considering that the experiment time should not too long, the search number of 2D image condition was reduced. In order to make the targets appear evenly in the different area of the screen, the screen was divided for four regions. According to mathematical plane right angle coordinate system rules, the upper right section of the screen was defined as region 1, the upper left section of the screen was defined as region 2, the lower left section of the screen was defined as region 3, and the lower right section of the screen was defined as regional 4. If the targets were showing near the fixation point or the boundary position, the subjects will be easy to find out the target. In order to avoid the impact on the experimental results, we will set the location of the target in the middle position of each region. And the searching area can be divided into 64 × 40 boxes with the size of 20 × 20 pixels (see Fig. 2). The background objects and targets were arranged as shown in Fig. 1. And the beginning of the task, the target didn't appear in 1st–7th rows, 34th–40th rows, 1st–12th columns and 53rd–64th columns in order to avoid extremely long search time. In addition, to avoid too short search time, the target didn't appear in 15th–25th rows and the intersection area of the 11th–29th rows and the 20th–44th columns. The area that the target did not appear was shown in Fig. 2(the dark area).

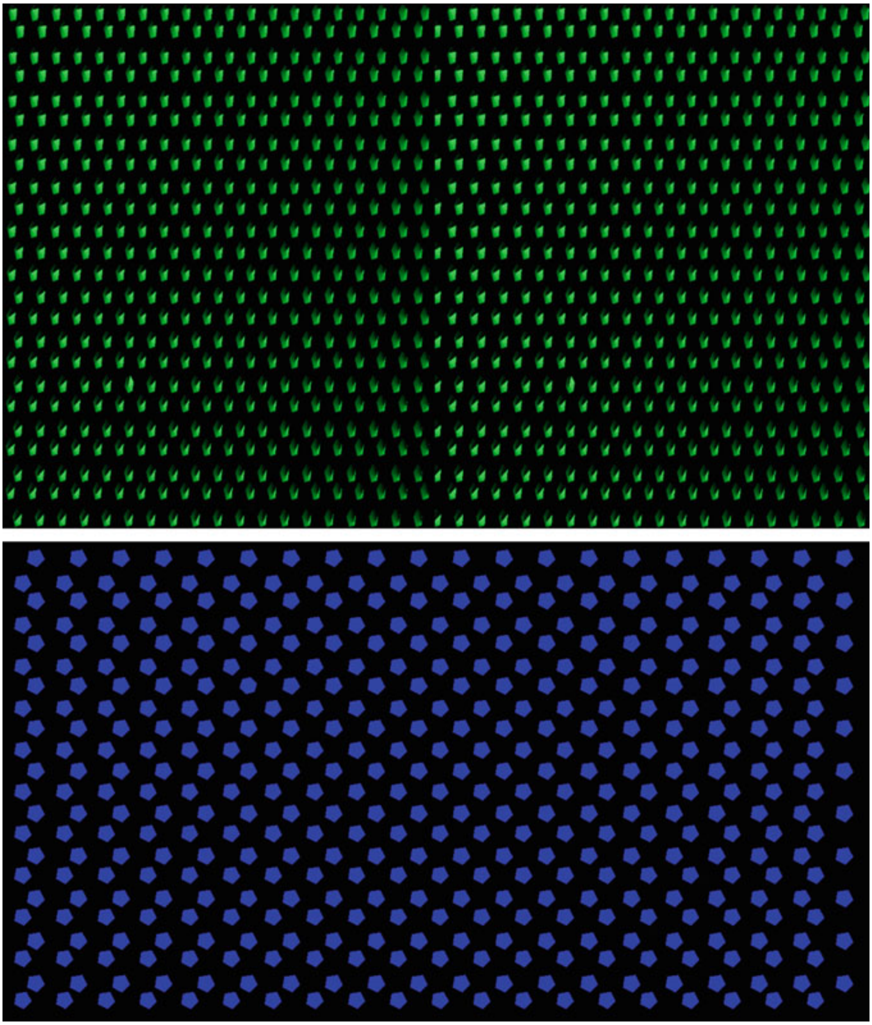


Fig. 1 Experiment pictures. The above picture is the *green* 3D picture, and the under picture is the *blue* 2D picture

By the E-prime software, the stimulus pictures were presented and the time of visual search time and the accuracy data of the every trial was recorded in the software. The experiment task was carried out in two kinds of 47-inch LED screen televisions and their matching 3D glasses, which one is 3D polarization TV and the other is 3D switch TV. The resolutions of all the display terminals were 1280 × 768 and their refresh rates were 60 Hz. The image mode of all the LED televisions sets exactly the same standard mode. Moreover, the Dell workstation host computer was connected with the LED by a data line. Participants were asked

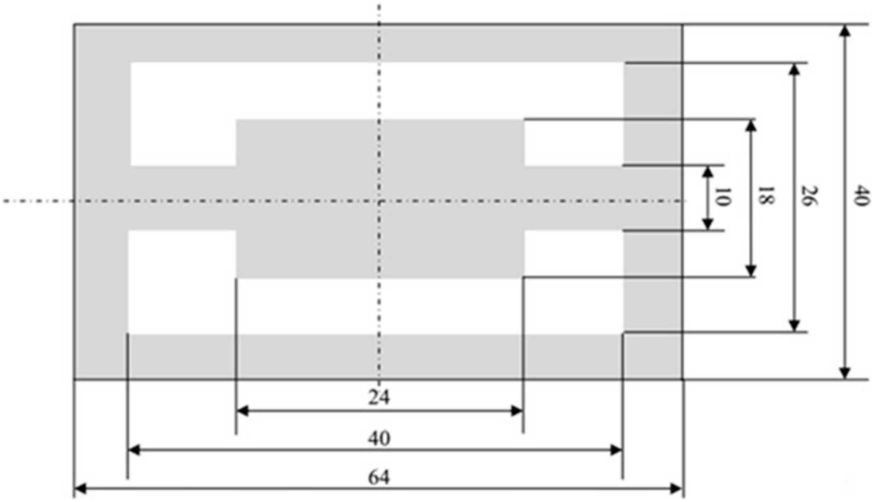


Fig. 2 The area that the target did not appear in each task

to seat on an adjustable chair before the screen at a distance of 2500 mm. The height of the chair was adjusted so that the participants' eyes were at the same height as the center of the screen.

2.3 Procedures

Experiments were conducted in a quiet laboratory environment that simulated home condition. It was installed in Human Factors and Ergonomics lab in China National Institute of Standardization. The visual search tasks were displayed on the LED TVs. After arriving at the laboratory, participants signed the informed consent and completed a general survey about their demographic information about name, sex, age, education, eye health condition, etc. Participants were asked to sit into the simulator to get ready for the test. First, the experimenter explained the procedure of the experiment and showed how to do the experiment task. Then participants began practice trials under the instructions of the experimenter to ensure participants understand the instruction and conduct the visual search task correctly. The practice trials were the same as the actual experiment trials except that there was no time pressure. If participants do not know the experimental process, then let him/her do practice again until him/her do practice proficiency. Then, they entered the formal experiment. Before each task, there was a red dot on the center of the screen. Participants were required to gaze at the dot and then clicked any key after which the task interface appeared. Once the target was found, the participants were required to click any key. Then a dialog appeared to let the participants choose the direction of the target's location (i.e. upper left, upper right, lower left or right under) by click the direction key. There were 6 groups of experiment trials in one single display to complete, and there was a five minutes break between each group. After completing the visual searching task, the participants were required to fill out the visual fatigue questionnaire. The same procedures have done on each displays. In order to avoid the impact of visual fatigue on the experimental results, the participant were asked to complete the experiment for two times, that is, one time was for completing one display test, the other time was to complete another display test, and the test order was balanced. Moreover, in each display testing, participants need to complete different color 2D and 3D image searching task and the presentation order of different color 2D and 3D image were balanced in different displays.

2.4 Data Analysis

The changes value of visual search and visual fatigue data was analyzed by IBM SPSS 20 Statistics software (IBM-SPSS Inc. Chicago, IL). The method of repeated-measure ANOVA analysis and t test were applied to the experiment data. The experiment was comparing 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 are obvious, and otherwise, they are not.

3 Results and Analysis

3.1 Visual Research Time

With regard to the search time, a repeated-measure ANOVA was applied to the search time data of the different conditions, a significant main effect of 2D/3D factors was found ($F(1,15) = 213.43$, $P < 0.001$), the main effect of displays was not significant ($F(1,15) = 0.73$, $P = 0.405$), and there was no significant interaction, $F(1,15) = 1.45$, $p = 0.248$. It indicated that under the 2D/3D search environment, search time was obviously difference, but it was no variation on various TV displays. The T test analysis showed that the average search time under 2D environment (41898 ms) was remarkably longer than the search time under 3D environment (6674 ms), $t(15) = 14.609$, $P < 0.001$ (see Fig. 2). The average search time of switch display under 3D environment was 6483 ms, and that of the polarization displays was 6866 ms, the result of T test analysis was $t(15) = 0.744$, $P = 0.468$; and under 2D environment, the former was 43384 ms,and the latter was 40314 ms, the result of T test analysis was($t(15) = 1.035$, $P = 0.317$). So it further showed that there was no significant difference whatever under 2D or 3D visual

Table 1 Comparison of averages and standard deviation about search time among different conditions

Types		Search time	Standard deviation
Switch display	3D	6483.38	449.86
	2D	43483.56	3117.32
Polarization display	3D	6866.19	524.90
	2D	40314.27	2819.68

search environment. The results indicate that there was no difference for various displays under the different search environment of 2D and 3D. Comparison of average search time under different conditions was as follows (see Table 1):

3.2 Accuracy

A repeated-measure ANOVA was applied to the search accuracy data of different conditions, a significant main effect of 2D/3D factors was found (F(1, 15) = 5.76, $P < 0.05$), and the main effect of displays was not significant (F(1, 15) = 0.628, $P = 0.441$), and there was no significant interaction, F(1, 15) = 0.73, $p = 0.407$. The results indicate that there was significant difference in search accuracy data under the 2D/3D search environment, and there was no significant difference of the search accuracy on different 3D displays. Further T test analysis showed that the average search accuracy under 2D environment (0.93) was remarkable lower than that of the 3D environment (0.96), t (15) = 2.40, $P = 0.030$ (see Table 2). Under 3D environment, there was no significant different between the average search accuracy of switch display(0.96) and that of the polarization display (0.96), t (15) = 0.094, $P = 0.923$; and there was also no environment under 2D environment between the Switch display (0.94) and the polarization display(0.92), t(15) = 1.195, $P = 0.238$. It indicates that no matter under the 2D or 3D visual searching circumstance, there is no significant difference. The results show that there was different about the search accuracy indicator between 2D and 3D search environment, but there was no difference about the search accuracy indicator between the two kinds of displays.

Table 2 Comparison of averages and standard deviation about response accuracy among different conditions

Types		Search time	Standard deviation
Switch display	3D	0.961	0.007
	2D	0.936	0.021
Polarization display	3D	0.960	0.013
	2D	0.917	0.022

3.3 Subjective Report

There were two sections of subject report data. The first part was visual fatigue, which including 9 aspects, namely eye burning, ache, strain, irritation, tearing, blur, double vision, dryness and headache, and the other part was comfort data, it contained evaluation of 4 aspects, namely task comfort, glasses comfort, difficulty of visual searching task and total satisfaction about the displays.

Visual Fatigue. The visual fatigue data was get by averaging the nine aspects of visual fatigue. A repeated-measure ANOVA was applied to the visual fatigue data of different conditions, a significant main effect of 2D/3D factors was found ($F(1, 2) = 17.997$, $P < 0.001$), and the main effect of displays was significant ($F(1, 2) = 5.066$, $P = 0.040$), and there was no significant interaction, $F(1, 2) = 0.128$, $P = 0.726$. The results showed that the degree of visual fatigue feeling under two visual search environment 2D and 3D was remarkable significant, and there was significant different subjective visual fatigue feelings on different displays. The planned comparisons revealed that the subjective visual fatigue of 2D visual search environment (34.5) was remarkable higher than that of 3D visual search environment (21.5), so it mean that the subjective visual fatigue feeling of 2D environment was far higher than the 3D visual fatigue. Furthermore, the visual search time of 2D environment was longer than that of 3D environment. When subjects finished visual search tasks on switch display, their subjective visual fatigue feeling value was 31; but when they finished the tasks on polarization display, they reported that the subjective visual fatigue feeling value was 25. Therefore, it indicated that a greater degree of visual fatigue feeling was caused by switch display. The subjective visual fatigue feeling values of different conditions were shown in the follow table: (Table 3).

Comfort. The comfort data was get by averaging the comfort values of task comfort, glasses comfort, visual comfort, visual searching task difficulty and total satisfaction about the displays (values of visual search task difficulty was reversed). A repeated-measure ANOVA was applied to the comfort data of different conditions, a significant main effect of 2D/3D factors was found ($F(1, 2) = 35.487$, $P < 0.001$), and the main effect of displays was significant ($F(1, 2) = 15.604$, $P < 0.001$), and there was no significant interaction, $F(1, 2) = 0.665$, $P = 0.428$. The results showed that under 2D and 3D visual search environments, subjects' subjective comfort feelings were different, and their subjective comfort feelings about various displays were different too. The planned comparisons revealed that the subjective comfort feeling value of 3D visual search environment (40.5) was

Table 3 Comparison of averages and standard deviation about the subjective visual fatigue feeling values among different conditions

Types		Visual fatigue	Standard deviation
Switch display	3D	25	13.40
	2D	37	17.54
Polarization display	3D	18	12.57
	2D	32	17.40

Table 4 Comparison of averages and standard deviation about the subjective comfort feeling among different conditions

Types		Comfort	Standard deviation
Switch display	3D	36	9.89
	2D	18	14.10
Polarization display	3D	45	12.92
	2D	31	16.83

significantly higher than that of 2D visual search environment (24.5). Therefore, it showed that the visual search task under 3D visual environment was more comfort than that of 2D visual environment, and visual search time of 2D visual search environment was longer than that of 3D visual search environment. On switch display, the subjective comfort feeling of visual search task was 27, and that of the visual search task on polarized display was 38. Therefore, it mean that subjects' subjective comfort feeling of polarized TV display was higher than that of the switch display. The contrast about subjective comfort feeling of different displays and 2D/3D was as shown in Table 4:

4 Conclusion

The research mainly tested subjects' visual search performance and subjective feeling of two indicators from two aspects about display types and 2D/3D visual environment. The statistical analysis results showed that there were significant differences of visual search task performance (search time and accuracy) between 2D and 3D visual environment, which the visual search performance of 3D visual environment was significantly higher than those of 2D visual environment. Therefore, it mean that the visual search tasks of 3D environment were finished quickly and accurately than those of 2D environment, and the different 3D displays did not affect the same visual search task performance. From the subjective feeling results, there were significant differences about subjective feeling (such as visual fatigue, comfort and satisfaction) of 2D and 3D visual search task. The planned comparisons showed that the subjective feeling of 3D environment was much better than that of 2D environment, and the subjective visual fatigue feeling degree of 3D environment was lighter and the evaluation of 3D environment was comforter than 2D environment. The subjective feeling of polarized 3D TV display was better than that of the switched 3D TV display. What's more, the degree of subjective visual fatigue and comfort feeling of polarized displays was better than switched 3D displays. The study provided the evidence that 3D visual search performance was better than 2D visual search performance, and meanwhile, it proved that different 3D display manufacturing technologies from different manufacturers did not affect the visual search performance, but it might affect the user's subjective visual fatigue and comfort feeling.

Acknowledgments The authors would like to gratefully acknowledge the support of the National Key Technology R&D Program of the Ministry of Science and Technology (2014BAK01B03), and China National Institute of Standardization through the "special funds for the basic R&D undertakings by welfare research institutions" (522015Y-3991).

References

1. Yu, R.F.: Yang, L.D.S: Age-related changes in visual lobe shape characteristics and their relationship to visual search performance. J Ergon. **57**(9), 1300–1314 (2014)
2. Zhang, Y., liu, N.A., Wu, X., Chang, J., Yu R.: Is dynamic visual search performance sensitivity to the visual fatigue and comfort of LED TV? A comparative experiment of eight LED TVs. In: International Conference on HCI International 2015. vol. 528, pp. 150–155. Springer, Los Angeles (2015)
3. BS EN ISO 9241-304-2008: Ergonomics of human-system interaction—Part 304: User performance test methods for electronic visual displays (2008)
4. Tsanga, S.N.H., Chana A.H.S., Yub R.F.: Effect of display polarity and luminance contrast on visual lobe shape characteristics. Ergonomics. **55**(9), 1028–1042 (2012)
5. Sheedy, J.E., Hayes, J., Engle J.: Is all asthenopia the same? Optom. Vis. Sci. **80**(11), November (2003)
6. Wiliams, L.G., Borow, M.S.: The effect of rate and direction of display movement upon visual search. Hum. Factors **5**(2), 139–146 (1963)
7. Drury, C.G., Forsman, D.R.: Measurement of speed accuracy operating characteristics. Ergonomics **39**(1), 41–45 (1996)

Aging Effects on Brain Efficiency in General Aviation Pilots

Zarrin Chua and Mickaël Causse

Abstract Understanding the effect of aging on brain efficiency and executive functions is important for high risk activities such as general aviation. In this study, ten private pilots in the age group 19–25 and ten in the 52–72 range completed the spatial working memory (SWM) and spatial planning and reasoning (One Touch Stockings, OTS) from the Cambridge Neuropsychological Test Automated Battery. The change in deoxygenated and oxygenated hemoglobin (HbO2) was measured. Younger pilots were found to be more efficient in the SWM task than the older group, with a smaller change in HbO2 and greater performance gain. However, aging has no significant effect on the OTS task efficiency, with both groups performing equally well. Analysis also suggests that there may be an effect on change in HbO2 due to flight hours.

Keywords Neuroergonomics · Working memory · Prefrontal cortex · Near-Infrared spectroscopy · Spatial reasoning

1 Introduction

The characterization of the effect of aging on brain efficiency and performance has been a constant subject of interest. Understanding this characterization is particularly important in potentially high risk environments such as general aviation [1–4]. Piloting relies on many of the executive functions, activities such as engaging working memory, route planning, and spatial reasoning are all known to be impacted by age [5, 6]. While research has suggested that there is a link between age and the likelihood of piloting error [7, 8], some studies tend to indicate that cognitive aging is more predictive of piloting performance than chronological age per se [3, 4]. Thus,

Z. Chua (✉) · M. Causse
Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse, France
e-mail: zarrin.chua@isae.fr

M. Causse
e-mail: mickael.causse@isae.fr

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_21

measuring cognitive performance in pilots at the time of their recurrent medical screening may be an excellent method to further increase flight safety. The prefrontal cortex is one of the brain structures that is more impacted by aging [9–11]. Because they are largely implemented in this brain region, executive functions are very sensitive to aging effect. Tools such as the Cambridge Neuropsychological Test Automated Battery (CANTAB) software have been developed to measure task performance when engaging executive functions. Previous research has shown activation in certain parts of the brain in conjunction with specific tasks. For example, the dorsolateral prefrontal cortex (PFC) is generally engaged during working memory [12, 13] and spatial planning and reasoning tasks [14, 15]. Development of instruments such as functional Near Infrared Spectroscopy (fNIRS) has also provided a mean to measure the brain activation in terms of changes in oxy- and deoxygenated hemoglobin. This non-invasive measurement has been gaining in popularity. However, little research has been completed on the characterization of the brain activation in the prefrontal cortex during the CANTAB tasks, and less so for pilots themselves.,

This paper presents a study that was designed to evaluate the effect of aging on the task performance and brain activation of private pilots during engagement of two executive functions in laboratory tasks. We begin by presenting the laboratory tasks and the experiment design and the statistical analyses conducted on the collected data. The second half of the paper discusses the impact of aging and recent flight experience on the performance and brain activation and concludes with a discussion of the significance of the results.

2 Experiment Design

The main control variables were the age group (young, older adults), difficulty level (4 or 6 depending on the test), and the optode location (16; Fig. 2). Changes in oxygenated (HbO2) and deoxygenated hemoglobin (HHb) and task performance were measured. Participants were seated in front of a tactile interface that displayed the CANTAB software suite and trained on the rules of the Spatial Working Memory (SWM) task prior to fNIRS installation and calibration. A ten second baseline reading was measured prior to starting the actual test itself. There were eight trials of the SWM task, with one trial each of $n = 6, 8$ boxes, and three trials each of $n = 10, 12$ boxes. Once the last run was completed, the software finished recorded task measures and the fNIRS data recording was stopped. This process was repeated for OTS, with the fNIRS headband retained between tasks. The OTS task had 24 trials with four trials each of $m = 1, 2, 3, 4, 5, 6$ maneuvers, in assorted order. It took approximately 45–60 min to complete the calibration and the two tasks. Participants performed the tasks in the exact order. The trial order for the two tasks could not be randomly varied between participants.

2.1 Tasks

The CANTAB software manual discusses the tasks in greater detail [16], but for brevity, the SWM task is an object retrieval game where the objective is to find n number of tokens without returning to a previously identified search space. As seen in Fig. 1, n number of boxes are randomly distributed on the screen, with one token hidden at a time. Users are taught that once a token has been found inside a box, that box will never hold another token for the rest of the trial. Task performance can be measured by the mean number of errors (number of times the user checks a box that has already held a token) and the strategy (number of times the user begins the search for the next token with a different box). A perfect performance would be 0 errors and a strategy of 0 (i.e. the user correctly chose the box that held each token every time).

The OTS task is a modified version of the classic Tower of Hanoi problem that solicits spatial reasoning and planning over the course of *m* maneuvers. Users must determine the minimum number of maneuvers necessary to replicate the top arrangement of balls with the bottom arrangement (Fig. 1). Users cannot advance to the next trial without having solved the current trial. The software provides instant feedback as to whether the user's response is correct or not. Task performance can be measured by the number of problems solved on the first attempt, the mean number of attempts until the correct answer is achieved, the mean latency to the first attempt, and the mean latency until the correct answer. A perfect (and unachievable within human performance bounds) performance would be 24/24 problems solved on the first attempt, an average of one attempt, and mean latencies of 0s.

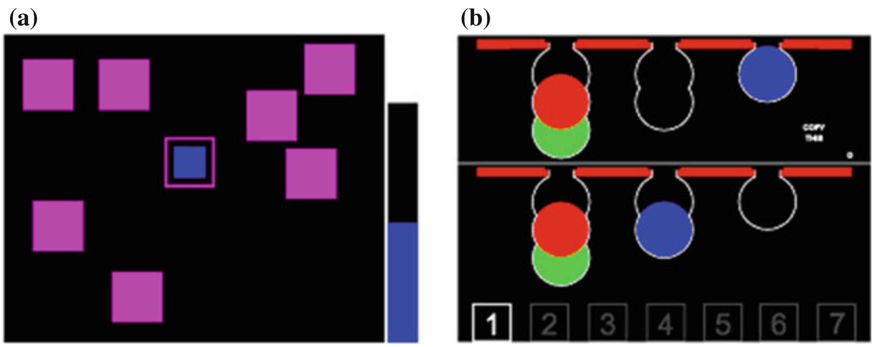


Fig. 1 CANTAB tasks. Figure 1a shows the SWM task at n = 8 boxes. The user has just found the fifth of eight tokens. Figure 1b shows the OTS task at m = 1 maneuver. The user has answered 1 for this trial. The option of 7 maneuvers is always available even though a 7-maneuver trial does not exist as to avoid users unfairly advancing to the next problem by selecting the highest number of maneuvers. Modified images from [16]. **a** Spatial Working Memory. **b** One Touch Stockings

2.2 FNIRS Equipment

The Biopac® fNIR100 system was used to measure the hemodynamics of the prefrontal cortex. This headband is equipped with 16 optodes (2.5 cm source-detector separation, Fig. 2) and records data at a frequency of 2 Hz. The data acquisition and analysis software from Biopac® was also employed: COBI Studio software (version 1.2.0.111) and fnirSoft (version 1.3.2.3). Changes in HbO2 and HHb (both in µmol/L) were measured using the modified Beer-Lambert Law with two peak wavelengths (730 nm and 850 m). Concentration measurements were band-pass filtered (pass band: 0.02–0.40 Hz) with a finite impulse response, linear phase filter with an order of 20 to further remove any slowly drifting signal components and other noise with other frequencies than the target signal [17]. Saturated channels (if any) were excluded. Several sources [18, 19] have stated that changes in HbO2 are more reliable and sensitive to changes in cerebral blood flow, thus the statistical analysis was conducted on only HbO2. Changes in HHb are shown but not statistically analyzed. All measures are changes in HbO2 concentration from a ten-second rest period baseline (measured prior to the start of each task) and are averaged over each difficulty level.

2.3 Participants

Twenty healthy adults were recruited for this experiment, all amateur pilots, with ten participants between the ages of 19–22 years (2 women, $\mu = 20.4$, $\sigma = 1.17$) and ten between 53–72 years (0 women, $\mu = 61.4$, $\sigma = 5.83$). Participants were recruited through flying clubs local to the Toulouse, France region. The average flight experience of the group was 2185.3 h ($\sigma = 5456.15$) and the average flight hours recently obtained in the two years prior to the experiment was 61.9 h ($\sigma = 112.4$).

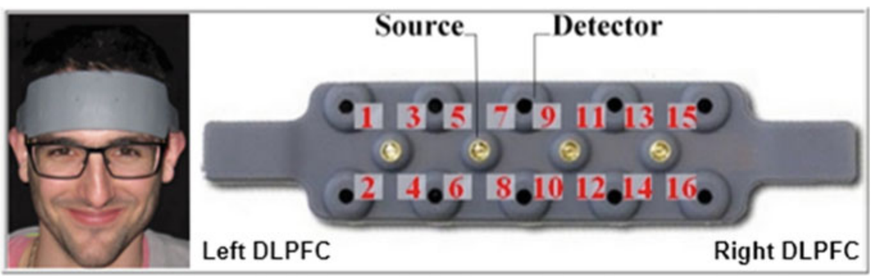


Fig. 2 FNIRS headband optode and source location

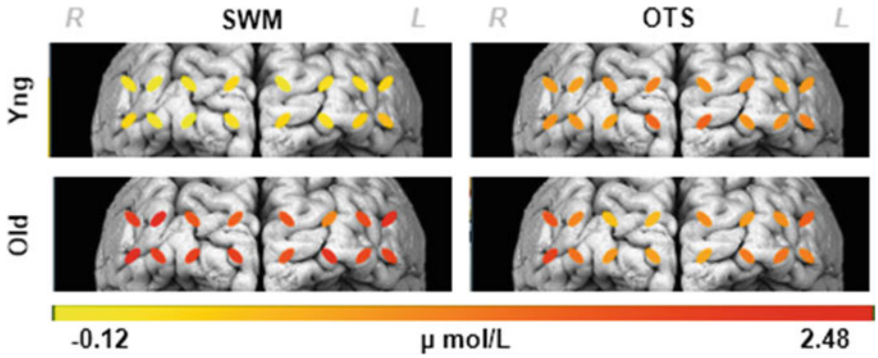


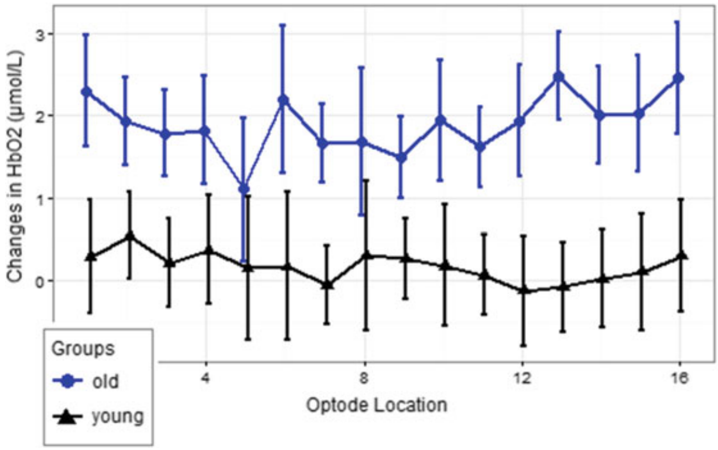
Fig. 3 Overall change in HbO2 in the PFC for both tasks and both groups

3 Results

The HbO2 data were analyzed using a three way mixed ANCOVA with group, difficulty, and optode location as main effects and recent flight hours as the covariate. A significant positive correlation exists between age and total flight experience ($r = 0.449$, $p < 0.05$) but no significant correlations exist between age and flight hours obtained in the prior two years. The CANTAB data were analyzed as either a two way mixed ANCOVA (group, difficulty, flight hours) or as a one way ANCOVA (group, flight hours). Significance for all tests was set at 0.05. All figures show vertical bars denoting the 95 % confidence interval. Figure 3 shows the distribution of average change in HbO2 between the two age groups and the two tasks.

3.1 SWM Brain Activation and Task Performance

Analysis of the changes in HbO2 during the SWM task showed that two main effects were significant: group ($F(1, 17) = 55.91$, $p < 0.000$, $\eta_p^2 = 0.767$, $\mu_{young} = 0.172$, $\mu_{old} = 1.901$) and difficulty ($F(3, 51) = 27.70$, $p > 0.000$, $\eta_p^2 = 0.620$, $\mu_1 = 0.379$, $\mu_2 = 0.491$, $\mu_3 = 1.352$, $\mu_4 = 1.925$). Optode location was not significant ($F(15, 255) = 0.50$, $p < 0.939$, $\eta_p^2 = 0.029$). Only the interaction between group × difficulty was significant ($F(3, 51) = 3.96$, $p < 0.013$, $\eta_p^2 = 0.189$; Figs. 4 and 5). The covariate flight hours was not ($F(1, 17) = 0.04$, $p < 0.844$, $\eta_p^2 = 0.002$). As seen in Fig. 6, HbO2 increases as difficulty increases. The older adults also experience greater change in HbO2 compared to young adults, with the greatest difference occurring at 8 boxes. At this difficulty level, young adults actually experience a decreased activation of HbO2 from the previous level. However, Tukey's HSD shows that within groups, the difference in HbO2 between each increase in difficulty level is not significantly different from the preceding



opt	old	yng
1	2.30	0.29
2	1.93	0.55
3	1.78	0.22
4	1.82	0.37
5	1.11	0.15
6	2.19	0.17
7	1.66	-0.06
8	1.68	0.30
9	1.49	0.27
10	1.94	0.18
11	1.62	0.07
12	1.94	-0.12
13	2.48	-0.08
14	2.00	0.02
15	2.02	0.11
16	2.45	0.30

Fig. 4 Changes in brain activation with respect to the optode location and the age group for the SWM task

level. The difference on HbO2 between groups within each difficulty level is significantly different ($p < 0.013$).

A MANOVA of the SWM task performance measures (mean between errors, strategy) showed that all main effects and their interaction were significant: group (Wilks' $\Lambda = 0.232$, $F(2, 16) = 26.461$, $p < 0.000$, $\eta_p^2 = 0.768$), difficulty (Wilks' $\Lambda = 0.150$, $F(6, 12) = 11.347$, $p < 0.002$, $\eta_p^2 = 0.850$), group × difficulty (Wilks' $\Lambda = 0.279$, $F(6, 12) = 5.180$, $p < 0.008$, $\eta_p^2 = 0.721$). The covariate flight hours was not significant (Wilks' $\Lambda = 0.922$, $F(2, 16) = 0.6774$, $p < 0.522$, $\eta_p^2 = 0.078$). Follow-up univariate ANOVAs showed that these results stood for mean between errors, but for strategy, the interaction group × difficulty was not significant. Older adults saw a decrease in performance compared to younger adults as difficulty

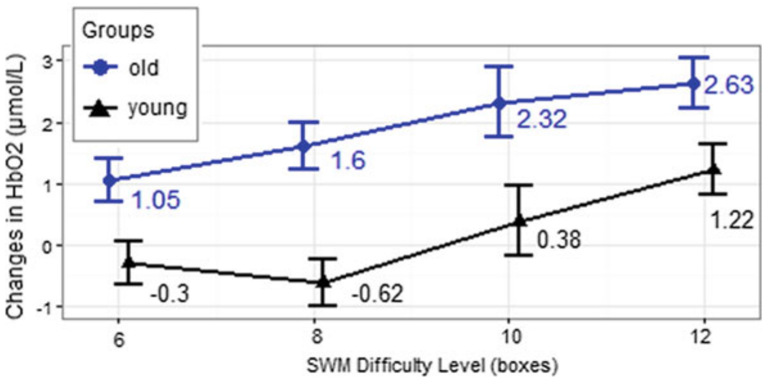


Fig. 5 Changes in brain activation with respect to difficulty and group in the SWM task

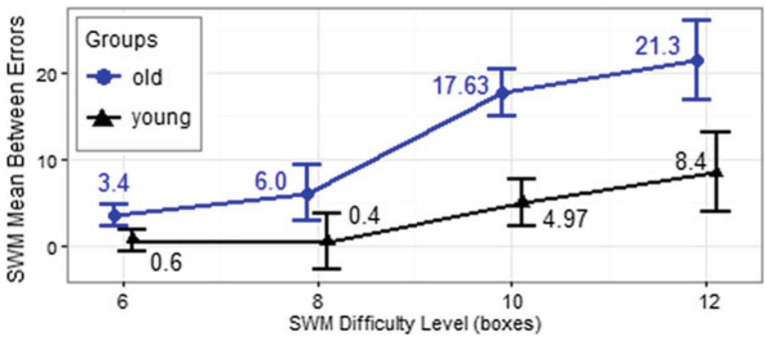
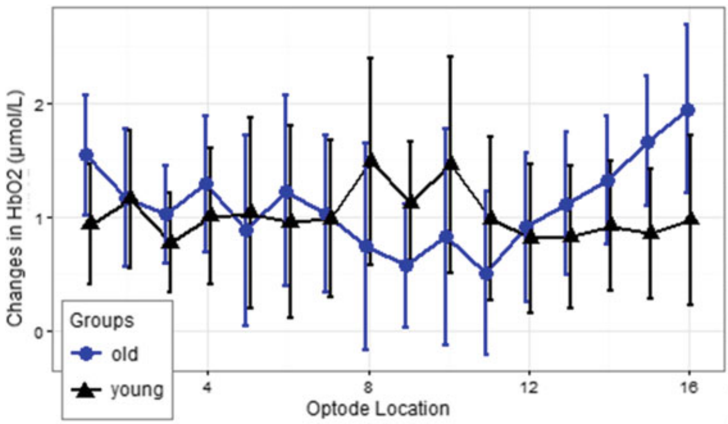


Fig. 6 Changes in task performance with respect to difficulty and group in the SWM task

increased, with honestly significant differences occurring in mean between errors between the two groups in n = 10, 12 difficulty levels ($p < 0.0001$, Fig. 6). The first three difficulty levels (n = 6, 8, 10) were not significant different for the younger adults. The changes in strategy comparing between difficulty levels were also all significant and linear.

3.2 *OTS Brain Activation and Task Performance*

Analysis of the change in HbO2 during the OTS task showed that only difficulty was significant ($F(5, 85) = 13.139$, $p_{adj} < 0.000$, $\varepsilon_{GG} = 0.398$, $\eta_p^2 = 0.436$). The other main effects and covariate were not significant: group ($F(1, 17) = 0.093$, $p < 0.877$, $\eta_p^2 = 0.001$), optode ($F(15, 255) = 1.355$, $p < 0.170$, $\eta_p^2 = 0.074$), flight hours ($F(1, 17) = 0.093$, $p < 0.765$, $\eta_p^2 = 0.005$). The only significant interaction was optode × group ($F(15, 255) = 2.638$, $p_{adj} < 0.001$, $\varepsilon_{GG} = 0.265$, $\eta_p^2 = 0.134$;



opt	old	yng
1	1.54	0.93
2	1.17	1.15
3	1.02	0.77
4	1.29	1.00
5	0.88	1.03
6	1.22	0.96
7	1.03	0.98
8	0.74	1.48
9	0.57	1.12
10	0.82	1.45
11	0.51	0.98
12	0.91	0.81
13	1.11	0.82
14	1.32	0.92
15	1.66	0.85
16	1.95	0.97

Fig. 7 Changes in brain activation with respect to optode location and group in the OTS task

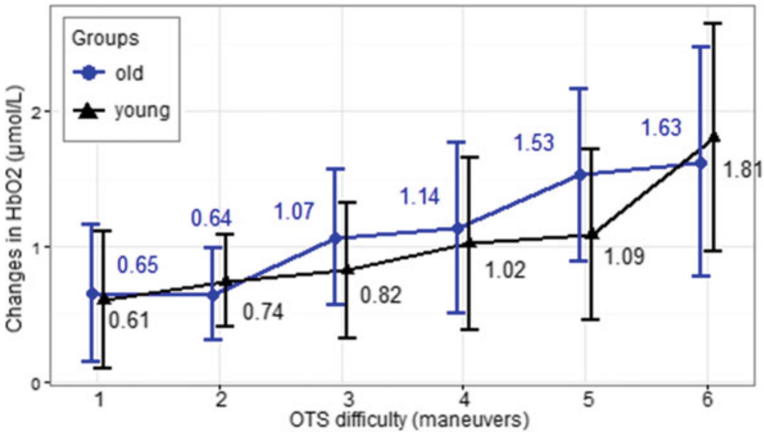


Fig. 8 Changes in brain activation with respect to difficulty and group in the OTS task

Figs. 7 and 8). Tukey's HSD showed that the change in HbO2 at the first five levels was not significantly different from the preceding level. At the highest difficulty level (m = 6), the change in HbO2 is significantly different from the other levels. As seen in Fig. 7, older and younger adults experienced activation in different locations of the PFC, with more activation in optodes 15 and 16 for the older adults and optodes 8 and 10 in the younger adults.

A MANOVA of three of the OTS task performance measures (mean number of attempts, mean latency to first attempt, mean latency to correct response) showed that only the group main effects was significant (Wilks' $\Lambda = 0.486$, $F(3, 15) = 5.279$, $p < 0.011$, $\eta_p^2 = 0.514$). Difficulty was not significant (Wilks' $\Lambda = 0.221$, $F(3, 15) = 0.221$, $p < 0.981$, $\eta_p^2 = 0.525$) nor was the covariate flight hours (Wilks' $\Lambda = 0.808$, $F(3, 15) = 1.186$, $p < 0.348$, $\eta_p^2 = 0.192$). Follow-up univariate ANOVAs showed that group was significant for the latency measures ($F(1, 17) = 15.725$ and 12.735, for first attempt and correct responses respectively; both $p < 0.002$ and with $\eta_p^2 \sim 0.4$), but not for the mean number of attempts. Difficulty was significant for all three of these measures. Flight hours was significant for the latency to first attempt ($F(1, 17) = 4.832$, $p < 0.042$, $\eta_p^2 = 0.221$). In general, as difficulty increased all participants needed more attempts before arriving at the correct answer, with the first three difficulty levels not being significantly different from each other. The mean number of attempts made by younger adults was comparable to older adults. Additionally, younger adults performed the task much faster than older adults. Pilots with more recent flight hours also appear to make the first attempt faster, but the effect is minor. Difficulty levels m = 5, 6 showed the greatest divergence in latency to both the first choice and the correct choice (Fig. 9). However, the overall performance between the two groups was equivalent. An ANCOVA on the number of problems solved did not show group to be significant ($F(1, 17) = 1.137$, $p < 0.301$, $\eta_p^2 = 0.062$) nor flight hours ($F(1, 17) = 0.044$, $p < 0.837$, $\eta_p^2 = 0.003$).

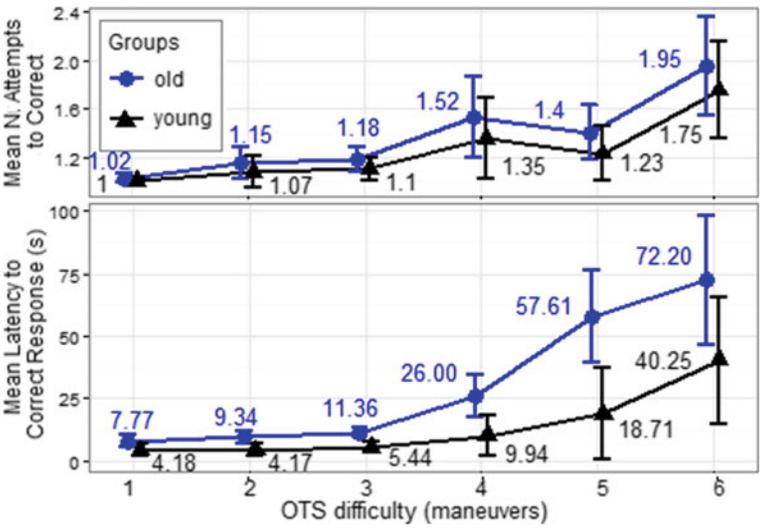


Fig. 9 Changes in task performance with respect to difficulty and group in the OTS task

3.3 Brain Efficiency and Consideration of Actual Age

Brain efficiency could be considered as a ratio between performance and resource utilization. In this case, resource utilization is analogous for the change in HbO2. The total SWM strategy and OTS mean number of attempts to correct answer scores were converted to Z-scores divided by the respective average change in HbO2. Group was a significant effect on the efficiency measure for SWM ($F(1, 17) = 9.312$, $p < 0.007$, $\eta_p^2 < 0.354$) but not significant for OTS ($F(1, 17) = 1.360$, $p < 0.260$, $\eta_p^2 < 0.074$). Younger adults were more efficient than older adults in completing the SWM task but there was no significant difference in efficiency between the two groups for the OTS task.

A multivariate linear regression was also conducted, but replacing group with age as a continuous predictor and dummy coding difficulty and optode (Eq. 1 for the regression model). In this analysis, the main effect of recent flight hours and the interaction with age were significant for the changes in HbO2 in the SWM and OTS tasks. Table 1 summarizes these findings. It should be noted that the reported F- and adjusted R^2 values are for the model described in Eq 1, but for simplicity, the coefficients for only age and age × flight predictors are listed). For both tasks, the factor flight hours is suggested to have a minor positive contribution to HbO2 (each hour of recent flight experience contributes less than 0.01 μmol/L of HbO2 in brain). Nevertheless, caution must be taken with these results, as the R-squared values for both regression models is quite moderate.

Table 1 Regression coefficients for changes in HbO2 (SWM and OTS)

Predictor	Coefficient	Std. Error	t-value	*p*-value
SWM: $F(129, 1150) = 8.538$, adj. $R^2 = 0.4319$				
age	0.0477	0.0132	3.6210	0.0003
flight	0.0070	0.0024	2.8810	0.0040
age × flight	-0.0001	0.0000	-3.0760	0.0021
OTS: $F(193, 1726) = 2.462$, adj. $R^2 = 0.1282$				
age	0.0122	0.0119	1.0280	0.3043
flight	0.0134	0.0018	7.4290	0.0000
age × flight	-0.0002	0.0000	-7.5640	0.0000

$$\begin{aligned} \text{HbO2} = \text{f}(&\text{age, recent flight hours, difficulty, optode, age} \times \text{difficulty,} \\ &\text{age} \times \text{optode, age} \times \text{flight, difficulty} \times \text{optode, age} \times \text{difficulty} \times \text{optode}) \end{aligned} \tag{1}$$

4 Discussion

This study characterized the effect of aging on the task efficiency of two executive functions, spatial working memory and spatial planning and reasoning, in two laboratory tasks. The change in HbO2 was measured as a neurological marker for brain activation while amateur pilots evaluated the SWM and OTS tasks from the CANTAB software suite. Three main results were found in this study. Globally, younger adults (less than 25 years of age) were more effective at spatial working memory than older adults. However, age does not play a significant impact on performance or brain activation with respect to spatial planning and reasoning. Lastly, recent flight experience (within two years of the task) does not have a significant impact on either performance in laboratory tasks or on brain activation.

Young adults were able to complete the SWM task better and with less brain activation than older adults. They were able to remember the position of previously found tokens and even at the highest difficulty level (n = 12), the number of errors made was comparable to the older adult performance at the second highest difficulty level (n = 8). The last two difficulty levels were the most challenging for the older adults, with nearly three times as many errors made compared to the first two levels. Interestingly, young adults saw a decrease in HbO2 in the first two difficulty levels. It is possible that the young adults were experiencing greater anticipation at the beginning of the task, particularly since the baseline was measured at the beginning of the SWM and not at the beginning of each trial. There was no significant change of HbO2 occurring at a particular optode across both groups. However, older adults did experience greater brain activation in the right and left dorsolateral PFC.

The aging effects are less evident in OTS task with the performance and brain activation of young adults statistically comparable to older adults. While younger adults had less changes in HbO2 and needed less attempts to reach the correct answer, these differences were not statistically significant. Both groups of adults saw an increase in HbO2 as difficulty increased. However, the location of this increase in HbO2 differed between groups. Younger adults saw greater activation in the medial PFC, whereas older adults experienced activation primarily in the right dorsolateral PFC. Additionally, young adults completed the OTS task significantly faster than older adults, with older adults taking nearly twice as much time to complete the task at the highest difficulty level as compared to young adults. However, rapidity in response did not translate to correct answers, as young adults answered less questions on the first try compared to adults.

In general, flight hours recently obtained played no effect on brain activation or performance for either the SWM or OTS task. However, when considered in tandem with age rather than group, this variable has a minor, but statistically significant effect on the change in HbO2. It is difficult to draw conclusions from this analysis given the small variety of ages and flight hours of the participants. Future studies should consider recruiting a greater range of both to further characterize this potential relationship.

Lastly, the neuroimaging results found with the fNIRS technique corresponds to studies conducted with similar paradigms but using other techniques. The dorsolateral PFC activation measured in OTS was also observed with functional magnetic [15], positron emission tomography [20], and transcranial magnetic stimulation [14]. Similar working memory task also showed increased dorsolateral PFC activation with fNIRS [12].

5 Conclusion

Understanding the brain activation and performance is critical for ensuring the safety of amateur pilots in a task that can have potentially dangerous consequences for errors. While this study has primarily focused on the changes on oxyhemoglobin concentration in the prefrontal cortex due to aging, such studies are useful for developing robust cockpit avionics or reducing task complexity. The results of this study have shown that greater brain activation and decreased performance in spatial working memory occurs in older pilots (57 years and up) but no significant changes occur in spatial planning and reasoning. However, aging does shift the localization of the brain activation from the medial to the right dorsolateral prefrontal cortex during spatial planning and reasoning. Recent flight experience may have a minor, but significant effect on brain activation when considered in conjunction with pilots' actual ages, but more data is needed to confirm this result.

References

1. Shao, B.S., Guindani, M., Boyd, D.D.: Fatal accident rates for instrument-rated private pilots. Aviat. Space Environ. Med. **85**, 631–637 (2014)
2. Taylor, J., Kennedy, Q., Noda, A., Yesavage, J.: Pilot age and expertise predict flight simulator performance: a 3-year longitudinal study. Neurology **68**, 648 (2007)
3. Causse, M., Dehais, F., Arexis, M., Pastor, J.: Cognitive aging and flight performances in general aviation pilots. Aging, Neuropsychol. Cogn. **18**, 544–561 (2011)
4. Causse, M., Dehais, F., Pastor, J.: Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. Int. J. Aviat. Psychol. **21**, 217–234 (2011)
5. Fabiani, M.: It was the best of times, it was the worst of times: a psycho-physiologist's view of cognitive aging. Psychophysiology **49**(3), 283–304 (2012)
6. Cabeza, R., Nyberg, L., Park, D. (eds.): Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging. Oxford University Press (2004)
7. Bazargan, M., Guzhva, V.S.: Impact of gender, age and experience of pi-lots on general aviation accidents. Accid. Anal. Prev. **43**(3), 962–970 (2011)
8. Hardy, D.J., Parasuraman, R.: Cognition and flight performance in older pilots. J. Exp. Psychol.: Appl. **3**(4), 313 (1997)
9. Turner, G.R., Spreng, R.N.: Executive functions and neurocognitive aging: dissociable patterns of brain activity. Neurobiol. Aging **33**, 821–826 (2012)
10. Yuan, P., Raz, N.: Prefrontal cortex and executive functions in healthy adults: a meta-analysis of structural neuroimaging studies. Neurosci. Biobehav. Rev. **42**, 180–192 (2014)
11. Rhodes, M.G.: Age-related differences in performance on the Wisconsin Card Sorting Test: a meta-analytic review. Psychol. Aging **19**, 482–494 (2004)
12. Fishburn, F.A., Norr, M.E., Medvedev, A.V., Vaidya, C.J.: Sensitivity of fNIRS to cognitive state and load. Front. Hum. Neurosci. **8** (2014)
13. Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B.: Optical brain monitoring for operator training and mental workload assessment. Neuroimage **59**(1), 36–47 (2012)
14. Basso, D., Lotze, M., Vitale, L., Ferreri, F., Bisiacchi, P., Belardinelli, M.O., Rossini, P.M., Birbaumer, N.: The role of prefrontal cortex in visuo-spatial planning: a repetitive TMS study. Exp. Brain Res. **171**(3), 411–415 (2006)
15. Robbins, T.W., James, M., Owen, A.M., Sahakian, B.J., McInnes, M., Rabbitt, P.: A neural systems approach to the cognitive psychology of ageing using the CANTAB battery. In: Rabbit, P. (ed.) Methodology of Frontal and Executive Function. pp. 208-231. Psychology Press Ltd, ISBN 0-203-34418-9 (2005)
16. CANTABeclipseTM Test Administration Guide. Manual Version 5.0.0. Cambridge Cognition, Ltd (2012)
17. Roche-Labarbe, N., Zaaimi, B., Berquin, P., Nehlig, A., Grebe, R., Wallois, F.: NIRS-measured oxy-and deoxyhemoglobin changes associated with EEG spike-and-wave discharges in children. Epilepsia **49**(11), 1871–1880 (2008)
18. Maidan, I., Bernad-Elazari, H., Gazit, E., Giladi, N., Hausdorff, J.M., Mirelman, A.: Changes in oxygenated hemoglobin link freezing of gait to frontal activation in patients with Parkinson disease: an fNIRS study of transient motor-cognitive failures. J. Neurol. **262**(4), 899–908 (2015)
19. Miyai, I., Tanabe, H.C., Sase, I., Eda, H., Oda, I., Konishi, I., Tsunazawa, Y., Suzuki, T., Yanagida, T., Kubota, K.: Cortical mapping of gait in humans: a near-infrared spectroscopic topography study. Neuroimage **14**(5), 1186–1192 (2001)
20. Wagner, G., Koch, K., Reichenbach, J.R., Sauer, H., Schlösser, R.G.: The special involvement of the rostrolateral prefrontal cortex in planning abilities: an event-related fMRI study with the Tower of London paradigm. Neuropsychologia **44**(12), 2337–2347 (2006)

Simulating Team Cognition in Complex Systems: Practical Considerations for Researchers

Michael McNeese, Nathan J. McNeese, Tristan Endsley, James Reep and Peter Forster

Abstract In contemporary society, teamwork is prominent as a critical component in many complex environments. Much of the success of teamwork is coupled with both team cognition and the contextual surround that a team is required to perform in. Over the last thirty years there have been various formulations of team simulations that represent team cognition and the context in various ways—some more beneficial than others. This paper examines numerous practical considerations, lessons learned, and insights developed over specific team simulations that our research group has engaged with over the years.

Keywords Team cognition · Simulation and modeling · Cognitive science

1 Introduction

Teams and teamwork have long been prevalent in many complex environments (aviation, command and control, healthcare, transportation, cyber) resulting in both successes and failures. In response to the importance of teamwork, team performance has been studied through a variety of methods and numerous technologies have been developed for augmenting collaborative activities. While teamwork has legitimately been inspired from cognitive perspectives [1, 2], teamwork is highly coupled to built and social environments (i.e. contextually-bound) [3, 4]. Yet, it has been the case that frameworks and simulations significantly under-represent real world practice, culture, and socio-political variables in the study of teams.

M. McNeese (✉) · T. Endsley · J. Reep · P. Forster
The Pennsylvania State University, Old Main, State College
PA 16801, USA
e-mail: mmcneese@ist.psv.edu

N.J. McNeese
Arizona State University, Tempe, AZ 85281, USA

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_22

1.1 Historical Background

This paper begins with the goal of retrospectively examining multiple simulations the first author has utilized for research, helped to develop, and/or validated over a thirty-year period. In this section, we will overview the team simulations we have developed and/or used, and what they have afforded researchers in terms of theory, models, and connected technologies. As one examines this historical strand of simulations it is insightful to note that the evolutionary pattern extends from primarily cognitively-bound tasks to becoming increasingly more contextually-bound [5]. This was not random, but resulted from interviews, concept mapping, and qualitative studies of real world operators and teams within given fields of practice (e.g., emergencies, crisis management, AWACS command and control airborne operations, intelligent analyst workers, cyber security) [6]. Hence development of team simulations has increasingly favored inclusion of context, perception and pickup of information directly, and even cultural specifications. Likewise, the evolutionary and in many cases extraordinary development of technology has also extended cognition from inside-the-head to be more distributed across time and context (see distributed cognition [3]; macrocognition, [7]. Indeed, current studies from our laboratory are heavily grounded in these research areas [8]. The objective of this paper is to take a broad look at these simulations and determine (1) What has been gained through research (i.e., what areas of cognitive science have been the focus for team simulations)?; (2) What are some of the salient considerations that have emerged?; and (3) What lessons learned and insights are valuable for other researchers?.

2 Emergent Complexity in Teamwork Simulation/Modeling

Practical teamwork is often tied to macrocognitive challenges [7] and often involves learning across time, getting to know teammates and how to compensate for individual differences, understanding team-to-team interdependencies, facilitating information sharing through limited visualization and communication, integrating cultural differences, and overcoming human-system integration limitations, to name a few. Indeed, the goal of understanding cognition within this broader team milieu has been paramount to enable designs that improve usefulness, adaptability, and resilience as the context unfolds.

Much of our work is historically related to the *Dynamic Decision Model* [9] in which individual decision-making is expanded into teamwork considerations. This resulted in a dynamic, intellective team decision-making simulation named TRAP [10], standing for Team Resource Allocation Problem. Within this simulation, time-contingent team interdependencies determine outcomes dependent on the joint success of human information processing strategies of each team member. TRAP was used to study cognitive variables (workload, biases, and information sharing) and technological impacts on team performance variables (e.g., use of large group displays). TRAP was typical of experimental psychology tasks but was computer-based and produced interactive opportunistic problem solving at the team level. In this sense, TRAP really focused on each team member being aware of analytical values underlying specific targets and fusing/sharing that information within the team milieu. TRAP integrated abstract thinking with emerging and uncertain information associated with "tasks", given specified team resource interdependencies. Its built-in complicatedness creates cognitive complexity. However, the simulation lacked a rich, applied, real world context and that limited its effectiveness. Also, TRAP studies were "one and done" and looked only "within a team" not "across teams". While this remains very typical of many team cognition experiments today, the approach unfortunately limits understanding of learning and dynamics within and across teams given multiple exposures to tasks and context.

The simulation CITIES [11] was developed in part to address TRAP weaknesses and afford a realistic setting for teamwork, information sharing, and team situation awareness by simulating teamwork between police and fire teams in crisis response. While different in context, it still maintained the strong underlying experimental psychology nature of TRAP. Of note is that CITIES was designed to look at team situational awareness (SA), specifically when distributed teamwork required technological means to coordinate/communicate (video teleconferencing, telephones, email, avatars were compared for performance gains). Recent work has expanded this into the NeoCITIES crisis management simulation [12], incorporating a wide array of cognitive, technological, communication, and information elements within studies, and also affording team-to-team interaction and longitudinal work with teams. Today, NeoCITIES exists as a sophisticated client-server information architecture utilizing an adaptive group interface for information display. NeoCITIES represents the logical congruence of TRAP and CITIES and is a prototypical team simulation environment. Other recent similar simulations, such as those involving UAVs, have detailed interactive team cognition in complex and dynamic domains with promising results [5, 13].

While NeoCITIES has been a very flexible simulation and has elements of cognition and context structured into task resource allocation problems, it still presents a relatively well-defined event structure that is dependent on emergent situations that are provided to participants. From that perspective—(like its predecessors) it emphasizes emergent team decision making with fairly well defined roles. Many theories and variables have been studied within this framework with much exactitude and thoroughness (see next section). However, it does not particularly capture the more ill defined situations that require a team to find, define, explore, and solve emergent problems. The Jasper task (developed as part of a broad series at Vanderbilt University) has been utilized to (1) emphasize problem solving tasks in teamwork (both distributed and collocated), (2) continue to place emphasis on collaborative technologies that facilitate learning, information sharing, and awareness, and (3) greatly increase the focus on contextualist, perceptual-based learning through the use of video-based macrocontexts [15]. Macrocontexts effectively blend cognitive and contextual elements, but provide opportunities for direct pickup of information, affording advances in problem ideation and solution (affordances, scaffolds) giving a unique excursion into very real world situations as demonstrated through Jasper's video-based story problems. Jasper was also predicated on foundational D = RxT physics problems as an underlying mathematical pretext. Therein, it was unique and enabled different kinds of collaborative interactions to test computer-supported cooperative work systems albeit in an innovative way. Jasper is by far the most contextualist-bound simulation we use. We have evolved NeoCITIES and Jasper in many ways over the last 15 years, and they have been our bread and butter go-to simulations to test theories and hypotheses in socio-cognitive science.

Understanding the challenges that the cybersecurity domain presents is a complicated endeavor. To that end, NeoCITIES has gone through several modifications, each with a specific research agenda. The first of these, CyberCITIES [14], was created to afford the capability to examine the effect that SA has on a Network Administrator's ability to respond to cyberattacks on a computer network. As with NeoCITIES, CyberCITIES continues to use the artificial concept of regret as opposed to modeling actual damage caused during an event. This concept, included in the original design of CITIES [11], is defined as "a measure of a player's opportunity lost in an event through inaction and slow or incorrect response" [15]. The use of regret as a measure allows the various frameworks to have a consistent measure across different scenarios irrespective to their contexts.

A second foray into the cybersecurity domain was with idsNETS, which focused on intrusion detection tasks of a single intrusion detection analyst. The interface of idsNETS differed significantly from previous versions of the NeoCITIES framework in that participants were responsible for not only identification of an effected network device and associated problem, but also for acknowledging the category and priority of the problem. The interface, while simplistic, provided a platform which could be extended to include more realistic tools, complex visualizations, and other such additions for further research.

Finally, teamNETS [16] served as a return to team research for our group. Using teamNETS, teams of three participants were each assigned to one specialty, intrusion detection, malicious software, or policy management. Participants, through monitoring events, take action by either filing a report to mitigate the problem or transfer the event to another player. This research was conducted to assess the impact that team knowledge structures have on team performance and collaboration using the same scoring model as the other NeoCITIES iterations [16].

While NeoCITIES and Jasper have endured, we have co-developed other simulations that vary teamwork, cognitive analytics, and contextual variation in unique ways. Briefly, two simulations have been developed to test intelligent agents-systems architectures within a complex team environment. R-CAST [17] provides intelligent analyst workers with team assistants in an emerging macrocognitive world. This world was very concrete and practical—IRAQI command and control tasks that required timely context switching and attention management. Hence, R-CAST equilibrated information fusion through a visual team interface that afforded cognitive augmentation to team workers. This augmentation was based on the principles of recognition-primed decision-making [18]. This was one of the first examples of testing a human-centered design of intelligent agents assisting teams in multiple tasks that required active switching of attention. This setting was cognitively complex but also grounded in the perturbations of contextual changes within a short timeframe. ABAIS [19] was another type of agent architecture that was specifically designed to assess beliefs and affect and then adapt a graphic user interface within an AWACS team environment. Both of these simulations were designed to test the intersections among interfaces, agents, and teamwork for outcomes that initiated resilience and flexible collaborative interactions. Both also represented human-centered approaches to complex systems where agent architectures were evaluated in highly real-world contexts. While these systems were important for innovation and forward thinking, the simulations behind them were not —for a variety of reasons. In addition to these two simulation systems, we have also worked in other areas of simulation including unique areas such as *design team negotiation*—TRACE—[20], reducing cognitive *conflict among a team of expert systems* [21], *fuzzy cognitive modeling* in emergent AWACS battle management [22]. It is interesting to note that the simulation basis for the AWACS crew utilized an extension to team-based distributed dynamic decision—DDD—simulation that was a descendent of the original dynamic decision model [9] that we began with.

As one examines these multiple simulations, the complex systems they model, and the scaled worlds that allow users to examine issues, problems, and context—grounded within a controlled lab setting—it is important to point out what is studied through the application of a given simulation.

3 Research Foundations in Cognitive Science

Much of the research utilizing the previously outlined simulation capabilities represents cross-sections of cognitive science, human-computer interaction, collaborative technology, and teamwork. This has been a very distinct niche for our research group, going all the way back to the period 1984–1988 when much of this began at the Air Force Aerospace Medical Research Laboratory, where colleagues were all working within the realm of the C^3 Operator Performance Engineering program (COPE) [23]. While this represents the embryonic beginnings, the continuance, expansion, and change through many research angles (theoretical orientation, application direction, measurement specification) and researchers have been rather vast. Yet, the cognitivist-contextualist, positivist-interpretivist, and human-centered/computation-centered tensions have remained constant, providing the requisite variety that has propelled cognitive science research into new realms.

One way to think about the research areas that have emerged is to cast them into distinct interrelated research contexts: *cognitive functions, socio-cognitive interaction, socio-technical systems* and *contextual surround.* Most of the experiments undertaken within a given study represent some kind of clustering of specific independent-control-dependent variables within each continua. Table 1 provides examples (representative but not exhaustive) of the type of theories, variables, or focus

Table 1 Continua of cognitive science research areas studied

Cognitive functions	
• Time pressure and temporality	• Hidden knowledge profiles
• Cognition, affect, and mood	• Decision making biases/heuristics
• Mental models/schema development	• Information access/retrieval
• Data analytics/integration	• Situation/contextual awareness
• Analogical problem solving	• Storytelling/narratives
• Information fusion	• Naturalistic decision making
• Perceptual anchoring	• Information weighting/value
Social-cognitive interaction	
• Team/shared situation awareness	• Team mental models
• Distributed cognition	• Information sharing
• Common operational picture	• Communication patterns
• Team-of-teams operations	• Cultural identity
Socio-technical systems	
• Electronic-mediated communications	• Intelligent agents
• Geo-spatial visualization technologies	• Large group displays
• Social networks/social media	• Decision aids
• Computer supported cooperative work	
• Adaptive computing/adaptive interfaces	
• Human computer interfaces	• Fuzzy cognitive maps
Contextual surround	
• Emerging complex situations	• Event uncertainty
• Information availability/timeframe	• Distributed work
• Levels of teaming	• Collocated work
• Scenario event interdependencies	• Perceptual pickup
• Constrained role-action structure	• Information overload

points that we have examined across many studies. While variables represent particular hypotheses, it has been useful to see how a particular simulation with unique simulation parameters and scenario-task structure answers a research question.

3.1 Team Cognition Measurement

While Table 1 examines some of the prominent areas we have explored, it is primarily geared towards research questions with independent variables. One of the most important and often daunting elements of team simulation is collecting and analyzing veridical measures that represent continua processes under investigation. This can often times be highly correlated with the phenomena occurring in the study (e.g., transactive memory evokes certain undeniable measures), but it is also bound to the scaled world simulation itself. Many of the simulations mentioned have been built to automatically collect measurements or easily derive distinct scores from functional components of the embedded task performance. This helps to provide multifold dimensions of dependent variables that can shed light on specific aspects of individual and team processes.

Within each of the above research areas, specific operationalization has occurred according to the parameters and demands within a given simulation. As one can see, many directions and research questions may be formulated within the combined continua. Hence research has emerged in unexpected directions at various times. As studies have been completed, much has been considered—much learned. Some of the best learning has transpired through problems that cause failure. Indeed, there are many lessons learned in different areas to contemplate and reflect on. The next section provides a breakdown of some of the major lessons learned that has been discovered while designing, building, and using the simulations discussed in the paper.

4 Salient Considerations, Lessons Learned, and Insights

This paper examines many of the considerations, complexities and lessons learned in team simulations that the authors have been involved in. This derives from concepts, and full-scale development of many experiments that (1) make use of information that is distributed across individuals, time, place, (2) generate scenarios that involve the use of collaborative technologies, (3) rely on designs of flexible client-server architectures and team interfaces, (4) are relevant to novice/expert and interdependent/heterogonous teams, and (5) develop unique team performance measures.

4.1 Lessons Learned and Practical Experience

Experimental Issues and Recommendations. Below, we outline some of the most prevalent challenges to experimentally studying team cognition within the laboratory.

First, requiring multiple people (often 3 or more) to show up to the lab at the same time is a challenge in of itself. Unfortunately, and often, one or more of the scheduled participants fails to show up. If this happens, then the experiment cannot be conducted. Over the years, we have developed mechanisms to anticipate these issues and ensure that the experiment can still be run. Often, we have an on-call participant list, meaning that if someone fails to show up we have a list of willing participants who we can call and ask them to show up in the lab immediately. Clearly this is not optimal, as the other participants have to wait for the new participant to show up, but it does overcome everyone wasting their time to begin with. Another mechanism that we have used in the past is to actually schedule an extra participant for the experiment. By doing this it increases the likelihood that we will have the correct number of team members available to run the experiment.

A topic that is often expounded on within the teamwork literature as a potential concern is the effect that individual differences within the team have on the generalizability of teamwork results. Teams that are typically used in the laboratory are often inexperienced (college students) and diverse in their gender, culture, etc. Because of this, many researchers have pointed out issues relating to statistical analyses being undermined due to the potential influence of inadequately weighted teams. For instance, gender is often a characteristic that is opined as being something that should be accounted for within initial team composition. Some researchers feel that teams should be equally mixed in terms of gender, not having teams that are mainly male or female. Now, the obvious and clear method for overcoming this is to counter-balance the teams before experimentation starts. The problem with counter-balancing is that this requires even more potential participants. Teamwork experiments already require an incredible amount of participants but if one were to counter-balance, even more participants are needed. Another issue related directly to the vast amount of subjects needed for completing teamwork research is that of statistical power. In many research circles, it is believed that the power must reach a certain threshold for the work to be statically relevant and valid. Yet, the problem relating to power and teamwork experiments is that in order to meet a high power, the researchers would then need to run hundreds upon hundreds of subjects through the lab.

In regards to the issues of counter-balancing and power for experimental teamwork studies, we take a logistical standpoint on them. Specifically, although we have counter-balanced in the past, we typically do not. First, as noted before, it is simply often not realistic to counter-balance teamwork characteristics or composition. Also, and more importantly, we feel that counter-balancing for specific characteristics is actually counter-intuitive to real world teamwork. Real teams are often not equally weighted for any specific characteristic, so why should we be promoting inaccurate and unrealistic teamwork in the lab setting? We feel that you should not, and we attempt to create an atmosphere of real world teamwork in our experiments. There are still more than enough statistical variables that can be pulled out and analyzed without the need for counter-balancing. As for power, we believe that power is important to consider but we do not hold firm the belief that it has to be astronomically high for the findings to be valid. We run enough experimental teams to result in what we feel is a respectable power, but we do not feel that power should be the end all be all.

Socio- Technical and Socio Political Considerations. The integration of the human-in-the-loop within the context of simulation environments is necessary and important for understanding the cognitive aspects of teamwork of sociotechnical systems. The affordances of simulation environments in complex sociotechnical environments (such as nuclear power stations, disaster response, crisis management, aviation) allows for the needs of the team to be understood and better designed for in the context. In the case of crisis management, the ingrained complexity of the environment puts significant challenges on responders cognitively and emotionally. Through the use of simulations, security and controllability of the contextual environment can be ensured.

When integrating the human into the team cognitive process, it is important to recognize the inherent biases that accompany them. Political, religious, and social belief systems develop different perspectives and motivations [25]. Differences such as cultural realities can significantly impact how participants react to a situation and share information. The case of the Asiana Flight 214 provides a poignant crisis example as it is presumed that the co-pilot refused to countermand the senior pilot in spite of the latter's obvious errors during approach.

Particularly in team settings, it is important to account for the influence of cultural biases into experimental design, as the inherent cultural perspectives that team members carry with them can dramatically influence interactions and outcomes of team performances. Unfortunately, oftentimes the cultural aspects of cognition are overlooked due to the complexity they add to experimental design. Endsley et al. [26] discuss many of the elements required to generate cultural relevance in the experimental design so that it is relevant to the participants (in this case, university students in the Unites States and in the United Kingdom), and yet also maintains fidelity to both the experiment through cross site parallelism and to the cultural context in which the simulation is being used. In the case of Endsley et al. [24], cultural relevance was obtained by adjusting the events within the scenario so that they were in British English as opposed to American English and that details were rendered salient to British participants, such as the renaming of a role "Hazardous Materials" to "Environmental Services/Army" which was culturally relevant to US participants, but not to British ones.

Performance Measure Considerations.

Quantitative. The determination of how well a team can perform is usually based on multiple dimensions in addressing differing components of what that team is trying to accomplish. Therein for many of the simulations addressed in this paper, quantitative performance generally consists of speed versus accuracy tradeoffs—as influenced by: (a) available time allotments to complete a task, (b) level of workload a task may demand, the degree of uncertainty-certainty associated with different tasks when they appear. There can be built-in variances associated with each of these elements which makes a team's performance easier or much more difficult, and the simulation usually has an architecture designed to flexibly manipulate these parameters according to objectives present in the study's hypotheses. To complicate matters, teamwork consists of individual work along with teamwork wherein individual work leverages different aspects of the above parameters. Yet, this must be successfully orchestrated in the midst of what other team members are doing simultaneously as well as with changes in contextual surround. Lessons learned suggest that dependent measures should be multi-composite wherein these different parameters can be utilized in a given formula (e.g., additive composites representative of a specific layer) that calculates an overall team performance score based on specified variables. The NeoCITIES task is an example of this, as the original composite score came from [11] and utilized a calculation that determined how well resources were applied to a given situation and whether they were able to extinguish the crisis at hand and then subjected this to a summary function. It has been our history to collect multiple dependent variables in addition to a summary team performance score to have a variety of lens to view what is going on.

Qualitative. Team based simulations provide many important things to practitioners in a variety of environments, providing significant insight into human behavior and cognition. Not least of these is that they provide a distinct set of shared experiences through which team processes can be explored in rich detail. The integration of qualitative measures into simulations such as these can provide a greater level of detail and insight into team cognitive behaviors and processes (and thus a fuller picture) beyond the quantitative measures alone, and should not be overlooked. Qualitative performances triangulated with other performance measures allow for "a deeper understanding of phenomena within its context, which may be used to inform other contexts" [25]. For example, team communication delivers and effective way to examine team cognition processes in context [26]. Through deliberative assessment and evaluation of content and flow of team communication data, an understanding of team processes and behaviors, successes and failures can emerge. Additionally, aspects of the environment, such as cultural, emotional, social and organizational perspectives can be revealed through qualitative assessment such that a richer integration of factors on teams can occur.

Scenario Development Considerations. When designing for team based simulations, it may be asked: what are the elements that make up a scenario and lend to its utility for both studying aspects of team cognition, and socio technical systems, and for training of users in an environment unencumbered by dangers and risks inherent in the real world? Effective scenario development requires identifying the problem, defining the scope of the scenario in terms of time frame and activities, recognizing the major stakeholders for which the problem is relevant, and mapping the contextual trends and drivers that establish the decisional environment. The creation of scenarios for a particular context involves the engagement of multiple research methods such as knowledge elicitation and interviews with subject matter experts, content analysis of policies, processes and past events (such as the terrorist events of the Boston bombing, natural disasters such as Hurricane Sandy, and routine emergency response events such as fires or chemical spills) which ground the events of the simulation in real world. Using ground truth as an initial context for scenario development improves relevancy and can add to the fidelity. However, the level of fidelity in a scenario depends upon the resources and technological applications that are available. The NeoCITIES simulation presents a sound medium fidelity in that it provides a level of realism such as exhaustion of resources but does not effectively account for human fatigue, which might be considered a characteristic to be included in a high fidelity operation.

Technical Considerations of Scenario Development. Once all of the elements for the scenarios have been defined, the technical modifications to the NeoCITIES framework can begin. This process, typically undertaken by someone that has a familiarity with technical programming, involves modifying the underlying source code for the application. Each of the elements of the scenario events (e.g. event title, event description, correct resource response, dispatch time, etc.) must be entered into the software. Additionally, if the research being performed requires a survey or technical briefing to measure hidden knowledge profiles, these elements must also be included. Upon completion of entering all of the required scenario elements, the software must be compiled and if necessary moved to the appropriate server location.

While the above details the process for scenario creation, there are also other considerations that must be acknowledged. First, depending on the context and measures being assessed, the interface itself may need to be modified. For example, when NeoCITIES was modified for the cybersecurity domain, the interface underwent significant changes related to this new context. Consequently, this process is potentially much more technically involved requiring someone with significant programming experience. Thankfully, the Adobe Flex environment provides a relatively easy mechanism for making interface modifications using drag-and-drop functionality if desired. As with any changes to the NeoCITIES framework, upon completion the software must be recompiled and deployed as necessary.

Cognitive Augmentation. Over the years, we have developed multiple decision aid tools and systems to help enhance both teamwork and more specifically develop team cognition. In sum, creating these cognitively augmented systems is incredibly difficult due to the issues inherently associated with teamwork. Take for instance, the development of a cognitive aid for an individual and consider how challenging the system development for that one product is. The aid must consider the individual's needs, motivations, relevant knowledge and information, trust levels, and awareness levels. This is a lot to account for. Now consider that same individual cognitive aid being then developed at a team level. This new "team level" cognitive aid must account for all of the previously outlined individual attributes, but now also must account for team level needs, motivations, trust levels, awareness levels, and communication and coordination mechanisms.

Developing team based cognitive aids must allow for adequate levels of both individual and team based work. Just because the focus is now on the team level, does not mean that individual work stops. Individual level work continues within the great collective teamwork. If the cognitive aid is focused on aiding the individual or the team too much, it can negatively affect awareness levels of both the individual and team. Too much focus on the team, such as having the aid constantly inform individual team members of team related information and knowledge can distract from individual work and awareness, which is necessary for adequate team level awareness. Likewise, too much focus on the individual's work and awareness will distract from that individual team member's team level awareness. The cognitive aid must understand the correct amount of both individual and team level knowledge and awareness it seeks to create. In order to create this correct balance, we recommend utilizing a human factors approach to first understanding the context, the team's goals, and the team itself. By doing this, a cognitive aid can then be developed that will hopefully assist the team to adequately complete their task. Understanding the context, the team, and the task, are all of paramount importance when developing team level cognitive augmentation.

5 Concluding Remarks

In this paper, we have presented a holistic representation on lessons learned over many years of studying team cognition in simulation. As we continue to use simulation to study team cognition, we must to continue to increase real world fidelity that is representative of real world tasks and situations.

References

1. DeChurch, L.A., Mesmer-Magnus, J.R.: The cognitive underpinnings of effective teamwork a metaanalysis. J. App. Psych. **95**, 32–53 (2010)
2. Hinsz, V.B., Tindale, R.S., Vollrath, D.A.: The emerging conceptualization of groups as information processors. Psych. Bulletin **121**, 43–64 (1997)
3. Holland, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput.-Hum. Interac. **7**(2), 174–196 (2000)
4. Salomon, G. (ed.): Distributed Cognitions: Psychological and Educational Considerations. Cambridge, United Kingdom: Cambridge University Press (1993)
5. Cooke, N.J., Gorman, J.C., Myers, C.W., Duran, J.L.: Interactive team cognition. Cogn. Sci. **37**(2), 255–285 (2013)
6. McNeese, M.D., Ayoub, P.J.: Concept mapping in the design and analysis of cognitive systems: a historical review. In: Moon, B.M., Hoffman, R.R., Novak, J.D., Cañas, A.J. (eds.): Applied Concept Mapping: Capturing, Analyzing and Organizing Knowledge, pp. 47–66 (2011)
7. Klein, G., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., Hollnagel, E.: Macrocognition. In: IEEE Intelligent Systems, pp. 81–85 (May/June, 2003)
8. McNeese, M.D., Mancuso, V., McNeese, N.J., Endsley, T., Forster, P.: An integrative simulation to study team cognition in emergency crisis management. In: Proceedings of the 58th Annual Meeting of the Human Factors and Ergonomics Society, Vol. 58(1), 285–289 (2014)
9. Pattipati, K.R., Kleinman, D.L., Ephrath, A.R.: A dynamic decision model of human task selection performance. IEEE Trans. Syst., Man, & Cybern. **13**, 145–166 (1983)
10. Brown, C.E., Leupp, D.G.: Team Performance with Large and Small Screen Displays (Report No. AFAMRI. TR-85.0:13). Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratory. (DTIC No. 158761) (1983)
11. Wellens, A.R., Ergener, D.: The C.I.T.I.E.S Game: a computer-based situation assessment task for studying distributed decision making. Simul. Games **19**(3), 304–327 (1988)
12. Hellar, D.B., McNeese, M.: NeoCITIES: A simulated command and control task environment for experimental research. In: Proceedings of the 54th Meeting of the Human Factors and Ergonomics Society, pp. 1027–1031. San Francisco, CA (2010)
13. Cooke, N.J., Gorman, J.C., Duran, J.L., Taylor, A.R.: Team cognition in experienced command-and-control teams. J. Exp. Psych.: Appl. **13**(3), 146 (2007)
14. Reifers, A.: Network access control list situation awareness. Doctoral Dissertation, The Pennsylvania State University, University Park, PA. (2010)
15. Heller, B.: An investigation of data overload in team-based distributed cognition systems. Doctoral Dissertation, The Pennsylvania State University, University Park, PA (2009)
16. Mancuso, V.F., McNeese, M.: TeamNETS: scaled world simulation for distributed cyber teams. In: HCI International 2013-Posters' Extended Abstracts, pp. 509–513. Springer, Berlin, Heidelberg (2013)

17. Fan, X., McNeese, M., Sun, B., Hanratty, T., Allender, L., Yen, J.: Human-agent collaboration for time stressed multi-context decision making. IEEE Trans. Syst. Man, and Cybern. (A), **90**, 1–14 (2009)
18. Klein, G.A.: Sources of power: how people make decisions. MIT Press, Cambridge, MA (1998)
19. Hudlicka, E., McNeese, M.D.: User affective & belief states: assessment and user interface adaptation. j. user modeling & user adapted. Interaction **12**, 1–47 (1998)
20. McNeese, M.D., Zaff, B.S., Brown, C.E, Citera, M., Selvaraj, J.A.: Understanding the context of multidisciplinary design: a case for establishing ecological validity in the study of design problem solving. In: Proceedings of the 37th Annual Meeting of the Human Factors Society, vol. 2, pp. 1082–1086. Santa Monica, CA. (1993)
21. Fraser, N.M., Hipel, K.W., Kilgore, D.M., McNeese, M.D., Snyder, D.E.: An architecture for integrating expert systems. Decis. Support Syst. **5**, 263–276 (1989)
22. Perusich, K.A., McNeese, M.D.: Using fuzzy cognitive maps for knowledge management in a conflict environment. IEEE Syst. Man Cybern. **36**(6), 810–821 (2006)
23. Snyder, D.E., Wellens, A.R., Brown, C.E., McNeese, M.D.: Three Paradigms for the Study of Multi-person and Human-machine Interaction. In: Proceedings of the IEEE Systems, Man, & Cybernetics Conference, pp. 480–481. Boston, MA. (1989)
24. Endsley, T.C., Reep, J., McNeese, M.D., Forster, P.K.: Conducting cross national research: lessons learned for the human factors practitioner. In: Proceedings of the Human Factors Ergonomics Society, vol. 59(1), pp. 1147–1151. Los Angeles, CA. (2015)
25. Fritze, P.: A three-phase model for mixed-methods development research in information systems. In: QualIT2004: International Conference on Qualitative Research in IT & IT in Qualitative Research. (2004)
26. Cooke, N.J., Duchon, A., Gorman, J.C., Keyton, J., Miller, A.: Preface to the special section on methods for the analysis of communication. Hum. Fact.: J. Hum. Fact. Ergono. Soc. **54**(4), 485–488 (2012)

Fatigue and Theories of Resources: The Energetic Metaphor of the Mind Functioning Might Be Reconsidered

José J. Cañas

Abstract Industrial accidents are explained often by saying that people are fatigued. Traditionally, Theories of resources have been used to explain this relation between fatigue and accidents. Those theories predict that when there is a lack on mental resources people are more prone to err. They also predict that mental fatigue depends on how long a person has been performing a task. Therefore, they predict that more time performing a task would let to more accidents. However, data from industry contradicts this hypothesis. When people are supposed to be more fatigued (i.e. at the end of the week) they have fewer accidents. This paper present some data from one ongoing research project aim to explain these results about industry accidents. Our results suggest that we should review the traditional theories of resources and propose new one that include some compensatory mechanism that supply extra resources when is needed (Hockey in Biol Psychol 45(1), 73–93, [1]).

Keywords Fatigue · Theories of resources energetic metaphor

1 Introduction

In psychology and other sciences, it is assumed that human behaviour and mental activity need "energy". Just as a car needs to consume the energy that comes from the combustion of certain gases to move, a human being needs the energy obtained from food to carry out their motor and mental activities. In a sense, we can say that in the life sciences, such as psychology, researchers have followed a "mechanistic" paradigm in which the functioning of the human machinery depends on the energy that is supplied to it. Traditionally, the energy that humans consume has been called "resources". Thus, we speak of "demanded resources", "resources available" or "resources provided by the person" to perform the task, which is often referred to as "effort" [1, 2].

J.J. Cañas (✉)
Faculty of Psychology, University of Granada, Campus de Cartuja, Granada 18071, Spain
e-mail: delagado@ugr.es

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_23

An important feature of this mechanistic paradigm is that it assumes that the resources are limited and that they are depleted when they are used. As a consequence of that, a phenomenon known as fatigue appears [2, 3]. Fatigue reduces mental alertness [4], after a prolonged "periods of demanding cognitive activity, requiring sustained mental efficiency" [5]. Therefore, it is assumed that fatigue affects performance on tasks: the fewer resources available to the task the worse performance will be. For that reason, fatigue has been considered responsible for many industrial accidents [6].

An intuitive application of the theory of resources would tell us that the main factor influencing fatigue has to be the time that the person has been performing the task. It is intuitive to think that the resources are depleted over time performing a task. The longer time performing a task, the fewer resources remain available and, therefore, the greater is the fatigue. For this reason, one of the variables that have been studied for some time as determinants of fatigue has been the "Time on Task" [7]. Thus, for example, it is intuitive to think that at the end of the working day it is more likely to have accidents due to fatigue caused by the exhaustion of resources.

However, there are some data on the time when accidents in the industry happen that do not show the effect of TOT. For example, data about accidents in Spain published by the Spanish "National Institute for Occupational Safety and Health" [8] show that there are more accidents early in the morning, on Monday, and in the first hours of the working shift (see Fig. 1). What these data from the field show is that accidents occur more frequently when resources are supposed to be intact. It seems as if performance is better and accidents less frequent when the person is supposed to be fatigued.

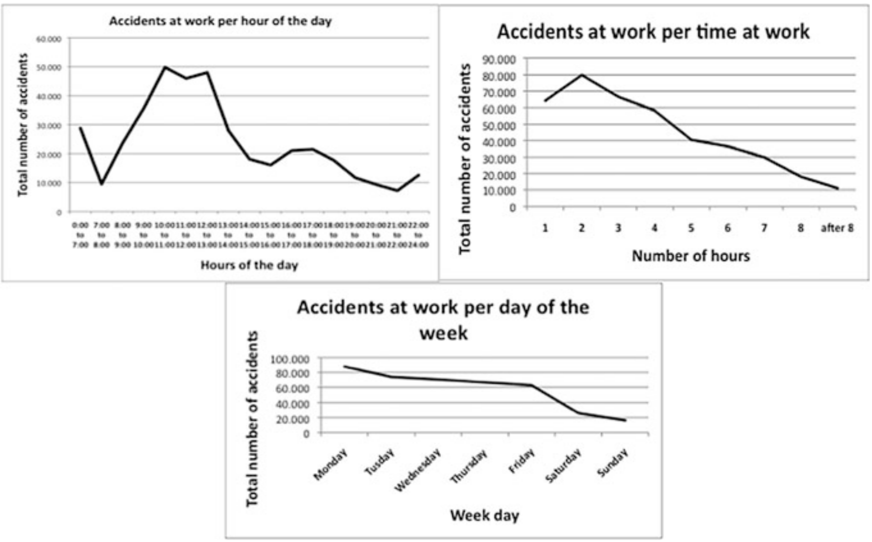


Fig. 1 Data from accidents in Spain published by the Spanish INSHT [8]

Along with these data, many researchers have also found that in situations where you should observe a decrease in performance because there are reasons to believe that the person has spent many resources after a long time performing a task, the observed results are different than expected: people are able to maintain a good level of performance, even if they declare they are fatigued (for a review see [9]). Therefore, it looks as if we need to reconsider the traditional resources theories explanation of fatigue as cause of accidents.

In this paper we describe some results of a research project that is being developed at the University of Granada which aims to prove that the concept of fatigue must be redefined from new theories of resources if we want to continue using it as a "cause of accident". The first step in this research project has been to define fatigue from the theories of resources by showing that the variable TOT (Time on task) has a real effect on any of the variables that have been proposed to measure fatigue.

The effects of TOT can be predicted from the classical theories of resources that implicitly include the decrement of resources with time in the definition of fatigue somehow. In all traditional resources theories it assumes that the decrease in performance due to time running a task is parallel to a depletion of physiological resources. This decrement on resources is reflected in physiological variables related to activation.

Resources theories have a strong support on what we know about the physiology of activation. For example, researchers have shown that there is a high correlation between changes in the amplitude of pupil diameter and the amount of resources used to perform a task [10–15]. In a recent review, [16] suggest that pupil dilation depends on the activity of one noradrenergetic system known as Locus coeruleus (LC) that acts as an inhibitory mechanism in the oculomotor parasympathetic system. The LC acts in one of two modes, in a tonic mode or in a phasic mode [17]. In the phasic mode of LC activity serves to process the task relevant stimuli. When that activity is higher so is the pupil dilation. Thus, as the stimuli complexity is greater and it requires further and deeper processing of information, the greater the dilation of the pupil is. In the tonic mode LC activity serves to explore the environment and to detect novel stimuli. This activity is carried out by increasing the diameter of the pupil. For this reason the tonic mode of LC activity has been related to the difficulty of the task, mental effort and the state of arousal and alertness. Thus, increased mental effort and increased arousal corresponds to a greater diameter of the pupil. Therefore, in our ongoing research project we are measuring pupil dilatation as an index of activation.

The participants in our experiments perform a simulated air-traffic control task. Air-traffic is a complex task were the effects of fatigue has been shown many times [18–20]. All participants are trained on the task to a point of almost no errors. The reason for that is that we want to observe how good performance is maintained after variables that are supposed to affect fatigue are manipulated. Thus, for example, we want to know how task complexity affect performance after perform the task for a long time.

According to new theories of resources, the variable that could wipe out the effect of TOT would be one that "activates" the person by taking her to use more resources to keep performance in a good level. This variable may be the complexity of the task.

Thus, for example, we can assume that the depletion of resources will be greater when the task is more complex. In an intuitive way of thinking, we can assume that a more complex task will need more resources and, even if the time passes, it would look as if the fatigue disappears. Therefore, as the new Resources Theories would predict TOT and task complexity would interact by reducing the effect of TOT when the task is more complex.

In this paper, I described the general procedure used in our ongoing research project and some data from one experiment designed to show how performance can be maintained almost perfect when the task is complex and it is performed for some time [21].

2 General Methodology of the Experiments

Psychology students from the University of Granada participated in our experiments. None of them have previous experience in air-traffic control. Their motivation for participation was to get extra credits for one of their courses.

The program used to launch the activity of air-traffic control, both in the training and in the experiment phases, is the ATC-lab Advanced software that simulate the air traffic control task and is publicly available from their authors [22] (see Fig. 2). It is equipped with a certain level of realism and it allows experimental control of many aspects of traffic scenarios. Realistic simulations with ATC-labAdvanced, means that the aspects represented on the screen appear in the real environment (including the dynamic motion of the aircraft) as the way to interact with one's environment simulated by the controller looks like a real operational scenario. Moreover, the experimental control can generate situations of air traffic needed for the purpose of research, that is, create standard settings, or isolate tasks or specific sub-tasks and test them. The xml code generated by the Software can be modified according to researcher needs, and is also possible to create the airspace with the features and the amount of traffic that researcher wants.

To create a traffic scenario we start by defining the "fixed" relevant characteristics of the simulation environment, that is, the size of the sectors, routes, and "waypoints". Subsequently, the aircrafts that is presented by scheduling them with their initial characteristics (altitude, assigned altitude, speed, time of appearance on stage and planned route). Once all the structural and dynamic characteristics of traffic scenario have been defined, we can obtain a log with the record of the performance for each participant. Participants have several tools available to perform the control task. Those tools could be use to, for example, change flight levels, speeds and headings of the aircrafts.

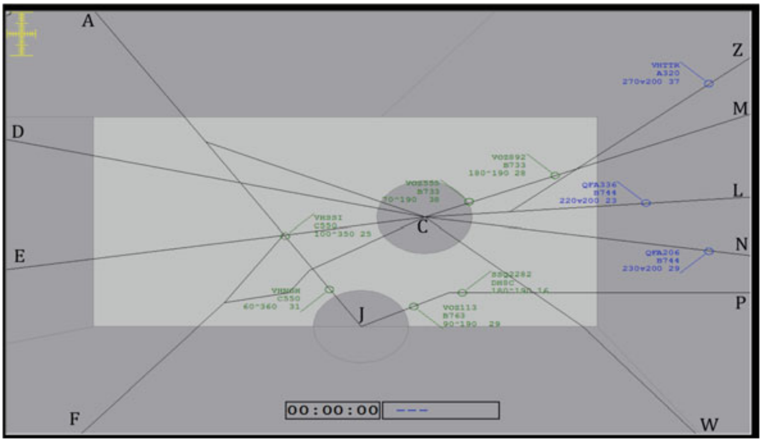


Fig. 2 ATC-labAdvanced radar screen with airspace with initial aircraft in the experimental stage

In our experiments, the pupil diameter is being recorded with an equipment based on an infrared system marketed by Tobii1 video system. The model used is the Tobii T120 with a sampling frequency of 120 Hz.

In a typical experiment, participants have two training sessions and one experimental session during three different days in the simulated air-traffic control tasks. During the training sessions participants learn the air-traffic control procedures in ATC-labAdvanced program. During each of the training sessions, participants perform the control task for one hour each day. Then, they perform a two hours experimental task on a different day.

The goal of the air-traffic task is to avoid conflicts that may arise between airplanes, or what is the same, to safe aircrafts from colliding. The first day the experimenter explain the objectives of the experiment to the participants and give them a training manual that they should read for about 20 min. The manual contained instructions informing them of important aspects such as altitude a aircraft should have in certain situations, what aircraft is under their control and how to avoid conflicts between two aircrafts.

For the 2 h practices we have been using in the existing scenarios included in the software of the air-traffic simulator. For the experimental session, we usually program new scenarios. In Fig. 2 we can see the initial radar screen participants see in the experimental session. The black letters are placed in the figure in order to explain the characteristics of the session. Those letters are not seen by participants.

We manipulated task complexity by presenting different number of airplanes. In the condition of low task complexity the scenario start with nine aircrafts, of which 6 were under the responsibility of the controller (green). Black letters indicate the points of origin of flights, which will cross the area under the responsibility in the controller. Then, 35 new airplanes appeared. In the high complexity task, the number of airplanes that appear during the session was 70.

The experimental design was a 23 × 2 Mixed Factorial. One of the independent variables manipulated was the time on task (TOT), which has been considered by intervals of 5 min (23 levels), the first interval of 5 min is taken as a baseline of each participant pupil diameter. The second independent variable was Traffic Density and it was measured by the number of aircrafs entering into the sector. It had two levels, one of High complexity (many planes) and one of Low Complexity (few airplanes). The dependent variables measured were the change in pupil diameter and the number of conflicts during the experimental session. The goal of an air-traffic controller is to avoid conflicts that may arise between airplanes, or what is the same, to safe aircrafts from colliding. Therefore, performance could be measured by the number of conflicts avoided by the participant. As more conflicts avoided, as better was the performance.

3 Some Results

As mentioned above, our purpose was to demonstrate that in order to keep an optimal level when performing the task for a long time is necessary to have a trigger factor to increase mental resources. To do this, our strategy is to train the participant in our experiments to reach that optimal level of performance and, then, we manipulate variables that could affect the level of activation, that is, the amount of resources invested in the task.

To make performance from the two Traffic Density conditions comparable we divided the number of conflicts for each participant by the number of possible conflicts in each 5 min interval. The number of possible conflicts was calculated as the combination of pairs of airplanes flying during the interval. Figure 3 shows

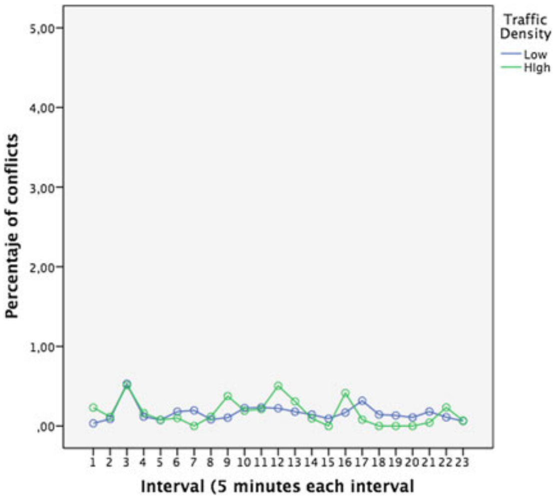


Fig. 3 Performance data from [21]

performance data from [21]. As we can see participants were able to keep performance almost perfect. The number of conflicts was less than 1 percent in both conditions of Traffic Density and along the 23 intervals of 5 min. Therefore, we can say that performance was not affected by TOT in the experiment.

Figure 4 shows the pupil diameter as a function of TOT and Traffic Density. The results clearly showed that the pupil diameter was reduced with the passage of time only when there was little traffic. When traffic was denser the pupil diameter remains the same throughout the 23 intervals of 5 min. If we take the diameter of the pupil as an index of activation we can say that activation was reduced when the task was not complex and performance can be maintained optimal with few resources. However, if the task was complex it was necessary to keep activation high to maintain optimum performance.

These results along with some others that are not predicted by traditional Theories of Resources have led to propose new resources theories in which a compensatory mechanism have been introduced that can lead to explain how it is possible to keep good performance even when a person is suppose to be fatigued [1]. This new resources theories are related to the Cognitive Activation Theory of Stress [23]. Thus, for example, if there is a situation in which the complexity increases and, therefore, the demand for resources is greater, the human organism can find additional supplies of resources that can be used to cope with the increased complexity by, for example, acting more actively on the task [24]. However, acting more actively on the task could not be sufficient. We analyzed the number of over responses of participants while they were performing the task to test this hypothesis. More precisely, we measured the number of times that they changed the route, the speed, and the altitude of airplanes during the task. As can be seen in Fig. 5, participants in both Traffic Density conditions showed the intensity of over activity. Therefore, we have to assume that the differences in activation shown in the pupil

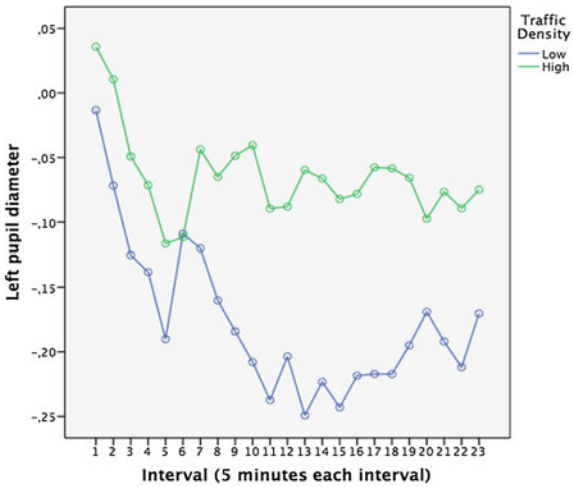


Fig. 4 Pupil diameter as a function of TOT and traffic density in [21]

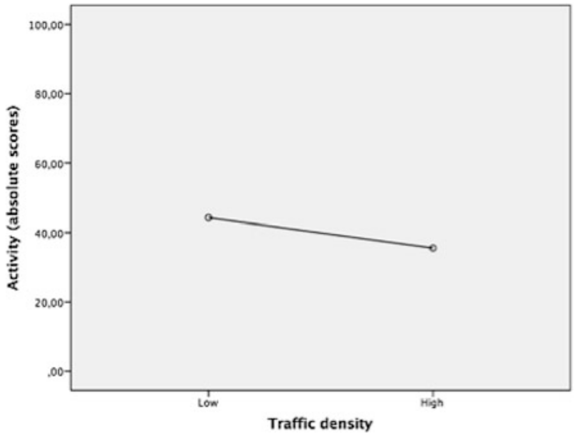


Fig. 5 Intensity of activity as a function of traffic density in Cañas et al. (in preparation)

dilatation must be affecting some internal process that is responsible for keeping performance optimal.

4 Discussion

What these data show is that the traditional explanation that accidents are caused by fatigue can be true but it is necessary to revise the postulates of the traditional theories of resources that have been proposed to explain this relationship between fatigue and accidents. If we take the physiological indices of consumption of resources while performing a task, we can see that the resources are "used" differently depending on variables such as the complexity of the task. This differential use of resources due to compensatory mechanisms can remove the effect of variables that have traditionally been associated with fatigue, such as the time on task. And that could be the reason for observing that there are fewer accidents at the end of the day or at the end of the week. The new theories of resources such as Hockey's Theory of Compensatory Control could be one of those theories [1].

References

1. Hockey, G.R.J.: Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol. **45**(1), 73–93 (1997)
2. Kahneman, D.: Attention and Effort. Prentice-Hall, Englewood Cliffs, NJ (1973)
3. Gopher, D., Donchin, E..: Workload: an examination of the concept. In Boff, K.R., Kaufman, L., Thomasm, J.P. (eds.) Handbook of Perception and Human Performance: Cognitive Processes and Performance (pp. 1–46) Wiley, USA (1986)
4. Grandjean, E.: Fatigue in industry. Br. J. Ind. Med. **36**(3), 175–186 (1979)

5. Lorist, M.M., Boksem, M.A., Ridderinkhof, K.R.: Impaired cognitive control and reduced cingulate activity during mental fatigue. Cogn. Brain. Res. **24**(2), 199–205 (2005)
6. Dawson, D., Ian Noy, Y., Härmä, M., Åkerstedt, T., Belenky, G.: Modelling fatigue and the use of fatigue models in work settings. Accid. Anal. Prevent. **43**(2), 549–564 (2011)
7. Stern, J.: The pupil of the eye: what can it tell us about 'mental processes'. Human Eng. Qual. Life (HQL) Q. **8**, 1–2 (1997)
8. INSHT (ed.): VII encuesta nacional de condiciones de trabajo de 2011. Madrid, INSHT (2012)
9. Hockey, R.: The psychology of fatigue: work, effort and control. Cambridge University Press (2013)
10. Kahneman, D., Beatty, J.: Pupil diameter and load on memory. Science **154**, 1583–1585 (1966)
11. Beatty, J., Lucero-Wagoner, B.: The pupillary system. Handbook of Psychophysiology **2**, 142–162 (2000)
12. LeDuc, P.A., Greig, J.L., Dumond, S.L.: Involuntary eye responses as measures of fatigue in US army apache aviators. Aviat. Space Environ. Med. **76**(Supplement 1), C86–C91 (2005)
13. Van Orden, K.F., Limbert, W., Makeig, S., Jung, T.P.: Eye activity correlates of workload during a visuospatial memory task. Hum. Factors **43**(1), 111–121 (2001)
14. Ahlstrom, U., Friedman-Berg, F.J.: Using eye movement activity as a correlate of cognitive workload. Int. J. Ind. Ergon. **36**(7), 623–636 (2006)
15. Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H., Laeng, B.: Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. J. Vis. **14**(4):1–20 (2014)
16. Laeng, B., Sirois, S., Gredebäck, G.: Pupillometry a window to the preconscious? Perspect. Psychol. Sci. **7**(1), 18–27 (2012)
17. Aston-Jones, G., Cohen, J.D.: An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu. Rev. Neurosci. **28**, 403–450 (2005)
18. Grandjean, E., Wotzka, G., Schaad, R., Gilgen, A.: Fatigue and stress in air traffic controllers. Ergonomics **14**(1), 159–165 (1971)
19. Luna, T.D., French, J., Mitcha, J.L.: A study of USAF air traffic controller shiftwork: Sleep, fatigue, activity, and mood analyses. Aviat. Space Environ. Med. **68**(1), 18–23 (1997)
20. Wilson, G.F., Russell, C.A.: Operator functional state classification using multiple psychophysiological features in an air traffic control task. Hum. Factors **45**(3), 381–389 (2003)
21. Cañas, J.J., López, A., Marchitto, M., Noriega, P.: Effects of time on task, task complexity, and effort on pupil dilatation (In preparation)
22. Fothergill, S., Loft, S., Neal, A.: ATC-labAdvanced: An air traffic control simulator with realism and control. Behav. Res. Meth. **41**(1), 118–127 (2009)
23. Ursin, H., Eriksen, H.R.: The cognitive activation theory of stress. Psychoneuroendocrinology **29**(5), 567–592 (2004)
24. Wickens, C.D.: Effort in human factors performance and decision making. Human Factors: J. Human Factors Ergon. Soc. **56**(8), 1329–1336 (2014)

A New Method for the Objective Registration of Mental Workload

Thea Radüntz

Abstract Complex and highly automated systems impose high demands on employees with respect to cognitive capacity and the ability to cope with workload. Prevention of over- and underload at workplaces with high cognitive demands can be achieved by objectively registering mental workload. Hence, the goal of this work is the development of such an objective method. We briefly introduce the so-called Dual Frequency Head Maps (DFHM) for registering mental workload by means of the electroencephalogram (EEG). Based on them, we obtain an index of mental state every 5 s ranging between the classes low, moderate, and high workload. Finally, we present results from a sample set of 54 people who executed cognitive tasks like switching and AOSPAN in a laboratory setting. We then verify the integrity of the new method by comparing the results with further workload relevant biosignal data, performance data, and the NASA-TLX questionnaire.

Keywords Mental workload · Electroencephalogram (EEG) · Signal processing · Pattern recognition

1 Introduction

Since the introduction of the PC in the 80s, the computerization of the modern working world is noticeably evolving ever faster. Although it aims to facilitate our life there is a growing consensus surrounding the negative consequences of inappropriate workload. This is related on the one hand to the impact of workload on employee health and on the other hand to increasing risks for the safety of other people due to elevated error rates. The consequences may arise from the inability to cope with increasing demands imposed on an individual's cognitive capacity and hence due to high mental workload [1–3]. However, the proliferation of automation

T. Radüntz (✉)
Federal Institute for Occupational Safety and Health, Unit 3.4 'Mental Health and Cognitive Capacity', Nöldnerstr. 40/42, 10317 Berlin, Germany
e-mail: raduentz.thea@baua.bund.de

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_24

can also be linked to monotonous tasks that reduce employees' arousal and induce underload [4–8].

There has been years of research targeting the registering of mental workload and the relationship between workload and performance. Nevertheless, there is no generally accepted, reliable, objective, and continuous method for mental workload registration that would allow the definition of an individual optimal workload range in which task solving is most efficient. Our aim is to develop such a method based on features from brain activity, the center of human information processing. Neuronal brain state monitoring can then be used for ergonomic evaluation and improvement of human-machine systems and hence contribute to the optimization of workload.

This article describes the development of a continuous method for neuronal mental workload registration. Cognitive tasks were conducted in a laboratory setting aiming to identify EEG features indicative of mental workload. For our task battery, we selected tasks reflecting executive functions [9]. Such functions are responsible for everyday actions demanding non-schema-based processing and requiring attentional control. Hence, the EEG features obtained should allow for the development of a generalized method that can measure mental workload independently from tasks.

2 Methods

2.1 Procedures and Subjects

The investigation took place in the shielded lab of the Federal Institute of Occupational Safety and Health. The experiment was fully carried out with each subject in a single day and consisted of a training phase where the subjects were familiarized with the tasks and the main experiment. Materials, procedures and the sample set have already been described in [10]. Briefly, the sample consists of 54 people between 34 and 62 years of age and shows high variability in respect to the cognitive capacity and hence to the expected mental workload.

We used several workload measuring methods. The collection of this additional workload indexing data was aimed at consolidating the development of our method by giving us the opportunity to control possible subject-dependent confounders but also further information in case of doubts at the end.

2.2 Tasks

Different cognitive task requirements were realized through the implementation of a task battery with the E-Prime application suite. Nine tasks of diverse complexity and difficulty inducing different levels of mental workload are included in our test battery [10, 11].

In this paper, we concentrate on the analysis and evaluation of four tasks: switch-PAR and switch-NUM as the easiest ones, switch-XXX as a switching task with working memory requirements, and AOSPAN as a demanding dual task (see Figs. 1, 2, 3, and 4). The latter is a translated version of the AOSPAN task developed by [12]. The analysis of rest measurements serves as a reference point measurement.

2.3 Subjective Ratings

Paired comparisons of the workload sources were conducted after each task during the training phase as the first part of the computerized version of the NASA-TLX questionnaire method [13]. Subjects were asked to rate the workload sources in 15 pairwise comparisons of NASA-TLX's six workload dimensions: mental demand, physical demand, temporal demand, performance, effort, frustration.

As part of the main experiment, following each task we then conducted the second part of the NASA-TLX, the ratings of its subscales. Subjects were asked to rate the task for each of the six workload dimensions within a 100-point range with 5-point steps. They indicated their rating by clicking on a 5-point step box with an optical mouse.

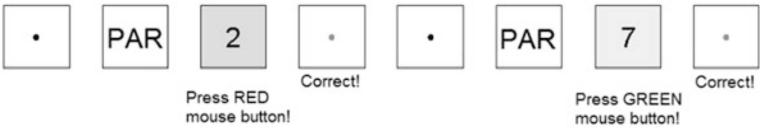


Fig. 1 Switch-PAR: The cue PAR appears on the screen followed by a number. This number is judged by its parity. The *RED mouse button* is pressed if the number is even, the *GREEN mouse button* if the number is odd

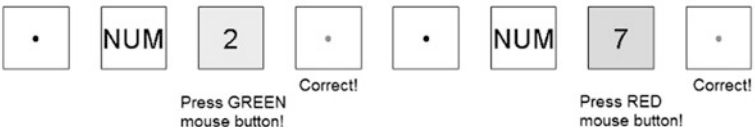


Fig. 2 Switch-NUM: The cue NUM will appears on the screen followed by a number. This number is judged numerically. The *RED mouse button* is pressed if the number is greater than 5, the *GREEN mouse button* if the number is less than 5

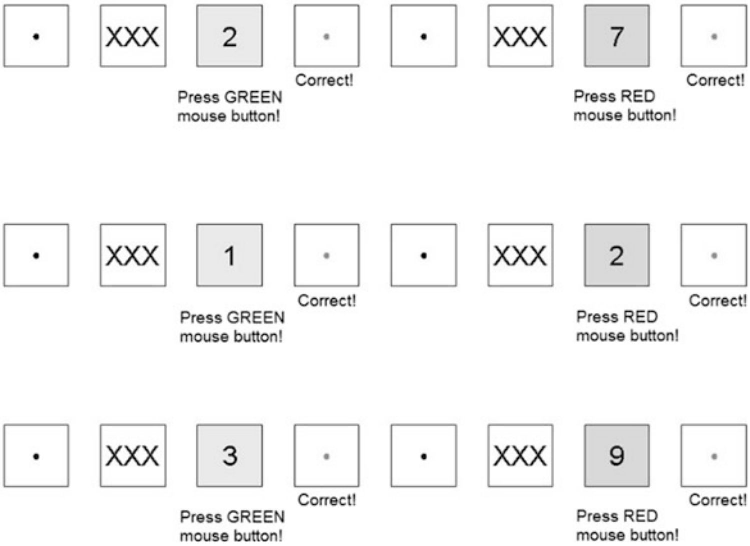


Fig. 3 Switch-XXX as a switching task with working memory requirements: This time, the sequence of the task cues is fixed: NUM, NUM, PAR, PAR, NUM, NUM, and so on… Subjects have to remember this sequence because now XXX appears in place of the cue. To review, NUM: *RED button* number greater 5, *GREEN button* number less 5; PAR: *RED button* even number, *GREEN button* odd number. If subjects lose their rhythm, a cue appears twice

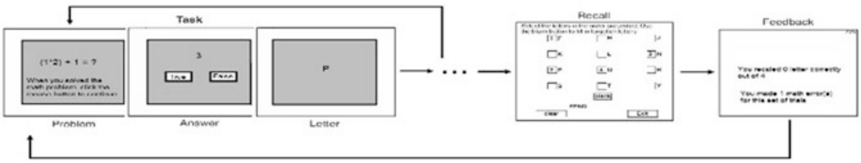


Fig. 4 AOSPAN as dual task (image adapted from [12]): memorize a set of letters in the order presented while simultaneously solving math problems. Trials consist of 3 sets of each set size, with the set sizes ranging from 3 to 7

2.4 *Physiological Measures*

In the main experiment during the execution of the tasks we registered the electroencephalogram (EEG), as well as further biosignal data (i.e. heart rate, blood pressure).

EEG. The EEG was captured by 25 electrodes placed at positions according to the 10–20-system and recorded with reference to Cz and at a sample rate of 500 Hz. For signal recording we used an amplifier from BrainProducts GmbH and their BrainRecorder software.

The recorded EEG signal is filtered with a bandpass filter (order 100) between 0.5 and 40 Hz. Subsequently, independent component analysis (ICA) is applied to the signal and the calculated independent components are visually inspected and classified as either an artifact or a signal component. The signal components are projected back onto the scalp channels. The artifact-corrected EEG signal is transformed to average reference and cut into segments of 10 s in length, overlapping by 5 s. Subsequently, the workload relevant frequency bands (θ: 4–8 Hz, α: 8–12 Hz) are computed over the segments using the Fast Fourier Transformation (FFT).

The newly developed method of dual frequency head maps (DFHM) is based on our analysis of the EEG spectra demonstrating an increase of the frontal theta band power and a decrease of the parietal alpha band power with increasing task difficulty level. Subsequently, labelling of the DFHM based on expert knowledge and classifier training is performed and workload is individually classified in the range of low load, moderate load, and high load [14]. The DFHM are computed for each EEG segment. Hence, the algorithm computes in an interval of 5 s a new workload index value. At the end, we calculate for each person and task three percentage values for the portion of the segments of each sector (LLS: low load segments, MLS: moderate load segments, HLS: high load segments).

Cardiovascular Parameters. Blood pressure was recorded continuously by the FMS Finometer Pro device. A finger cuff was placed around the subject's finger and systolic and diastolic blood pressure as well as the heart rate were detected automatically. The recorded data was processed in the time domain.

2.5 Performance

We concentrated on the analysis of the individual accuracy rates for all four tasks. For AOSPAN, correct responses include the number of sets in which the letters are recalled in correct serial order and correct math problem solving.

2.6 Statistical Analysis

Six ANOVAs were carried out utilizing repeated measures design, one within-subject factor (portion of LLS, portion of HLS, systolic BP, HR, accuracy rate or NASA-TLX), six levels (the four tasks and the two rest measurements) for the factors portion of LLS and HLS, systolic BP, and HR, or four levels (the four tasks) for the factors accuracy rates and NASA-TLX. Differences between the levels were examined and tested with a post hoc test (Bonferroni).

3 Results

3.1 Subjective Ratings and Performance

Subjective Ratings. Figure 5 shows the average workload index for the selected tasks switch-PAR, switch-NUM, switch-XXX, and AOSPAN as representatives of two low, a moderate and a high workload tasks. Workload means changed significantly during the experiment (Greenhouse-Geisser F(5.96; 316.01) = 65.023, $p < 0.001$). Post hoc analysis revealed significant changes of the subjectively rated mean workload index between all tasks apart from the two easy switch tasks among each other.

Performance. Figure 6 shows the average accuracy rates for the selected tasks switch-PAR, switch-NUM, switch-XXX, and AOSPAN. Accuracy rate means changed significantly during the experiment (Greenhouse-Geisser F(3.71; 196.67) = 173.256, $p < 0.001$). Post hoc analysis revealed significant changes of the mean accuracy rates between all tasks.

3.2 Physiological Measures

EEG. Analysis of the classified EEG segments demonstrates a proportional increase of the high load segments and a proportional decrease of the low load segments with increasing task difficulty level. The means of LLS and HLS changed significantly during the experiment (Greenhouse-Geisser F(2.79; 148.29) = 70.73, $p < 0.001$; Greenhouse-Geisser F(2.75; 145.54) = 65.22, $p < 0.001$). Results obtained from the assessment of the EEG segments are presented in Figs. 7 and 8.

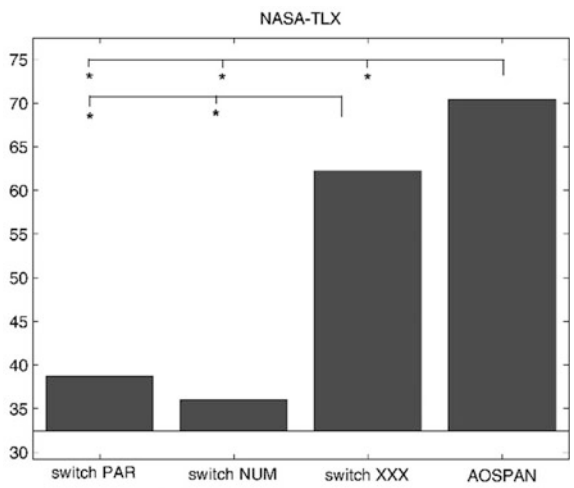


Fig. 5 NASA-TLX computed for switch-PAR, switch-NUM, switch-XXX, and AOSPAN over 54 subjects

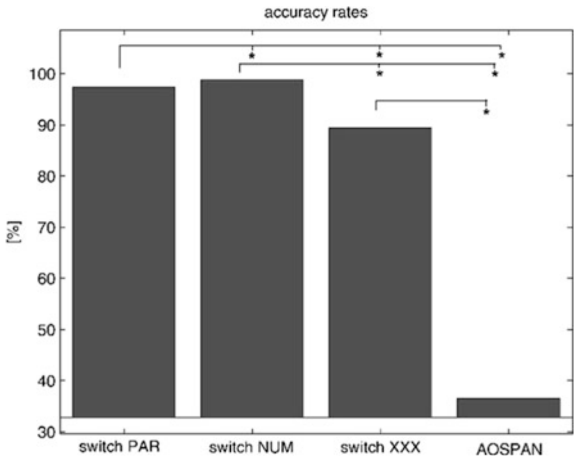


Fig. 6 Accuracy rates computed for switch-PAR, switch-NUM, switch-XXX, and AOSPAN over 54 subjects

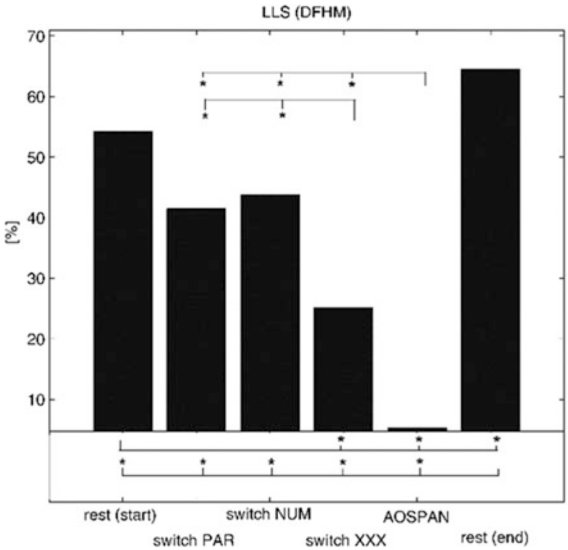


Fig. 7 EEG—proportion of LLS computed for switch-PAR, switch-NUM, switch-XXX, and AOSPAN over 54 subjects

Post hoc analysis of the proportion of HLS showed that the means were significantly larger the more difficult the tasks were. Significant differences were identified between all tasks and the rest measurement at the end as well as between the rest measurement at the beginning and all tasks except the switch-NUM task. No significant differences could be found between the two easy switch tasks nor among the rest measurements at the beginning and end of the experiment.

The proportion of LLS revealed similar significant changes between the four tasks. In respect of the rest measurements, significant differences could be found between the measurement at the end and all other measurements. No significant

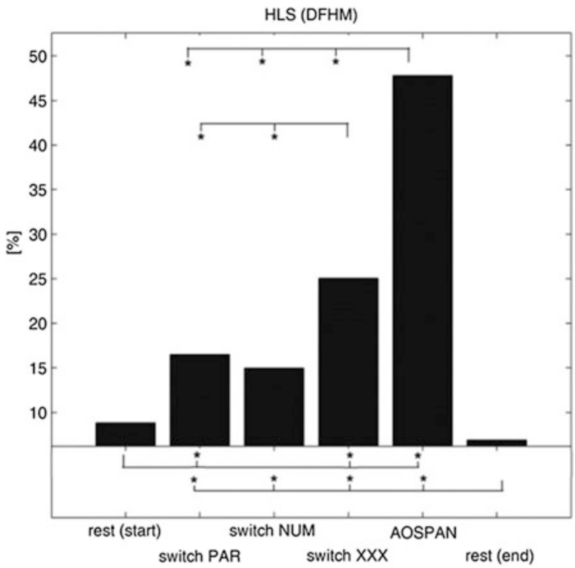


Fig. 8 EEG—proportion of HLS computed for switch-PAR, switch-NUM, switch-XXX, and AOSPAN over 54 subjects

changes could be identified between the rest measurement at the beginning and the two easy switch tasks.

Cardiovascular Parameters. Both systolic BP and HR differed between the measurements significantly (Greenhouse-Geisser F(4.45; 235.65) = 17.62, $p < 0.001$; Greenhouse-Geisser F(5.89; 312.26) = 20.92, $p < 0.01$).

HR during the rest measurement at the end was, according to post hoc analysis, lower than during all four tasks. HR during the rest measurement at the beginning was significantly lower than switch-NUM, switch-XXX, and AOSPAN. Furthermore, significant changes in HR could be found between all tasks except for switch-XXX and AOSPAN.

Systolic BP means were significantly larger during the AOSPAN task than in switch-PAR, switch-NUM, and the rest measurements. Additionally, they were significantly larger during switch-XXX than in the two easier switch tasks and the rest measurements. No significant changes could be found between the easy switch tasks switch-PAR and switch-NUM. Furthermore, there were no significant changes between the rest measurements at the beginning and at the end, between the rest measurement at the end and the two easier switch tasks, and between the rest measurement at the beginning and switch-PAR.

Results of HR and systolic BP are presented in Figs. 9 and 10.

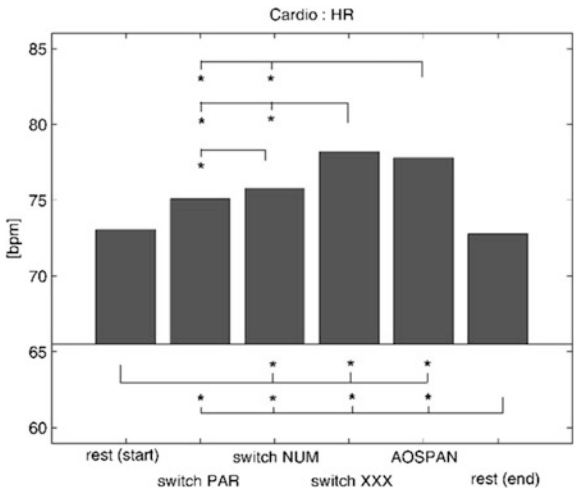


Fig. 9 Heart rate computed for switch-PAR, switch-NUM, switch-XXX, and AOSPAN over 54 subjects

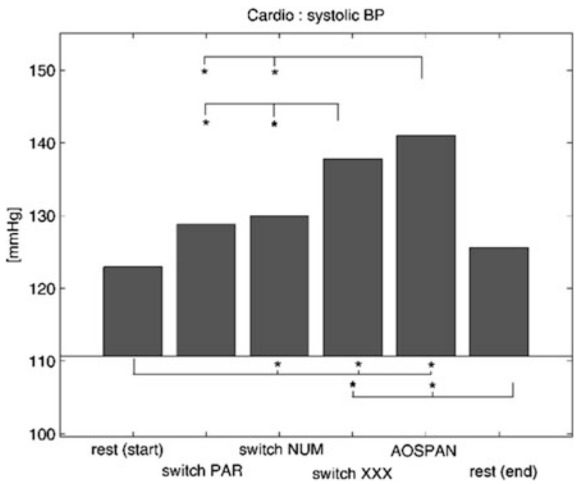


Fig. 10 Systolic BP computed for switch-PAR, switch-NUM, switch-XXX, and AOSPAN over 54 subjects

4 Discussion

The registration of mental workload by means of the EEG is the central issue addressed by this paper. We induced different levels of mental workload on the basis of a task battery, but for the sake of convenience we have concentrated here on the switch-PAR, switch-NUM, switch-XXX and AOSPAN tasks. The cognitive requirements of the first two tasks are quite low and the tasks can be assumed to be representative of an easy task. The switch-XXX task is more demanding due to higher requirements on the working memory and rule switching. It can be classified as a moderate to difficult task but not as challenging as the AOSPAN task.

The AOSPAN task demands memory control while dealing with distraction due to the math problem solving. It is a dual-task with high workload requirements.

Subjective ratings derived from the NASA-TLX questionnaire demonstrate significant workload differences between the more demanding tasks (switch-XXX and AOSPAN) and the easy tasks. No significant difference could be identified among the subjects in respect of their experienced workload between the two easy tasks switch-PAR and switch-NUM.

Accuracy rates show significant differences between all tasks but remarkably larger breaks between the difficult AOSPAN task and all others but also between the moderate task and the two easy tasks. Although there is a significant difference between switch-PAR and switch-NUM, it is fairly clear that the two tasks are located in the low workload level compared to the other two tasks. However, it can be seen that the switch-PAR task is slightly more difficult than the switch-NUM task.

Cardiovascular parameters indicate significant differences between the more demanding tasks (switch-XXX and AOSPAN) and the two easy tasks. They also show significant differences between both demanding tasks and the rest measurements at the beginning and the end of the experiment. What is more, HR indicates small differences between both easy tasks but also between the rest measurement at the end and all other tasks. Surprisingly, no significant difference can be observed among the more difficult tasks. Here we have to ask whether the cardiovascular parameters are not able to finely distinguish among more demanding tasks, potentially due to a ceiling effect.

The EEG and the frequently observed variability of the θ- and α-band according to attention, fatigue and mental workload, constitute the theoretical background for the new method of DFHM. The obtained index can be used for neuronal mental state monitoring and ranges between low, moderate, and high workload. Results analyzing the proportions of the HLS and LLS are in concordance with the results expected based on difficulty levels resulting from the requirements of the tasks. The most demanding AOSPAN task contains significantly less segments of low load than the other tasks and the rest measurements. The switch-XXX task includes less LLS than the easy switch tasks and the rest measurements, while the easy switch tasks include less than the rest measurements. Furthermore, the rest measurement at the beginning has less LLS than the rest measurement at the end indicating that the workload at the beginning is a bit higher than at the end of the experiment. All these differences were found to be significant.

In respect of the HLS, the AOSPAN task again shows substantially higher values than all other measurements. Also in consideration of its small proportion of LLS, AOSPAN is a high mental workload task. Switch-XXX includes significantly higher proportions of HLS than the easy switch tasks and the rest measurements. Finally, both easy switch tasks comprise less segments of high load than the rest measurements. Interestingly, there is no significant difference of HLS's proportion among the rest measurement at the beginning and the switch-NUM task, which we already have noticed to be slightly easier than the switch-PAR task according to the accuracy rates. Furthermore, there is no significant difference in the proportion of

HLS among the rest measurements, although there is some indication that the activation of the subjects at the beginning is a bit higher than at the end, similar to the findings from the analysis of the proportion of LLS. Hence, we can assume that the index is able to distinguish between very small gradual differences, in particular when both HLS and LLS are considered simultaneously.

To sum up, our results from the new DFHM method for measuring mental workload are solidly in line both with the accuracy rates and with the subjective ratings. Furthermore, they are in concordance with our expectations deriving from the known task requirements and also with the cardiovascular parameters.

Acknowledgments I would like to thank Dr. Sergei Schapkin, Dr. Patrick Gajewski, and Prof. Michael Falkenstein for selection of the battery's tasks. I would like to thank Mr. Ludger Blanke for technical support during the timing tests for the tasks. In addition, I would like to thank Ms. Xenija Weißbecker-Klaus, Mr. Robert Sonnenberg, Dr. Sergei Schapkin, and Ms. Marion Freyer for general task testing and for conducting the laboratory experiments. Furthermore, I would like to thank Ms. Marion Freyer for daily operational support and our student assistant Jon Scouten for proofreading. I would also like to express my sincere appreciation to Dr. Gabriele Freude for her valuable and constructive suggestions.

More information about the project where our EEG data was acquired can be found under the following link: http://www.baua.de/de/Forschung/Forschungsprojekte/f2312.html?nn=2799254.

References

1. Kompier, M.A.J., Kristensen, T.S.: Organisational work stress interventions in a theoretical, methodological and practical context. In: Dunham, J. (ed.) Stress in the Workplace: Past, Present and Future, pp. 164–190. Whurr Publishers, London (2001)
2. Landsbergis, P.A., Cahill, J., Schnall, P.: The changing organisation of work and the safety and health of working people: A commentary. J. Occup. Environ. Med. 61–72 (2003)
3. NIOSH, N.: The changing organization of work and the safety and health of working people. Tech. Rep. 2002-116, National Institute for Occupational Safety and Health (NIOSH) (2002)
4. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation induced complacency. Int. J. Aviat. Psychol. **3**, 1–23 (1993)
5. Parasuraman, R., Mouloua, M., Molloy, R.: Monitoring automation failures in human machine systems. In: Mouloua, M., Parasuraman, R. (eds.) Human Performance in Automated Systems: Current Research and Trends, pp. 45–49. Hillsdale, NJ, Earlbaum (1994)
6. Hacker, W., Richter, P.: Psychische Fehlbeanspruchung. Psychische Ermüdung, Monotonie, Sättigung und Stress (Spezielle Arbeits- und Ingenieurpsychologie in Einzeldarstellungen), vol. 2, 2 edn. Springer, Berlin (1984)
7. Debitz, U., Gruber, H., Richter, G.: Psychische Gesundheit am Arbeitsplatz. Teil 2: Erkennen, Beurteilen und Verhüten von Fehlbeanspruchungen, vol. 2, 3 edn. InfoMediaVerlag (2003)
8. May, J.F., Baldwin, C.L.: Driver fatigue: the importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transportation Research Part F 12, 218–224 (2008), last accessed on 03 Nov 2011
9. Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D.: The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. Cogn. Psychol. **41**(1), 49–100 (2000), http://psych.colorado.edu/~tor/Papers/Unity_Diversity_Exec_Functions.pdf. Last accessed on 17 Mar 2014
10. Radüntz, T., Freude, G.: Towards a continuous method for mental workload registration. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics. LNCS, vol. 9174,

pp. 176–187. Springer International Publishing (2015). http://dx.doi.org/10.1007/978-3-319-20373-7_17
11. Radüntz, T.: Neuronal mental workload registration during execution of cognitive tasks. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics. LNCS, vol. 8532, pp. 59–69. Springer International Publishing (2014)
12. Unsworth, N., Heitz, R.P., Schrock, J.C., Engle, R.W.: An automated version of the operation span task. Behav. Res. Methods **37**(3), 498–505 (2005). http://webs.wofford.edu/boppkl/coursefiles/Thesis/articles/UnsworthHeitz_2005_autoospan.pdf. Last accessed on 02 Apr 2014
13. Hart, S.G., Staveland, L.E.: Development of the NASA TLX: results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North Holland, Amsterdam (1988)
14. Radüntz, T.: Kontinuierliche Bewertung psychischer Beanspruchung an informationsintensiven Arbeitsplätzen auf Basis des Elektroenzephalogramms. Ph. D. thesis, Humboldt-Universität zu Berlin, Department of Computer Science, Berlin. (2016). http://edoc.hu-berlin.de/docviews/abstract.php?id=42402

Part V
Assessing Cognition and Performance

Effect of Time Pressure on Work Efficiency and Cognitive Judgment

Junpei Yabuki and Hiroshi Hagiwara

Abstract It is often necessary to work under time constraints. Previous studies reported that a reasonable amount of time pressure can improve performance. However, they did not reveal what kind of pressure should be applied. This study investigated the ability of subjects to perform under time constraints and different kinds of pressure. We investigated work efficiency, cognitive judgment and the effect of time pressure on brain activity using an electrocardiogram (ECG) and functional near infrared spectroscopy (NIRS). In this experiment, we used four types of time pressure. We applied each type of time pressure to subjects during the performance tests. For all performance tests, the results for auditory time pressure were better than those for no time pressure. In conclusion, we found that the subjects performed better when working under auditory time pressure.

Keywords NIRS · Time pressure · oxyHb · Work efficiency · Cognitive judgment

1 Introduction

It is often necessary to work under time constraints. These time constraints create pressure on people, which affects work efficiency [1]. On the other hand, previous studies reported that a reasonable amount of time pressure can improve performance [2]. However, they did not reveal what kind of pressure should be applied. Accordingly, in our study, we considered a visual stimulus and an auditory stimulus. This study investigated the ability of subjects to perform in timed conditions

J. Yabuki (✉)
Graduate School of Information Science and Engineering, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan
e-mail: is0143xp@ed.ritsumei.ac.jp

H. Hagiwara
College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_25

under different kinds of pressure. The purpose of this study was to investigate work efficiency, cognitive judgment and the effect of time pressure on the brain activity and HF (high frequency component of heart rate variability) using an electrocardiogram (ECG), functional near infrared spectroscopy (fNIRS) and the results of a performance evaluation.

2 Experimental Method

2.1 *Subjects*

Thirteen healthy, non-medicated college students (11 men, 2 women; 20–22 years old) participated in the study. All subjects provided written informed consent prior to participation. The subjects refrained from excessive eating and drinking the night before the experiment. In addition, the subjects refrained from engaging in prolonged or strenuous exercise on the morning of the experiment. Blood hemoglobin concentrations in the brain were measured with the NIRStation (Shimadzu, Japan), a near-infrared imaging device. The measurement sites for NIRS are shown in Fig. 1. In this report we focused on the frontal probe and analyzed data from channel 1 (right frontal cortex), channel 10 (center frontal cortex), channel 19 (left frontal cortex), all three of which are involved in cognition or judgment, and channel 25 (dorsolateral prefrontal cortex, near Brodmann area 46), which is involved in attention and control. In addition, ECGs were measured with a Polymate (Digitex Lab, Japan). The sites of the ECG electrodes were monitored with a 4-lead ECG. Before attaching the electrodes, skin cream was applied to the subject's skin to reduce impedance.

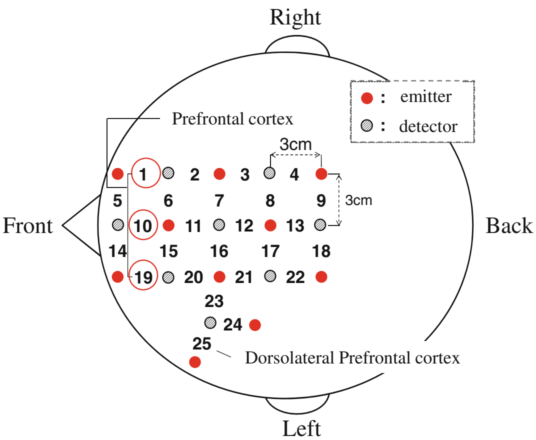


Fig. 1 Schematic indicating the mounting position for the probes. Probe spacing is 3 cm. We analyzed channels *1*, *10* and *19* (prefrontal cortex) and channel *25* (dorsolateral prefrontal cortex), circled in *red* on the schematic

2.2 Tasks

We used two tasks in the experiment: a click-and-drag task and a Stroop test. In the click-and-drag task (Fig. 2a), the participant was required to click on a black square on the left side of the screen and drag it onto a black target square on the right side of the screen [3]. In this experiment, two levels of usability were provided, with target squares of 4 and 8 pixels (px). These sizes were chosen to give clear differences in the size of the target. A total of 100 black squares had to be dragged onto the targets. In the Stroop test, four images were displayed, as in Fig. 2b. The subjects were required to decide whether a word matched a color and a picture. They were asked to indicate the number of matches using buttons numbered 1–4. The click-and-drag task and the Stroop test were conducted for 2 min each. The click-and-drag task and the Stroop test were designed to measure monotonous work efficiency and cognitive judgment, respectively.

2.3 Method for Applying Time Pressure

In this experiment, we used four types of time pressure. The first was no time pressure (NTP). The second was auditory time pressure (ATP). In ATP, the time remaining for the task was communicated to the subject verbally for remaining

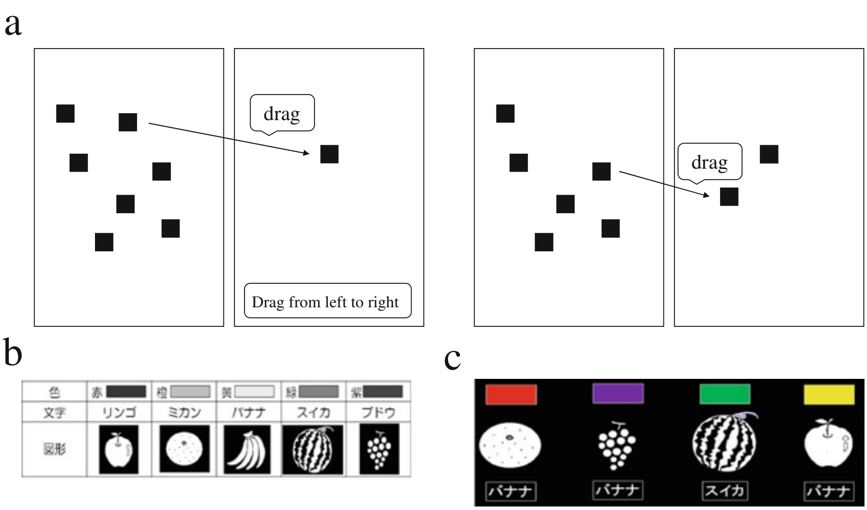


Fig. 2 **a** The click-and-drag task: participants were required to drag the *black square* from the left side of the screen on to a target that appeared at a random position on the right side of the screen. **b** The Stroop test: the five images that were displayed during the test. **c** An example of the display screen during the Stroop test

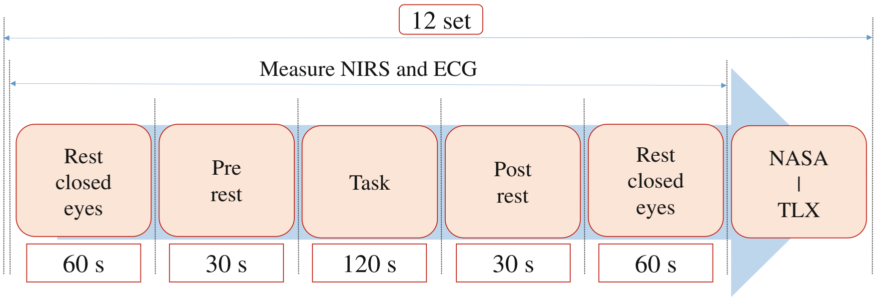


Fig. 3 Schematic of the experimental design. For baseline correction, participants were placed under a control condition (rest eyes closed) for 60 s before and after the tasks

times of 120, 90, 60, 50 and 40 s. When there was 30 s remaining, the remaining time was communicated at intervals of one second. The third method was visual time pressure (VTP). In VTP, the remaining time for the task was communicated visually with a countdown timer. The fourth method was auditory and visual time pressure (AVTP), which was a combination of the ATP and VTP methods. We applied each type of time pressure to subjects during the performance tests.

2.4 Experiment Protocol

The experimental protocol is shown in Fig. 3 and was as follows: 60 s rest (closed eyes), 30 s rest (eyes open), 120 s conducting task (performing click-and-drag task or Stroop test), 30 s rest (eyes open) and 60 s rest (eyes closed). The NASA Task Load Index (NASA-TLX) was evaluated after conducting each task. The order the tasks were performed in was randomized in every experiment to exclude an order effect.

3 Analysis

3.1 NIRS

Using NIRS, it is possible to measure the oxygenated hemoglobin concentration (oxyHb) [3], deoxygenated hemoglobin concentration, and total hemoglobin concentration. In this study, we focused on oxyHb, because oxyHb correlates with changes in the cerebral blood flow. NIRS waveforms carry important information, because NIRS data are not absolute but relative to baseline values. Therefore, to assess NIRS waveforms, calculating the average oxyHb is not sufficient. In our previous study [4], we used δoxyHb to assess the NIRS waveform. First, we

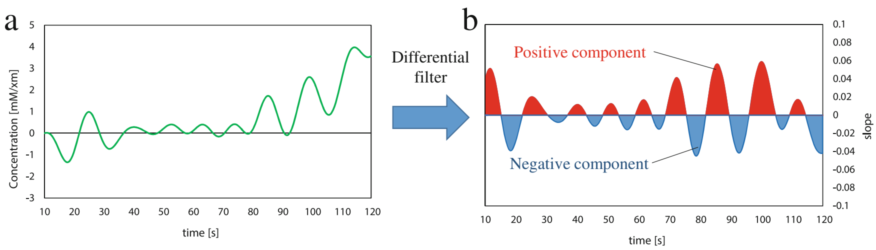


Fig. 4 **a** oxyHb data after applying a BPF; **b** oxyHb data after applying a differential filter. The initial 10 s of a task were excluded when calculating the δoxyHb because the neurovascular coupling reaction takes 6–10 s

applied a bandpass filter (BPF; 0.001–0.1 Hz) to oxyHb (Fig. 4a). After applying the BPF, we applied a differential filter to the oxyHb (Fig. 4b). Following the application of the differential filter to the oxyHb, we defined a positive component as a result greater than zero and a negative component as a result less than zero. We defined δoxyHb as the sum of the positive component and the negative component. By using this measure of δoxyHb, we can assess the change in the oxyHb during a task. A positive (negative) value of δoxyHb indicates an overall increase (decrease) in the oxyHb level.

3.2 *Performance Test*

The performance during the click-and-drag task was measured with the number of squares moved in 2 min. The performance during the Stroop test was evaluated with the number of correct answers and the positive reaction rate (the fraction of answers which were correct) in 2 min of the task. The effect on performance of the different types of time pressure was investigated by comparing the performance under NTP with the other time pressure conditions for each task.

3.3 *Other Indices*

Subjects were also assessed with the NASA-TLX [5], which evaluates 6 measures of workload: mental demand, physical demand, temporal demand, overall performance, frustration level and effort. In this study, we were interested in the temporal demand and frustration level.

The high frequency component of heart rate variability (HF) measured by the ECG represents the activity of the parasympathetic system.

4 Results

4.1 Click-and-Drag Task

Figure 5 shows the results of the click-and-drag task. When the size of the squares was 4 px, the number of points moved did not differ with the different methods for applying time pressure. When the size of the squares was 8 px, more squares were moved under the ATP and AVTP methods than under the NTP method. Comparing the number of squares moved for the NTP and ATP methods, borderline statistically significant differences were found ($p < 0.1$). In the comparison of the NTP and AVTP methods, statistically significant differences were found ($p < 0.05$).

4.2 Stroop Test

Figure 6a shows the number of correct answers in 2 min for the Stroop test, and Fig. 6b shows the positive reaction rate for the Stroop test. Both plots were normalized for each subject. The numbers of correct answers for the ATP and AVTP methods were greater than for the NTP method. However, there were fewer correct answers for the VTP method than the NTP method. In the comparison of the NTP and AVTP results, borderline statistically significant differences were found ($p < 0.1$). The positive reaction rates for the ATP and AVTP methods were greater than for the NTP method. For the VTP method, the positive reaction rate was lower than with the NTP.

4.3 ECG (HF)

Figure 7 shows the average ECG HF values. The changes in the average ECG HF values were investigated between 61–90 s and 91–120 s to determine if the

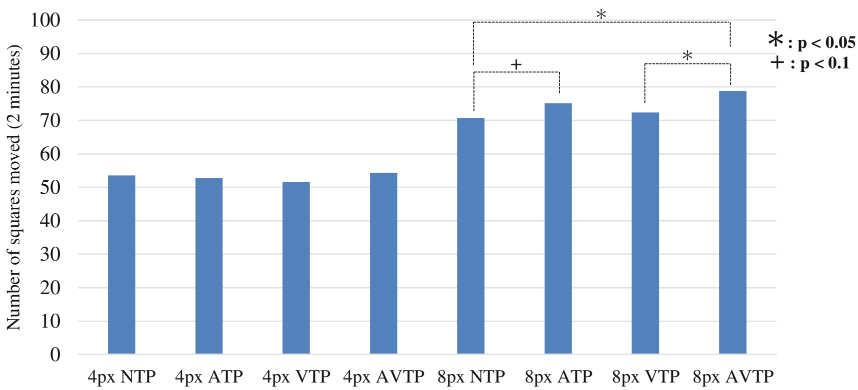


Fig. 5 Results for the click-and-drag task under the different time pressures

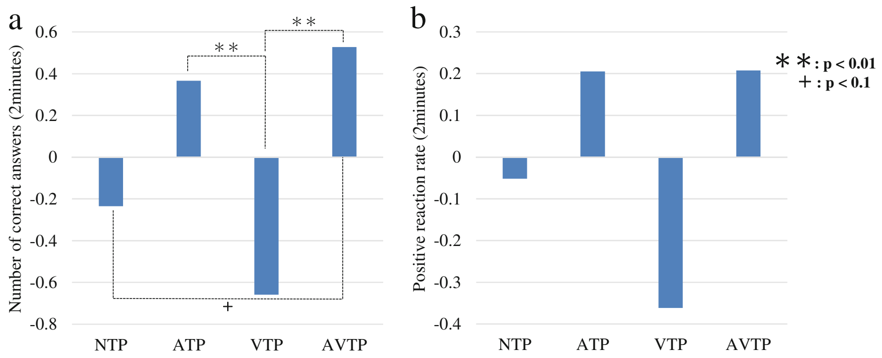


Fig. 6 **a** Mean change in the number of correct answers in the Stroop test. **b** Mean change in the positive reaction rate for the Stroop test. The results in **a** and **b** have been normalized

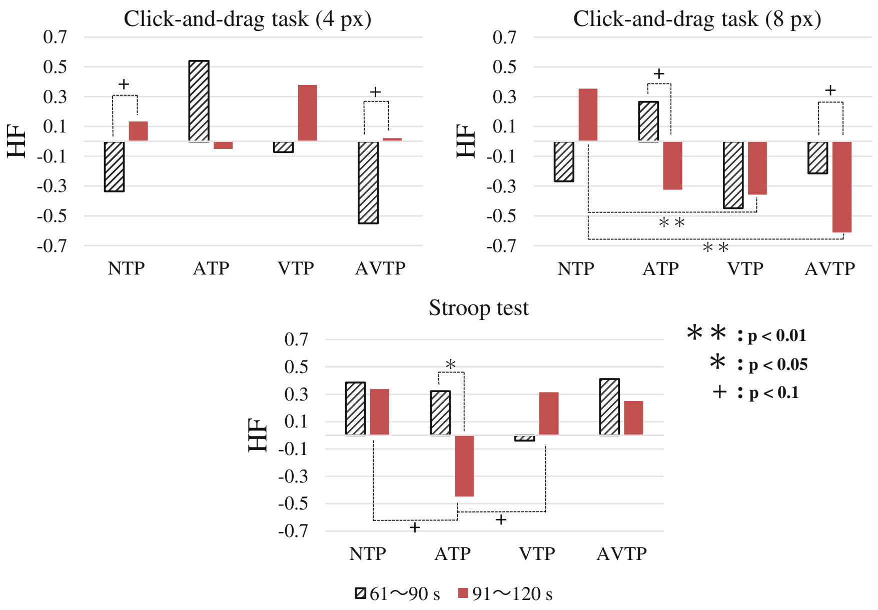


Fig. 7 HF values measured by ECG. All results were normalized for each subject

participants showed any stress due to time pressure. In the click-and-drag task with 4 px squares, the HF decreased (increased) under ATP (AVTP). In the click-and-drag task with 8 px squares, HF increased for both ATP and AVTP. Similarly, for the Stroop test, HF increased under both ATP and AVTP. In the click-and-drag task with 4 px squares, comparing the HF over the periods of 61–90 s and 91–120 s with the start time under ATVP, the differences were of borderline statistical significance ($p < 0.1$). Comparing the same periods in the click-and-drag task with 8 px squares, both the ATVP and ATP conditions were of borderline statistical significance ($p < 0.1$). In the Stroop test, only the results under ATP for these time periods were of even borderline significance ($p < 0.1$).

4.4 NIRS

Figure 8 shows the δoxyHb values for channels 1 and 25. Applying ATP and AVTP resulted in an improvement in the performance test. It was also investigated whether subjects felt increased pressure between 61 and 90 s and 91 and 120 s. In channel 1, in the click-and-drag task with 4, 8 px squares, and the Stroop test there

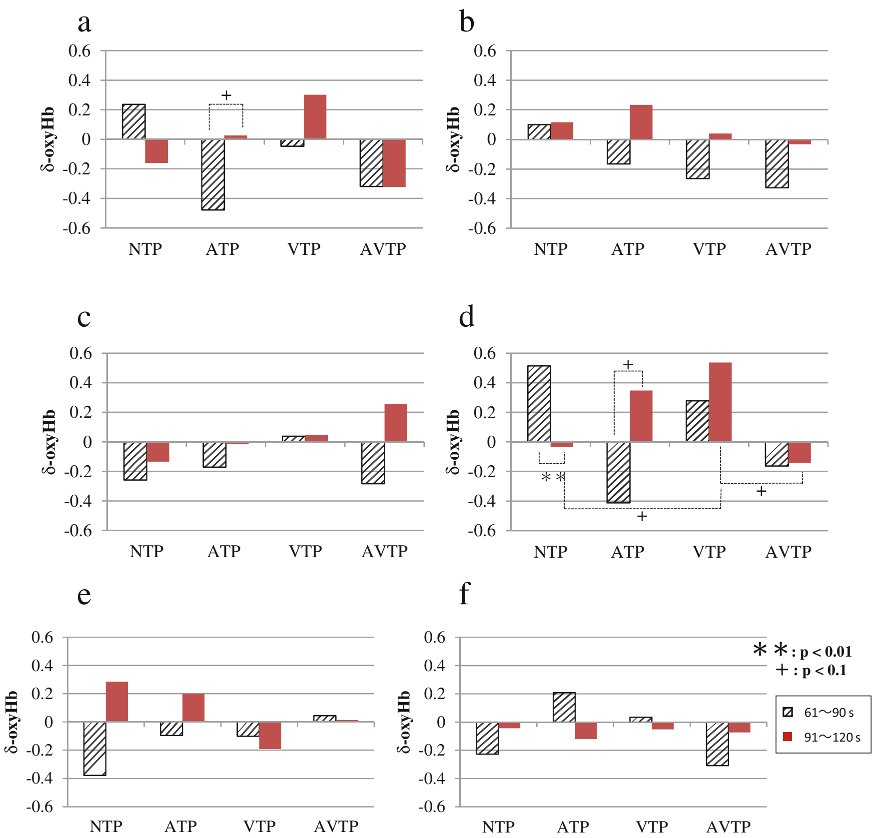


Fig. 8 **a** δoxyHb values for channel 1 in the click-and-drag task with 4 px squares. **b** δoxyHb values for channel 1 in the click-and-drag task with 8 px squares. **c** δoxyHb values for channel 1 in the Stroop test. **d** δoxyHb values for channel 25 in the click-and-drag task with 4 px squares. **e** δoxyHb values for channel 25 in the click-and-drag task with 8 px squares. **f** δoxyHb values for channel 25 in the Stroop test

was an increase in the oxyHb under both the ATP and AVTP methods. In channel 1, in the click-and-drag task with 4 px squares, in the comparison of the 61–90 s and 91–120 s periods under the ATP method, borderline statistically significant differences were found ($p < 0.1$). Similarly, in channel 25, in the click-and-drag task with 4 px squares for both the ATP and AVTP methods, subjects showed an increased oxyHb. The opposite behavior was seen in channel 25 for the Stroop test under ATP. In channel 25, in the click-and-drag task with 4 px squares, in the comparison of the 61–90 s and 91–120 s periods, marginally statistically significant differences were found ($p < 0.1$) for the change in pressure.

4.5 NASA-TLX

Considering the NASA-TLX for temporal demand in all tasks, the scores under ATP and AVTP were higher than those under NTP. The differences in the temporal demand scores for ATP and AVTP with NTP for all tasks were statistically significant ($p < 0.01$). Also, in the Stroop test, the frustration scores under ATP and AVTP were higher than those under NTP and VTP. Statistically significant differences were found between the ATP and AVTP scores with the VTP scores for frustration ($p < 0.01$).

5 Discussion

The subjects' performance did not differ with the different types of time pressure in the click-and-drag task with 4 px squares. However, when the squares were 8 px, the performance under the ATP and AVTP methods was better than under the NTP method. These results suggest that people are more responsive to time pressure when working on relatively simple tasks. In the Stroop test, the frustration scores under the ATP and AVTP methods were higher than those under the NTP and VTP methods. This result suggests that there may be a tendency to feel more stress under auditory temporal pressure than under visual temporal pressure.

For all tasks, under the ATP and AVTP methods, as the time limit for the task is approaching, the δoxyHb values of channels 1 and 25 were shown to increase. Also, as the time limit of the task was approaching, the ECG HF values tended to decrease. In addition, for all tasks, the results were improved with the application of auditory pressure. These results suggest that applying a light auditory temporal pressure can improve performance when conducting simple tasks.

Acknowledgment 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. Lin, Y.L., Hwang, S.L.: The application of the loglinear model to quantify human error. Reliab. Eng. Syst. Saf. **37**(2), 157–165 (1992)
2. Yamazaki, T., Karashima, M., Saitou, M.: The Influence of Time Constraints on the Performance in Decision Making. Jpn. J. Ergon. **38**, 320–321 (2002)
3. Hoshi, Y., Kobayashi, N., Tamura, M.: Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. J. Appl. Physiol. **90**, 1657–1662 (2001)
4. Iwasaki, H., Hagiwara, H.: Availability and future prospects of functional near-infrared spectroscopy (fNIRS) in usability evaluation. Human Factors Ergon. **27**, 6368–6377 (2012)
5. Miyake, S., Kamishiro, M.: Subjective mental workload assessment technique, an introduction to NASA-TLX and SWAT and a proposal of simple scoring methods. Jpn. J. Ergon. **29**(6), 399–408 (1982)

Cognitive Architecture Based Platform on Human Performance Evaluation for Space Manual Control Task

Yanfei Liu, Zhiqiang Tian, Yuzhou Liu, Junsong Li and Feng Fu

Abstract Based on ACT-R cognitive architecture, the paper proposes an integrated research and development platform to investigate astronaut's cognitive behavior. By applying this platform, astronaut's cognitive behavior model can be built; by abstracting the parameter of astronaut's behavior, cognition and special task, the characteristic of astronaut's cognitive processing can be analyzed; by software emulation technics, the astronaut's cognition and behavior simulation platform is constructed, and the astronaut's performance for special spaceflight task can be evaluated. As a study case, spacecraft manual control rendezvous and docking (RvD) task is chosen to investigate influential factors on human performance. The results show that, both the model's simulating cognitive process and model's running result are consistent with real situation, and the model's specific parameter can map certain cognitive characteristic.

Keywords Cognitive architecture · Performance evaluation · Cognitive modeling · Cognitive behavior

Y. Liu · J. Li · F. Fu
Department of Computer Science and Technology,
Zhejiang Sci-Tech University, No. 928, 2nd Avenue,
310018 Xiasha, Xiasha Higher Education Zone, Hangzhou, China
e-mail: yliu@zju.edu.cn

J. Li
e-mail: ljscg@sina.com

F. Fu
e-mail: fufeng@zstu.edu.cn

Z. Tian (✉)
National Key Laboratory of Human Factors Engineering,
Astronaut Research and Training Centre of China, No. 26,
Beiqing Road, 100094 Haidian District, Beijing, China
e-mail: tianzhiqiang2000@163.com

Y. Liu
Community College of Allegheny County, 3 Allegheny Center,
Pittsburgh, PA 15212, USA
e-mail: rcdl@me.com

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_26

1 Introduction

New space explorations are rapidly developing ever than before nowadays. However, for the astronauts boarding on the spaceship/space-station, they are exposed to numerous stressors during spaceflights, such as microgravity, confinement, and radiation, all of which may impair human cognitive capabilities. While some critical operations for spaceflight, such as handling robotic arms, extravehicular activities, and driving the spacecraft, etc., fault operation may cause serious disasters. The cognition for crewmembers will affect task's performance; therefor essential operating skills on spaceflight must be developed as a team member in an environment of highly dynamic, fast-changing, and even sometimes unpredictable environment.

To improve crewmembers' performance, a lot of researches have been conducted. Patricia M. Jones and Edna Fiedler illustrate a selected set of topics from three broad frameworks that represent both health and performance issues in human spaceflight [1]. The first one of them is the 2005 National Aeronautics and Space Administration bioastronautics roadmap [2]. The other framework is organized around stressors and stresses. Stressors are defined as exogenous events and stresses are the organism's response [3]. The third one is the 2009 risk list from National Aeronautics and Space Administration's Human Research Program [4]. A simplified representation of these issues is also shown with Diagram, which organizes these issues and risks as a health–task performance continuum. However, due to limitations for experimental conditions, uncertainty and poor features of experimental results in the study of human cognitive behavior for spaceflight, experimental researches are difficult to implement in reality. In the meantime, for the restriction and deficiency of the studies on human mind, using a computer modeling and simulation method to investigate human cognition to improve performance becomes a new method of study on human factors in spaceflight tasks. Especially nowadays modeling based cognitive architecture (such as ACT-R) to explore human performance in complex systems is generally accepted. Zhang et al. pose time-fuel optimal control model to modeling human control strategies to simulated rendezvous docking tasks [5] and Liu et al. evaluate human performance by constructing assessment platform based on architecture [6].

2 Human Performance Evaluation Based on ACT-R

Cognitive modeling is to produce a computational model for how people perform tasks and solve problems, based on psychological and Biological principles. Cognitive architectures are theories of cognition that try to capture the essential representations and mechanisms that underlie cognition [7]. There are several well-known cognitive architectures such as Soar, EPIC and CLARION etc. Each of them has its unique advantages.

2.1 Features of ACT-R

Adaptive Control of Thought—Rational (ACT-R) is one of the most typical and widely used cognitive architecture, it has been used to successfully model a variety of behavioral phenomena and has proven particularly successful at modeling tasks with a demanding cognitive component [8]. Dozens of ACT-R models have been developed and empirically validated through an active user community in academic, industrial, and military research laboratories. One important feature of ACT-R that distinguishes it from other theories in the field is that it allows researchers to collect quantitative measures that can be directly compared with the quantitative measures obtained from human participants. With the research continues, ACT-R evolves ever closer into a system which can perform the full range of human cognitive tasks: capturing in great detail the way we perceive, think about, and act on the world.

There are three major competing categories of cognitive architecture: symbolic architectures, subsymbolic (emergent) architectures, and hybrid approaches as the mixture of the two [9]. ACT-R is the hybrid cognitive architecture that incorporates a productions system—a computer system consisting of a series of rules (productions), a database (or "working memory") of facts, and an algorithm—as its symbolic structure. The subsymbolic structure of ACT-R is represented by a set of massively parallel processes that can be summarized by a series of mathematical equations; these subsymbolic equations control many of the symbolic processes (i.e., the processes of the system changing from one state to another). ACT-R can be regarded as the central processing unit of the production system that is surrounded by perceptual and motor systems. The architecture places an emphasis on knowledge learning and rule selection mechanisms [10]. Compared with other cognitive architectures, ACT-R has a number of benefits.

Firstly, it has an extensive theoretical foundation based on neurobiology and neuroimaging research as well as psychology. Each ACT-R component is designed to correspond to a specific area of the human brain, the function of which has been experimentally tested using methods such as functional magnetic resonance imaging (fMRI). In ACT-R, elements of declarative knowledge are named chunk which represent WMEs (for working memory elements) and the chunk's production utilities are computed and used in conflict resolution.

Furthermore, ACT-R has been used successfully to create models in many areas of academic research [11]. It has been applied in education, human-computer interaction, and especially in tactical training to create computer-generated fighting forces [12]. Numerous international institutions in the U.S. have adopted ACT-R to explore cognitive behavior in humans, including the National Aeronautics and Space Administration, the Defense Advanced Research Projects Agency, and the National Institute of Mental Health at the National Institutes of Health.

Other advantages of ACT-R include high-quality software documentation, tutorials, and sample models, as well as readily available training and support. The user interface is highly interactive, easy to use, and allows for detailed data analysis. The open source architecture enables additional modules to be added easily.

Overall, ACT-R provides an excellent tool for exploring models of human performance and an excellent platform for analyzing goals, operators, methods, and selection (GOMS) rules in complex systems [13].

2.2 ACT-R Cognitive Modeling

The ACT-R–based cognitive modeling is intended to mirror the real cognitive process by careful construction of both the task and the model's parameters. While the model is running, a monitor displays each of the cognitive behaviors being simulated as well as their status, including progress to the goal, knowledge retrieval, production rule execution, conflict resolution processes, and motor actions. The most important work of cognitive modeling so far is extraction for declarative knowledge, procedural knowledge and model's parameter. The declarative knowledge is some conceptions definition for the task or some facts, such as the operations, the vehicle's status and the relationship between crosshair-cursor's size and vehicle's distance etc. for manual control RvD task. The procedural knowledge is lots of rules for decision-making of the model.

To validate the model, the simulated cognitive behavior and experimental results were compared with those of actual manual tasks. Additionally, the ACT-R control panel tool was used to analyze how well the model reproduced the real manual control tasks. The results of this comparison indicated that the model's processes and results were representative of actual human cognitive behavior exhibited during real-world control tasks.

2.3 ACT-R Model Based Human Performance Evaluation

By running cognitive behavior model, the operator's cognitive produce and control behavior was reproduced. We can examine the model output and efficiency to assess the model's performance. To get best effects, sometimes we take measures of changing the model's parameters. By comparing the model running result and true human operations, the model's effectiveness was verified. Finally By adjusting model's parameter and comparing model running consequence under different adjusted parameters, the human performance was evaluated according to certain criteria.

Due to the ACT-R's design has implemented its each module mapping to human's brain functional regions, and ACT-R can show and record each buffer's activities status corresponding to irrelevance module at each cognitive period, each cognition procedure and its effect on human performance can be calculated and mapped to certain cognitive function part. By using visualization methods, the human performance can be displayed and special task's cognitive behavior procedure can be simulated by computer software.

By combination of related influence of influencing factors and according to the factor's weight in human performance, human performance assessment report was generated. Furthermore, we could adjust the various model parameters to simulate a variety of conditions and predict the expected results of human operators. Thus, we could test each variable and assign a weight to each variable's potential influence on human performance.

3 A Platform on Human Performance Evaluation

To perform more sophisticated studies on human performance during complex spaceflight tasks, An ACT-R cognitive architecture based platform on human performance evaluation for space manual control task is designed and implemented. The platform is set of analysis software tools to evaluate human performance for on-orbit astronaut's control tasks. It is created with a three main functional cluster that included cognitive modeling and model evaluation cluster, model-based human performance evaluation cluster, and model fusion and cognition visualization cluster. Its function include modeling and simulation for space controlling task's cognitive behavior, biomechanical simulating for controlling operations, human performance evaluation and analysis, 3D visualization for cognitive behavior, multiply model's integration, database interface management and communication, and resource schedule management for whole system. Figure 1 shows the framework of the cognitive architecture based platform for human performance evaluation of spaceflight manual control task.

3.1 Cognitive Modeling and Model Evaluation Cluster

Cognitive modeling and model evaluation cluster are the fundamental part of the platform, and it encompasses cognitive behavior modeling module, model's verification and evaluation module, and cognitive model libraries.

Cognitive behavior modeling module conduct declarative knowledge's definition by analyzing experimental record and empirical knowledge; model's procedural knowledge are constructed through extracting relationship between cognitive process and manual control behavior, and model's parameters are refined according to boundary conditions and task's specific characteristics. Based on declarative knowledge, procedural knowledge, and model's parameters, by adhering to ACT-R programming paradigm the specific control task's cognitive model is built.

Model's verification and evaluation module is to verify the model's validation and how the cognitive model fitting the human cognitive procedure. The model's validation is verified from two different stages. The first stage is the model can

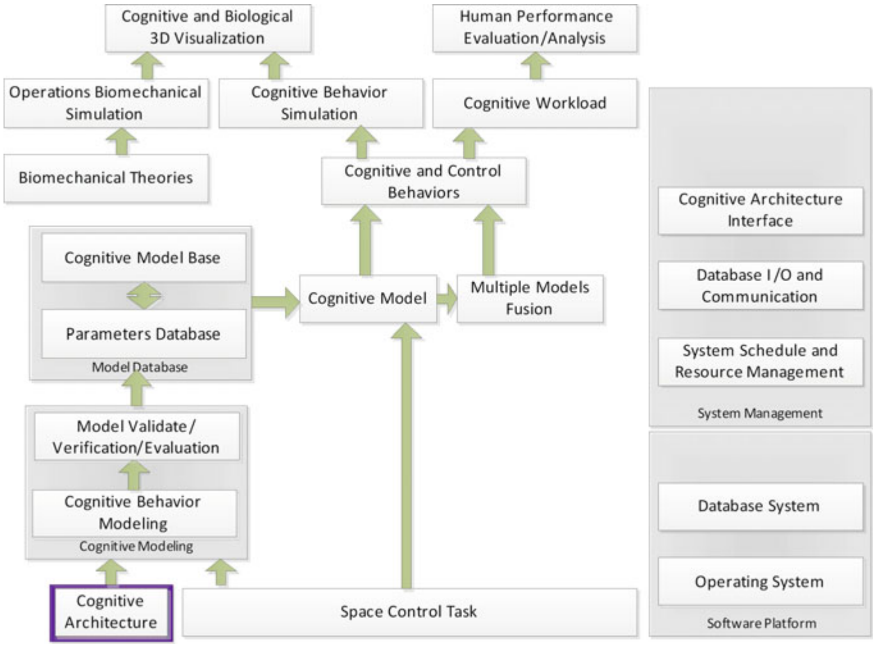


Fig. 1 Cognitive architecture based platform on human performance evaluation for spaceflight manual control task

successfully complete the model's task which model is built based on. The second stage is the model's every single operation while it running coincides with actual cognitive behavior which it appears in experiment in details. By running different cognitive behavior model, the cognitive process and control behavior will be redisplayed accordingly. By comparing the process and results between model running and actual control process the model's validation is verified in two aspects of model accomplishing control task and cognitive behavior details.

Cognitive model libraries are model bank for different kind of spaceflight manual control tasks or even various models for same task. While these models are applying to task performance evaluation, this makes sure that the best suitable model would be chosen for certain task.

3.2 Model-Based Human Performance Evaluation Cluster

In ACT-R, the buffer trace module records the actions which occur through the buffers while a model runs and can report that information in a text display or return a list of Lisp structures which encode the actions that took place. By calculating irrelevance buffer's active scale, the corresponding brain region's active time can be obtained. By doing so, the brain's workload can be counted.

This cluster is composed main work of task analysis, model selecting and modeling and performance evaluation. By task analysis, the target task is classified one or several control task with which there are models in cognitive model libraries. The goal of model selecting and modeling is obviously to select model from model libraries to fit current classified task, and modeling work is to implement cognitive modeling for target task. By adjusting model's parameter and comparing model running consequence under different adjusted parameter, the human performance is evaluated according to certain judging criteria. By combination of related influence of influencing factors and according to the influencing factor's weight in human performance, construct the human performance assessment method and standard, and human performance assessment report will be generated according to design's purpose.

3.3 Model Fusion and Cognition/Behavior Visualization Cluster

This cluster is the high functional module for human performance evaluation. It aims to show human's cognitive process picturesquely. Due to the cognitive architecture vividly portrays human's cognitive process which includes vision perception, information processing, knowledge retrieve, conflict resolution, rules trigger and motor generation, by computer simulation technics the human's cognition/behavior as they implement task can be animated on the computer screen. Model fusion is designed to integrate model for the purpose of making the model more suitable for target task. The cognition/behavior visualization module is to make human's cognitive behavior and their actions redisplay on the computer monitor by using animation. For this purpose human biomechanical motor model for certain operations should be built up.

3.4 System Schedule and Resource Manage Cluster

This cluster in the platform makes sure that the whole system can be run smoothly. It contains system's run-time environment, resource management and schedule, Database I/O and communication modules, and system maintenance interface etc.

The run-time environment describes the software responsible for actually running the platform. Resource management is in charge of resource allocation and recycling, and schedule is responsible for calling of other modules. Database I/O bring about the interface function to access database effectively. Communication modules conduct data or information exchange between modules. And system maintenance interface realize information's maintenance (i.e. creating, retrieving, updating, and deleting information) and parameter's configuration (i.e. defining, setting-up, or modifying the model's parameter) function.

4 Manual RvD Human Performance Evaluation

As a study case, spacecraft manual control RvD task is chosen to investigate human performance during spaceflight mission. To build cognitive model, Firstly, extract model's declarative knowledge by experimental methods and empirical knowledge, and then construct model's procedural knowledge through analysis of relationship between cognitive process and control behavior, after that refine model's parameters according to boundary conditions and task's specificity. Based on declarative knowledge, procedural knowledge and model's parameters Manual RvD cognitive behavior model is built, and model's validation is verified by comparing the process and results between model's running and actual control.

The following sections illustrate the main step for performance evaluation for manual RvD control task.

4.1 ACT-R Based Cognitive Workload for RvD Control Task

To evaluate the performance of manual RvD control task, run the RvD model and record actions which occur through the buffers while a model runs. Figure 2 is the ACT-R buffer trace using manual RvD control task cognitive mode:

In Fig. 2, chunk name indicates there is a new chunk set in the buffer, request between two "+" characters indicates there is a request, a series of "=" characters indicates the buffer is modified, a series of "−" characters indicates the buffer is cleared, a series of "*" characters indicates the module is busy, and if it is filled with spaces it indicates other case.

4.2 ACT-R Based Human Performance Evaluation for Manual RvD Task

To some extent, the tracing buffer is busy shows that brain region is working. We assume that there is workload in that region. By counting the total active time for different tracing buffer, each total active time of all tracing buffer will be obtained. The total active time of that buffer can be taken as the hint for that workload. Figure 3 Show the workload of manual RvD task tracing buffer for four skillful grade (skilled, less skilled, unskilled and novices).

```
         |    PRODUCTION     |      GOAL       | VISUAL-LOCATION |      VISUAL      |      MANUAL      |
   0.000 |  +    START    + | .   ACC-GOAL    | .VISUAL-LOCATIO |                  |                  |
   0.025 |  ************** | .               | .               |                  |                  |
GOAL REQUEST VISUAL-LOCATION
   0.050 |  + FIND-DRONE + | .   =======     | .VISUAL-LOCATIO |                  |                  |
   0.075 |  ************** | .               | .               |                  |                  |
FIND DRONE LOCATION
   0.100 |  ************** | .   =======     | .    -------    | +MOVE-ATTENTI+   |                  |
   0.125 |                 | .               |                 | **************   |                  |
   0.150 |                 | .               |                 | **************   |                  |
   0.175 |                 | .               |                 | **************   |                  |
   0.185 |  +ATTENDED-DRO+ | .               |                 | .    DRONE0      |                  |
   0.210 |  ************** | .               |                 | .                |                  |
   0.235 |  + CHOOSE-AXIS+ | .   =======     | .VISUAL-LOCATIO | .    DRONE1      |                  |
   0.260 |  ************** | .               | .               | .**************  |                  |
THE DRONE ROLL IS -3.022526e-5 YAW IS 0.0014250036 PITCH IS 9.175262e-4 AXIS Y
#|Warning: Creating chunk Y of default type chunk |#
   0.285 |  +CONTROL-MOVE+ | .   =======     | .               | .**************  |                  |
   0.310 |  ************** | .               | .               | .**************  |                  |
PRESS KEY "d"
   0.335 |  + CHOOSE-AXIS+ | .   =======     | .               | .**************  | +  PRESS-KEY +   |
   0.360 |  ************** | .               | .               | .**************  | **************   |
THE DRONE ROLL IS -3.022526e-5 YAW IS 0.0014250036 PITCH IS 9.175262e-4 AXIS Y
   0.385 |  ************** | .   =======     | .               | .**************  | **************   |
   0.410 |                 | .               | .               | .**************  | **************   |
   0.435 |                 | .               | .               | .**************  | **************   |
   0.460 |                 | .               | .               | .**************  | **************   |
   0.485 |                 | .               | .               | .**************  | **************   |
   0.510 |                 | .               | .               | .**************  | **************   |
   0.535 |                 | .               | .               | .**************  | **************   |
   0.545 |                 | .               | .               | .**************  | **************   |
   0.570 |                 | .               | .               | .**************  | **************   |
   0.595 |                 | .               | .               | .**************  | **************   |
   0.620 |                 | .               | .               | .**************  | **************   |
   0.635 |  +CONTROL-MOVE+ | .               | .               | .**************  |                  |
   0.645 |  ************** | .               | .VISUAL-LOCATIO | .    DRONE2      |                  |
```

Fig. 2 ACT-R buffer trace using manual RvD control task cognitive model

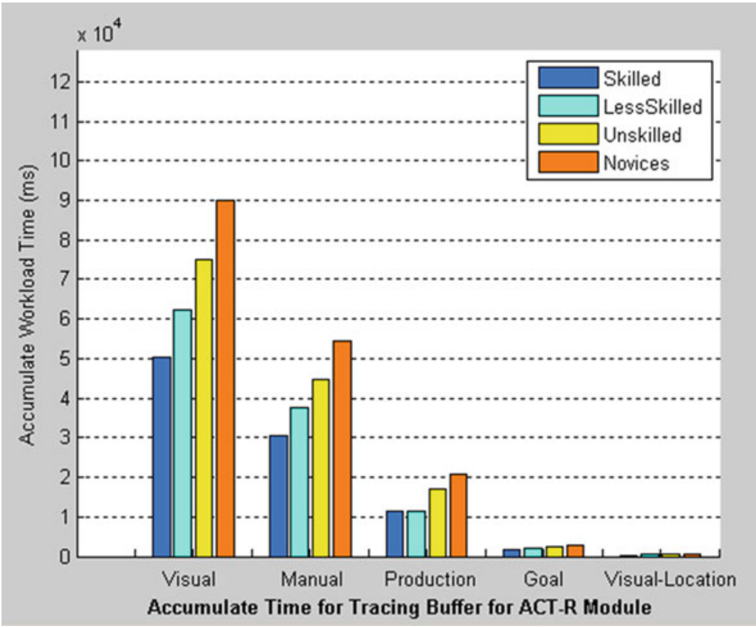


Fig. 3 Workload of manual RvD task tracing buffer for different skillful grades

From Fig. 3, it shows that there are different operational performance for different skillful subjects. That is to say that the skillful parameter in ACT-R model can represent operator's skillful grade and can be mapping skillful characteristic.

4.3 ACT-R Based Cognitive Workload Mapping on Human' Brain Region

Each module in ACT–R is designed to associate with distinct cortical regions. The goal buffer is associated with the dorsolateral prefrontal cortex (DLPFC), the retrieval buffer is associated with the ventrolateral prefrontal cortex (VLPFC), the manual buffer is associated with the adjacent motor and somatosensory cortical areas, and one of the visual buffers is associated with the dorsal "where" path of the visual system, while the other, associated with the ventral "what" system. The predefined mapping of modules on brain regions was based on a reading of the

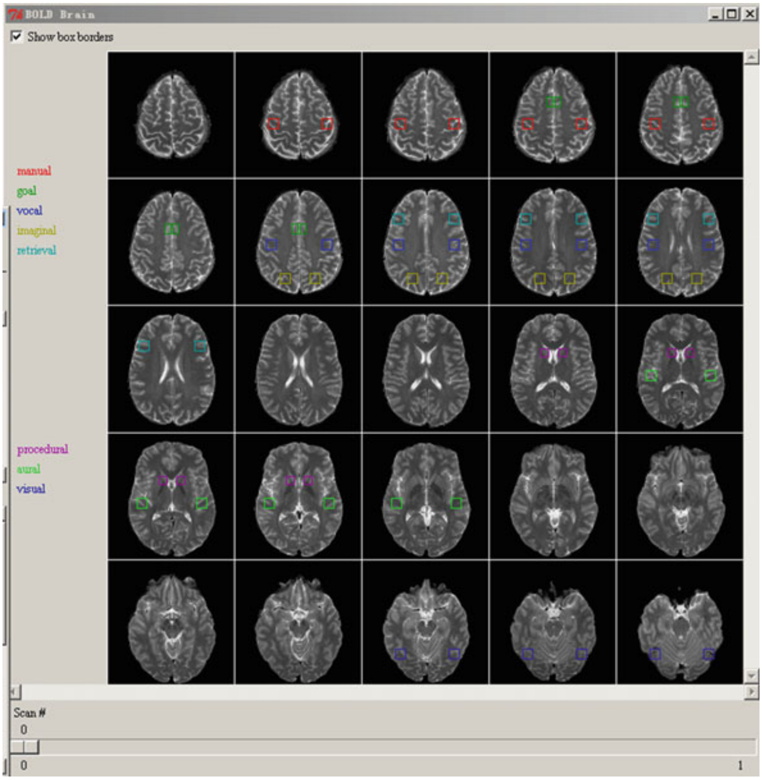


Fig. 4 Mapping buffer activity on the human brain region according to BOLD response values from comparison to fMRI data for manual RvD control task

literature, and even more recently research show evidences for these associations five ACT-R modules by fMRI (functional magnetic resonance imaging) data-driven mapping [14].

In ACT-R, the BOLD module provides the modeler with a command for predicting blood-oxygen-level dependent (BOLD) response values from the activity of the buffers for comparison to fMRI data. As the model running, the "BOLD Brain" window shows image slices of a reference brain and displays the BOLD data values for the buffers as color coded boxes in the images in the areas with which the corresponding buffer has been associated. Figure 4 is buffer activity mapping on the human brain region according to BOLD response values from comparison to fMRI data for manual RvD control task.

The BOLD module also provides the modeler 3D picture for mapping tracing buffer's activity on corresponding human brain region. The modeler can zoom in and zoom out the image, or move the brain rotation only by mouse operation.

5 Conclusion

To explore sophisticated studies on cognitive behavior in complex spaceflight tasks, the paper proposes an integrated research and development platform for investigating astronaut's cognitive behavior and performance evaluation on the basis of ACT-R cognitive architecture. The components of the platform are detailed on the base of ACT-R's features. As a sample human performance for different skillful grade operator with manual RvD control task are examined. The cognitive workload are mapping to human' certain brain functional region.

This paper's main job is two-folded. The one is it proposes and implements a performance assessment and simulation platform for space manipulation task. The other one is manual RvD task's cognitive performance is evaluated and the workload is simulated, and model's skillful parameter can map human cognitive characteristic.

Acknowledgments This work is supported by Zhejiang Provincial Natural Science Foundation under Grant No. LY12C09005, Y1110477, National Natural Science Foundation of China under Grant No. 61100183, 6110503 and 973 Program of China under Grant No. 2011CB711000.

References

1. Zhang, S.Y., Tian, Y., Wang, C.H., Huang, S.P., Fu, Y., Chen, S.G.: Modeling human control strategies in simulated RVD tasks through the time-fuel optimal control model. In: Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management, pp. 661–670. Springer International Publishing (2014)
2. Jones, P.M.: Reviews of Human Factors and Ergonomics 6, pp. 172. Sage Press, California (2010)

3. Aeronautics, National, Administration, Space: Bioastronautics Roadmap: A Risk Reduction Strategy for Human Space Exploration (NASA Special Publication SP-2004-6113). National Aeronautics and Space Administration, Washington, DC (2005)
4. Kanas, N., Manzey, D.: Space Psychology and Psychiatry, 2nd edn. Space Technology Library. Microcosm Press, Springer, New York (2008)
5. National Aeronautics and Space Administration. Human Research Program. http://humanresearch.jsc.nasa.gov
6. Liu, Y.F., Tian, Z.Q., Zhang, Y., Sun, Q., Li, J.S., Sun, J., Fu, F.: COMPAss: a space cognitive behavior modeling and performance assessment platform. In: Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management, pp. 630–636. Springer International Publishing, Switzerland (2014)
7. Taatgen, N., Anderson, J.R.: The past, present, and future of cognitive architectures. Top. Cogn. Sci. **2**(4), 693–704 (2010)
8. Fleetwood, M.D., Byrne, M.D.: Modeling the visual search of displays: a revised ACT-R model of icon search based on eye-tracking data. Human-Comput. Interact. **21**(2), 153–197 (2006)
9. Farkaš, I.: Indispensability of computational modeling in cognitive science. J. Cogn. Sci. **13** (12), 401–435 (2012)
10. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychol. Rev. **111**(4), 1036 (2004)
11. Best, B.J., Lebiere, C., Scarpinatto, K.C.: Modeling synthetic opponents in MOUT training simulations using the ACT-R cognitive architecture. In: Proceedings of the 11th Conference On Computer Generated Forces And Behavior Representation, vol. 7, p. 33 (2002)
12. Dimperio, E., Gunzelmann, G., Harris, J.: An initial evaluation of a cognitive model of UAV reconnaissance. In: Proceedings of the Seventeenth Conference on Behavior Representation in Modeling and Simulation, pp. 165–173 (2008)
13. Byrne, M.D.: ACT-R/PM and menu selection: applying a cognitive architecture to HCI. Int. J. Hum Comput Stud. **55**(1), 41–84 (2001)
14. Borst, J.P., Menno, N., Taatgen, N.A., van Rijn, H., Anderson, J.R.: Using data-driven model-brain mappings to constrain formal models of cognition. PLoS ONE **10**(3), e0119673 (2015)

Multi-level Cognitive Cybernetics in Human Factors

Daniel N. Cassenti, Katherine R. Gamble and Jonathan Z. Bakdash

Abstract Cybernetics provides a framework for understanding the behavior of closed-loop systems, including the feedback control intrinsic to cognitive systems (Smith and Smith in continuing the conversation: a newsletter of ideas in cybernetics. Greg and Pat Williams, Gravel Switch, KY, [1]). We propose adopting our interpretation of the cybernetics concept of feedback control of cognition by integrating across metacognition, performance, computational cognitive modeling, and physiological levels of analysis. To accomplish this objective, we tie cognitive variables to each level of analysis, including: (1) metacognition—self-evaluation of cognition; (2) performance—objective measures of progress toward a goal state; (3) physiology—indications of cognitive function (e.g., heart rate variability as an index of the level of task engagement); and (4) cognitive models—prediction and understanding of empirical results using sequences of cognitive steps. We call this integrative approach, Multi-Level Cognitive Cybernetics (MLCC). In this paper, we define the MLCC framework, discuss how MLCC can inform the design of adaptive automation technologies, and discuss the benefits of the MLCC approach in human factors.

Keywords Cybernetics · Cognitive psychology · Human factors · Computational modeling · Research methods

D.N. Cassenti (✉) · K.R. Gamble · J.Z. Bakdash
Human Research and Engineering Directorate, U.S. Army Research Laboratory,
459 Mulberry Point Rd., Aberdeen Proving Ground, Adelphi, MD 21005, USA
e-mail: daniel.n.cassenti.civ@mail.mil

K.R. Gamble
e-mail: katherine.r.gamble.ctr@mail.mil

J.Z. Bakdash
e-mail: jonathan.z.bakdash.civ@mail.mil

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_27

1 Introduction to Cybernetics

In the 1940s, Norbert Wiener's conception of cybernetics [1, 2] formalized a new research field for systems operating in feedback loops. His ideas influenced a multitude of disciplines, including the physical sciences, mathematics, life sciences, and social sciences [3]. Below, we briefly describe the history and foundations of cybernetics and focus on the field's relationship to the study of human behavior.

1.1 History

Wiener [2] defined cybernetics as the study of, "control and communication in the animal and machine." In his conceptualization, animals (including humans) and feedback-controlled machines are examples of systems that interact with each other and within the larger environment. He coined the term for this field from the Greek word for "steersman," inferring that we make adjustments to goal-directed behavior based on feedback from the environment.

1.2 Basic Concepts of Cybernetics

Communication is a theme of cybernetics. Though the term communication is broadly defined, cybernetics operationalizes communication as the processes by which a system, whether animal or machine, controls its interactions with the environment through the use of information. For example, an animal's motor behavior changes a variable in the environment. The animal's sensory organs receive the information about the altered environment. This communication takes place within closed-feedback loops; the system that initiated an action (whether originating with the organism or environment) affects the receiving system, which alters its output and then influences the originating system again [2].

An example of this closed-loop process is a warm-blooded animal's internal thermostat [4] controlled by the hypothalamus, which triggers heating (e.g., shivering) or cooling (e.g., sweating) at certain temperatures. As the temperature changes, the internal thermometer registers the change and stops heating or cooling when it gets past some threshold, thus the adaptive mechanism uses feedback to control itself.

1.3 Human Behavior and Cybernetics

There has always been some focus on human behavior in the cybernetics community, and many cybernetics efforts placed strong emphasis on how humans interface with the environment using technology [3]. However, this emphasis on technology was perceived by some researchers to mean that the field was too far mechanistic and precluded the concepts of autonomy and agency, for example. In response, cyberneticists began a new movement in second-order cybernetics [see 5].

According to second-order cybernetics [5], the scientist, by virtue of observation, becomes part of the closed-feedback loops of organisms (humans) to environment. That is, the observer affects the environment and the environment affects observer. The perspective of second order cybernetics was less mechanistic than the initial approach of the discipline "… by emphasizing autonomy, self-organization, cognition, and the role of the observer in modeling a system" [5].

2 Introduction to Multi-level Cognitive Cybernetics

Multi-Level Cognitive Cybernetics (MLCC) is a research approach based on concepts from cybernetics and cognitive psychology, which seeks to determine the optimal set of human behavioral triggers for adaptive automation that leads to the most effective relationship of user and technology in reaching a shared goal state. MLCC integrates metacognitive, performance, physiological, and computational (modeling) levels of analysis within closed feedback loops. The MLCC approach includes running empirical studies with a number of human behavior measures for data collection and levels for analyzing these data.

2.1 Cognitive Aspects of MLCC

Cybernetics is relevant to the study of cognition because both have a focus on the role of feedback in shaping human behavior [6]. For example, the processes of working toward a goal and adapting to changing environments [1] are central to cognitive science. In MLCC, the user and technology (i.e., the environment or the interface with the environment) work together to meet a common goal. To optimize goal or task completion, human behavior can be used as input to help technology adapt, while the technology's adaptive feedback may help improve the user's performance in a circular-causal relationship.

The first major movement for including closed-feedback loops in cognitive psychology came from Gibson's ecological perception [7]. Neisser's perception-action cycle (PAC) [8] also modeled the closed loop connecting the human with the environment, but without the direct perception component of ecological perception, instead discussing how information is processed. This made closed loop an important part of studying cognition and therefore opened the door to MLCC.

This adaptive closed feedback loop can be broken down into two subsystems (Fig. 1). The cognitive (or human) side has multiple stages, comprised of: (1) *perception* to process the sensations from the technology; (2) *information processing* to mentally transform perception into goal-oriented options; (3) *decision making* to select among these options; and (4) *motor control* to implement the chosen solution (see [9]). Technology also has four stages, including: (1) *input stream* designed to receive the information sent by the user; (2) *digital computation* to transform the input into meaningful units; (3) *user behavior assessment* to determine if adaptive automation must be triggered; and (4) triggering or not triggering *adaptive automation* algorithms.

2.2 Similarities and Differences with Other Approaches

MLCC is related to Human Computer Collaboration (HCC; [10]). HCC investigates how to integrate user and technology in pursuit of joint goals to the point where a human plus a computer outperforms either alone (e.g., weather forecasting). MLCC will study human-system collaboration in a more multi-faceted manner, by incorporating multiple levels of input (human feedback), such as performance and physiology, rather than only overt behavior (e.g., button presses).

MLCC also draws from the Ecological Interface Design (EID; [11]) approach. EID is built on the assumption that a technological interface directly affects a user's cognitive processes, and thus an interface can be adjusted to reduce high cognitive demands to improve task performance. As with HCC, MLCC goes beyond EID by examining more than the perceptual stage of cognitive processing. However, Vicente's system of Cognitive Work Analysis (CWA; [12]), a tool common for EID, which may aid researchers employing MLCC to determine which stages of processing need adaptive automation.

2.3 Levels of Human Behavior Analysis

For adaptive automation to be effective, the technology must know when to trigger automation [13]. While most studies focus on external triggers, we suggest that, as shown in Fig. 1, triggers based on cognitive variables can be sufficient triggers for automation.

There are methods at each level of analysis identified that can be used to assess user state and to trigger automation, all of which are related to cognitive psychology (additional detail in 4.1). Here, we describe these methods and the costs and benefits of each. (1) Metacognition: the user assesses their need for an aid. Although

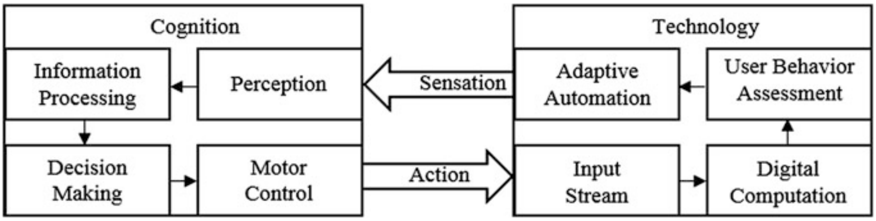


Fig. 1 The stages of cognition and adaptive technology in the MLCC framework

this gives the user control over automation, humans are typically poor at accurately assessing their own cognitive states [14]. (2) Performance: accuracy or response time levels trigger automation. Performance data are objective indicators of when an aid is needed, but do not directly indicate the cognitive variables causing low performance, and therefore what aid might be most helpful. (3) Physiology: measures of internal state fluctuations can trigger automation. Physiology can relate to cognition, such as heart-rate variability indicating level of engagement [15] or pupil diameter indicating mental workload [16], but may have multiple causes. (4) Computational cognitive modeling: post-data-collection simulation of the processes that led to the results. Computational cognitive modeling attempts to explain how phenomena are produced, but it rarely accounts for multiple causative processes as modelers tend to settle into a single, well-fitted iteration of a model rather than seeking alternatives, which could result in inaccuracy. Because each level of MLCC has limitations when used in isolation, the MLCC approach suggests integrating the four levels to enable a comprehensive measure of human behavior that lead to the best trigger of automation and the most optimal feedback loop.

3 Adaptive Automation and Triggering Functions

Human factors can greatly benefit from a focus on closed-feedback loops between the human and environment, which can better inform the design of technology. Behavioral psychologists historically focused on the human's ability to adapt to the environment, but how the environment adapts to the human is equally important to understand.

Automation is intended to enhance *human* performance, often by reducing workload (subjective, physical, and cognitive). However, in nearly all situations, the user must supervise the automation for errors, no matter how infrequent they may be. This can create an even more difficult problem called "automation bias," where users assume the technology's inerrancy and stop paying attention. Alternatively, a user may feel the constant need to monitor automation, a test of vigilance, harming not helping performance [17].

MLCC recognizes that adaptive automation occurs in a closed feedback loop, where there is a reciprocal relationship between a human and technology, as each acts upon and in response to the other. That is, each uses input received from the other to adjust their behavior and, thus, their output back to the other system. Autocorrect in spelling is an example of adaptive automation. Technology offers output, spelling suggestions, to a human, who will in turn provide feedback about what output will be more useful for improved future performance, and the technology will use the human's output to improve its next output.

The following sections describe four different levels in MLCC, which will help designers decide when and what adaptive automation should trigger during task completion. We use two cognitive tasks as scenarios to illustrate the levels: resource allocation (i.e., supplying units with resources in a command and control task) and text processing (i.e., drawing reasoned conclusions from a text feed).

3.1 Metacognition

The simple solution to communicating with technology is via metacognition, or a direct action from the user following deliberate self-assessment of cognitive processing. This may be done in three ways. First, the user could produce an overt behavior, such as pressing a button, to inform the technology that an automated aid is needed, and the technology would then respond by engaging automation. Although, this appears to be a simple process to assess whether adaptive automation should be triggered, metacognition is a complex cognitive process to determine just-in-time activation of an aid, even if it is not difficult for the technology to process a button press. Second, the user may a priori inform the technology, which skills he or she feels less confident performing, and the technology could engage aids that perform those tasks only. Lastly, the user could perform practice trials of the task and assess where aid would be needed, so that the technology would perform those tasks in the future. If the user cannot distinguish assigned supplies from new ones in the resource allocation task, metacognition could isolate this as the problem and engage automation that color codes the used from new supplies.

3.2 Performance Measures

Parasuraman et al. [18] conducted a study using performance thresholds as triggers for adaptive automation. Performance measures of accuracy, reaction time, or both, are used as input to adaptive automation. Feigh et al. [13] recommend the use of performance-based thresholds, which are effective for tasks and domains with well-specified limits. While the use of performance as a trigger is intuitive as an objective measure of success in reaching a goal, poor performance could be explained by multiple factors, and a threshold cannot determine what these factors are. Thus, while performance thresholds can trigger adaptive automation, a direct mapping of why performance faltered would lead to the optimal aids being triggered. For example, if a performance trigger activates a decision making aid in the text processing task, but the deficit is in visual attention, then the aid will not likely improve performance. If the technology integrates both metacognitive and performance triggers, the user could press buttons to engage a text perception aid and disengage a decision-making aid, assuming the user can correctly assess what is needed.

3.3 Physiological Measures

Physiological measures are often associated with different cognitive processes [19], and therefore act as triggers in response to different variables within the MLCC framework. We address five types of physiological measures for MLCC. (1) Heart rate variability (HRV): the range of heart rate fluctuations can be indicative of task engagement [15]. Decreased HRV may indicated a lack of engagement, which could suggest an overreliance on automation, and thus a need to return task control to the user. (2) Eye tracking: a record of eye movements that can measure where attention is focused [20]. The user who is not focusing on the proper portions of the screen may need highlighting automation to aid in the redirection of gaze. (3) Pupil diameter (PD): a wider pupil diameter is indicative of greater mental workload [16]. An increase in PD may turn on automation, while a decrease in PD may return tasks back to the user. (4) Electrodermal activity (EDA): detection of electrical changes in skin conductance. High EDA may indicate increased arousal. The technology may take over or relinquish tasks based on whether arousal levels are too high or too low in order to maximize performance. (5) Electroencephalogram (EEG): recordings of electric activity measured at the scalp. Cassenti et al. [21] found that latency Event Related Potentials (ERPs) could be modeled as the timing of perceptual encoding and memory (i.e., context) updating, respectively. Large latency from the ERPs could indicate whether automation is needed to aid the perception or decision making stages, respectively.

Because physiological variables are related to various cognitive processes, they could lead to specific recommendations for automation. These measures are objective, and therefore great options for automation triggers, though some of them are still in early stages of development for real-time assessment and processing. While some of these technologies continue to develop to where they can become reliable triggers of automation, they can presently be used in post-processing to better understand adaptive automation. In the text processing task, users may have an increased EDA when a lot of information appears in the feed and engage a perceptual aid. However, if the technology integrates the performance level with EDA processing, it may decide that the increased arousal is not hurting performance and avoid distracting the user with the aid.

3.4 Cognitive Modeling

Cognitive modeling is a computational approach for modeling processes of the mind; it can predict performance and may be informative for adaptive automation. Computational modeling is a method of simulating empirical results in a modeling system with better fits between empirical and model results increasing the explanatory power of the model corrected for number of variables. In the case of MLCC, we specifically refer to cognitive modeling. The cognitive stages within the MLCC framework may be isolated as problem areas based on a cognitive model of engagement with the technology, and subsequent adaptive automation employment may be used to target those stages that presented the most difficulty. The following are some examples of adaptive automation targeted (based on explanations from models) to the particular stages from Fig. 1. (1) Perception: highlight areas of importance in the visual interface of the technology. (2) Information processing: Given a textual feed, visually enhance the elements that the technology deems most helpful. (3) Decision making: in a resource allocation task, provide resource management recommendations. (4) Motor control: automate the operation of complex motor behavior, such as marking an area of interest for surveillance.

There are many options for cognitive modeling systems. ACT-R [22] has three primary properties that make it most conducive to the MLCC approach: (1) the goal of both ACT-R [22] and MLCC is to explain human behavior (as opposed to more artificial intelligence oriented modeling systems); (2) ACT-R is the most thoroughly researched cognitive modeling system, (3) ACT-R and MLCC are based on the closed-feedback loop between environment and human. Of the options, we recommend the use of ACT-R. In the resource allocation task, cognitive modeling may indicate that the heaviest mental workload requirements are required in the decision making stage. If both the findings from ACT-R modeling and PD physiological processing are integrated in the technology, increased PD would act as a manipulation check that workload is high and validating the need to deploy a decision making aid.

4 Benefits and Limitations of the New Approach

According to the MLCC approach, empirical studies should be conducted to collect a host of human behavioral measures, and run post-experimental analyses and computational models to determine the human-behavior variables that would most benefit from adaptive automation across a variety of tasks. The objective of this approach is not to collect an extremely large amount of data to find significant results, but instead to: (1) collect data to down select among the levels to determine the optimal triggers for adaptive automation; (2) discover whether multiple measures within or between levels predict each other and optimization of performance (see [23] for statistical guidance on comparing statistical models). This approach could facilitate advancement of adaptive human-technology integration. Here, we discuss the benefits of using MLCC.

4.1 Integrating Human Behavior Measures in Human Factors

Ultimately, MLCC is a research approach through improved communication between user and technology in their joint pursuit of goals. This requires the technology to sense and process user behavior to determine whether and what adaptive automation to trigger. The results from studies that integrate the different levels within the MLCC approach will determine what measures are most important to collect. We predict that in many cases, multiple levels used together will help adaptation between technology and user, thus allowing the human factors developer to integrate several sensory capabilities into the technology to aid various cognitive stages. The strength of the MLCC approach is in the integration of these levels, such that each can provide support or scaffolding when others are less efficient or could be improved. The methods of determining how triggering of adaptive technology leads to optimal performance fall into two categories within MLCC. First, metacognitive (i.e., deliberate engagement of automation and self-assessment of skills) and performance-based (i.e., accuracy and response time) measures can be used to engage automation in real time when needed. Second, due to the current limitations in technologies, the physiological recordings described above and computational modeling can be analyzed using post-processing, but not necessarily real-time processing, with the hope that future developments will allow real-time processing, as well.

MLCC focuses on the data that the human user produces, which allows technology to adapt along with the human, who can continuously learn as they engage with the technology. By integrating the four levels of analysis proposed above into MLCC, we offer a new approach to human factors design research that can still be effective if one or more levels are found not be applicable to triggering adaptive automation. If this is the case, the designer can lean on other levels to aid design.

4.2 Improved Feedback Process

Before the advent of human factors, designers and engineers typically created their systems with little consideration of how well the technology fit the human user. Instead, they relied on the human to learn (i.e., adapt to) the technology and in many cases still do today. Human factors emerged as a field because of accidents in aviation [24], which stressed the enormous cost of neglecting psychological needs in the design of technology. Early human factors research rested on designing technology that fit human needs, but the technology resulting from early human factors applications did not necessarily adapt itself as the user performed a task. Technological design took another step forward with the rise of adaptive automation research, which considered how technology could adapt to a human. We hope that MLCC will strengthen human factors research by using cybernetics concepts to improve how feedback from the human to technology triggers the appropriate adaptive automation at the right time to optimize performance.

By Smith's [1] conceptualization of cybernetics, systems (e.g., technology and the user) can be mutually adaptive to one another in the pursuit of a goal. In MLCC, we consider both the adaptation that the human use is capable of, as well as advancing closer to the point where technology is just as adaptive as the user. With a focus on sensing human behavior and using this input to drive technology design toward the optimization of performance, MLCC provides more possibilities for feedback to technology and allows for concurrent adaptation of systems in pursuit of goals.

4.3 An Approach to Improve Design of Technologies

Empirical research that follows the MLCC approach will collect an array of data, including metacognitive, performance, and physiological measures to determine which levels, individually or together, improve triggering of adaptive automation. Analyses will indicate which of these measures are most helpful in providing critical feedback for a given technology. The technology will, in turn, need to be designed with the means to sense and process these data. If physiological variables are deemed important to improving feedback from the user, the technology must be designed with the capabilities to measure, transfer, and interpret the specific type of data. For example, measures of HRV would require the use of a heart rate monitor that can transmit data in real-time. Technology that includes wearable physiological sensors could be adapted into design.

4.4 Improved Capabilities of Human-Technology Pairings

As a result of applying MLCC, the user and the technology will be two parts of a closed feedback loop, such that they will not work in isolation, but rather will adapt and improve together. The overall objective of MLCC is therefore to rely on both human and technology adaptation to optimize performance. The focus on how to accomplish this objective is an understanding of which variables within the MLCC framework are most conducive to achieving goals with optimal performance.

Theoretically, any of the levels in the MLCC structure can determine which stages of the cognitive process are most in need of adaptive automation. Practically, with the exception of EEG and possibly eye tracking, we believe none of the suggested measures can isolate stage of cognitive processing by current research perspectives. The remaining measures can determine that adaptive automation is needed, but not necessarily what stage of processing the aid should help. Computational modeling takes on added importance for determining the stage of processing where a user has difficulty, and therefore what type of adaptive automation should be deployed, as cognitive modeling relies explicitly on chained cognitive steps (i.e., productions) to reach a goal state. With modeling providing guidance on cognitive stages, adaptive automation may be applied to aid the cognitive stage that is most in need of help, thus users and technology will have more optimal capabilities to achieve goals.

4.5 Limitations

Integration across multiple levels of analysis presents several challenges. Different levels of analysis occur on timescales varying by orders of magnitude and may be weakly related to each other. This poses difficulties in data analysis and interpretation of results, and potentially informing appropriate adaptive automation triggers. Nevertheless, the strength, or lack thereof, among multiple levels of analysis and performance may indicate tasks where adaptive automation is unlikely to be effective.

4.6 Conclusions

The MLCC approach cites basic cybernetics concepts to formulate an approach to adaptive automation in human factors. As the field of human factors moves toward mutual adaptation rather than just human adaptation, MLCC represents a step forward. With the integration of multiple levels of cognitive methods, human factors designers can use the best information for optimizing performance and goal completion.

References

1. Smith, T.J., Smith, K.U.: The cybernetics basis of human behavior and performance. In: Continuing the Conversation: A Newsletter of Ideas in Cybernetics, pp. 1–28. Greg and Pat Williams, Gravel Switch, KY (1988)
2. Wiener, N.: Cybernetics or Control and Communication in the Animal and the Machine. Wiley, New York (1948)
3. Klein, R.R.: The Cybernetics Moment: Or Why We Call Our Age the Information Age. Johns Hopkins University Press, Baltimore (2015)
4. Ashby, W.R.: An Introduction to Cybernetics. Wiley, New York (1956)

5. Heylighen, F., Joslyn, C.: Cybernetics and Second Order Cybernetics. Ency. Phys. Sci. Tech. **4**, 155–170 (2001)
6. Simon, H.A.: The Sciences of the Artificial, vol. 136. MIT press, Cambridge, MA (1996)
7. Neisser, U.: Cognition and Reality: Principles and Implications of Cognitive Psychology. WH Freeman, New York (1976)
8. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
9. Kaber, D.B., Wright, M.C., Prinzel, L.J., Clamann, M.P.: Adaptive automation of human-machine system information-processing functions. Hum. Fact. **47**, 730–741 (2005)
10. Terveen, L.G.: Overview of human-computer collaboration. Know.-Bas. Sys. **8**, 67–81 (1995)
11. Vicente, K.J., Rasmussen, J.: Ecological interface design: theoretical foundations. IEEE Trans. Syst. Man Cyber **22**, 589–606 (1992)
12. Vicente, K.J.: Cognitive Work Analysis. Erlbaum, Mahwah, NJ (1998)
13. Feigh, K.M., Dorneich, M.C., Hayes, C.C.: Toward a characterization of adaptive systems: a framework for researchers and system designers. Hum. Fact. **54**, 1008–1024 (2012)
14. Carlson, R.A.: Experienced cognition. Erlbaum, Mahwah, NJ (1997)
15. Segerstrom, S.C., Nes, L.S.: Heart rate variability reflects self-regulatory strength, effort, and fatigue. Psych. Sci. **18**, 275–281 (2007)
16. Recarte, M.A., Nunes, L.M.: Effects of verbal and spatial-imagery tasks on eye fixations while driving. J. Exp. Psy. App. **6**, 31–43 (2000)
17. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse. Abuse. Hum. Fact. **39**, 230–253 (1997)
18. Parasuraman, R., Barnes, M., Cosenzo, K., Mulgund, S.: Adaptive automation for human-robot teaming in future command and control systems. Technical Report, Army Research Laboratory, USA (2007)
19. Byrne, E.A., Parasuraman, R.: Psychophysiology and adaptive automation. Biol. Psy. **42**, 249–268 (1996)
20. Duchowski, A.: Eye Tracking Methodology: Theory and Practice, vol. 373. Springer, New York (2007)
21. Cassenti, D.N., Kerick, S.E., McDowell, K.: Observing and modeling cognitive events through event-related potentials and ACT-R. Cog. Sys. Res. **12**, 56–65 (2011)
22. Anderson, J.R., Lebiere, C.: The Atomic Components of Thought. Erlbaum, Mahwah, NJ (1998)
23. Shappell, S.A., Wiegmann, D.A.: A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System. Ashgate Publishing, Burlington, VT (2012)
24. Burnham, K.P., Anderson, D.R.: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag, New York (2003)

Accessibility Evaluation: Manual Development and Tool Selection for Evaluating Accessibility of E-Textbooks

Yu Ting Sun, Aaron K.M. Manabat, Mei Ling Chan, Isis Chong and Kim-Phuong L. Vu

Abstract The growing availability of digital learning materials and their integration in many facets of the education system have created a need for evaluating the accessibility and usability of e-learning materials. In some cases, these digital resources are simple conversions of the original, printed materials into electronic formats. In other cases, the digital versions of the materials take advantage of the interactive abilities of the online and electronic media to enhance students' learning experiences. Regardless of format, however, the content needs to be accessible. Although accessibility guidelines and accessibility evaluations tools are available to users, there is no comprehensive accessibility evaluation technique to help guide users in selecting the most accessible learning materials. In this study, we surveyed existing accessibility tools, selected a recommended tool set, and created a manual for post-secondary educators, students, and other stakeholders to use for evaluating the accessibility of e-textbooks.

Keywords Accessibility evaluation · Web accessibility · Disability · Assistive technology · E-textbook

Y.T. Sun · A.K.M. Manabat · M.L. Chan · I. Chong · K.-P.L. Vu (✉)
California State University Long Beach, 1250 Bellflower Blvd, Long Beach, CA 90840, USA
e-mail: Kim.vu@csulb.edu

Y.T. Sun
e-mail: Debby130403@gmail.com

A.K.M. Manabat
e-mail: Aaronkristoffer@gmail.com

M.L. Chan
e-mail: Chong.isis.a@gmail.com

I. Chong
e-mail: Chan@meilingchan.com

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_28

1 Introduction

Most universities in the United States are adopting online components in their education curriculum. Many studies have shown that online courses could produce equivalent learning outcomes compared to traditional, face-to-face courses [1]. However, a major benefit from online learning is that it provides greater access than traditional classrooms [2]. For example, if students are able to access their textbooks online and retrieve class information online, they will have convenient access to their learning materials. In addition, students could save money due to the reduced costs associated with producing electronic materials. However, the online course materials must be accessible for the students to be able to maximize their potential use.

Originally, web accessibility was thought to be primarily important for people with disabilities because it allows this group of users to perceive, understand, navigate, and interact with online materials. It has been demonstrated, however, that improving web accessibility not only benefits people with disabilities, but it can also benefit other groups of people, including older users and users accessing information under restrictive conditions or even on mobile devices [3]. Although federal regulations are available for providing minimum accessibility requirements [4], when considering e-learning for post-secondary education, broader accessibility issues need to be considered. First, it is important to determine how accessible a learning material is for individuals who use assistive technologies versus those who do not. Beyond this comparison, measuring the degree of accessibility provided by different e-learning materials is also key. A final issue to consider is whether the optimal accessibility tools identified for one type of format (e.g., HTML) are ideal for use with other formats (e.g., EPUB or PDF). In this study, we take the first step at examining these issues by evaluating the accessibility tools that are available for checking different accessibility components of e-textbooks in different formats.

E-textbooks are usually beneficial for students because, in comparison to printed textbooks, e-textbooks cost less and can be accessed from different locations and from various web-enabled devices (e.g., desktop computer, laptop, tablet, or mobile device). For example, the California school system (California State University, University of California, and California Community College) has enacted legislation to promote the use of e-textbooks, with the "goal of making higher education in California more affordable by providing faculty and students access to free and lower-cost instructional materials" [5]. This legislation supports students' access to free e-textbooks and will likely to increase the number of users who regularly access e-textbooks for their classes. Although EPUB, HTML and PDF are common e-textbooks formats, at the time of this writing, no accessibility tools or guidelines for evaluating the accessibility of e-textbooks across all of these formats have been released.

Current guidelines are available for web-based content, but these guidelines may not be directly applicable to specific e-textbook formats. Several accessibility evaluation tools are available on the market for people to use for conducting

Table 1 Comparisons of candidate non-assistive color and contrast tools based on price and compatibility (yes or no) with different e-textbook formats

Tools/software	Price	EPUB	HTML	PDF
Color contrast analyzer (CCA, stand alone tool)	Free	Yes	Yes	Yes
WebAIM color contrast checker	Free	Yes	Yes	Yes
Google chrome extensions—color contrast analyzer	Free	No	Yes	No
Check my colors	Free	No	Yes	No

accessibility evaluations, but there has been little research illustrating the effectiveness of each of these tools [4]. Moreover, many of these accessibility tools cannot be used in an automatic fashion (i.e., press a button and receive results). Rather, the tools need to be used by human accessibility evaluators and these evaluators have to perform manual checks and make judgements of accessibility based on the outcomes. Most importantly, we are not aware of standard accessibility techniques that available for evaluating e-books in different formats. To fill this gap in the literature, the goal of this project was to develop a methodology for evaluators to determine the level of accessibility of e-textbooks in different formats.

We will present two phases of the present project: (1) accessibility tool review and (2) instruction manual creation. In the first phase, we compiled an inventory of existing tools that were available to evaluate different aspects of e-books, using a SkillsCommons checklist. The SkillsCommons checklist includes a list of 15 checkpoints that evaluators should assess when determining the overall accessibility of a product (refer to Tables 1, 3 and 4, for details about each checkpoint). The SkillsCommons accessibility checkpoints were developed by the MERLOT (Multimedia Educational Resource for Learning and Online Teaching) project, and WCAG 2.0 guidelines. The checklist has been used for evaluating the accessibility of electronic text and media on the web [6]. In the second phase, we describe the methods used to develop accessibility manuals for e-textbook evaluation.

2 Phase One: Accessibility Tool Review

2.1 Method

A total of 10 human factors students were involved in the checkpoint tool/method review and evaluation phase. The students worked in 4 groups of 2–3 students each. Their task was to find existing tools that can be used to evaluate a subset (3–4) of the SkillsCommons accessibility checkpoints (see Fig. 1 for flow model). Some of the checkpoints could be assessed using non-assistive technologies, and others using assistive technologies. We will describe the methods used for each of these technologies separately below. Each group presented their findings to the entire group and two subject matter experts (SMEs). In consultation with the two SMEs, a subset of tools were selected to be used in the evaluation process based on the

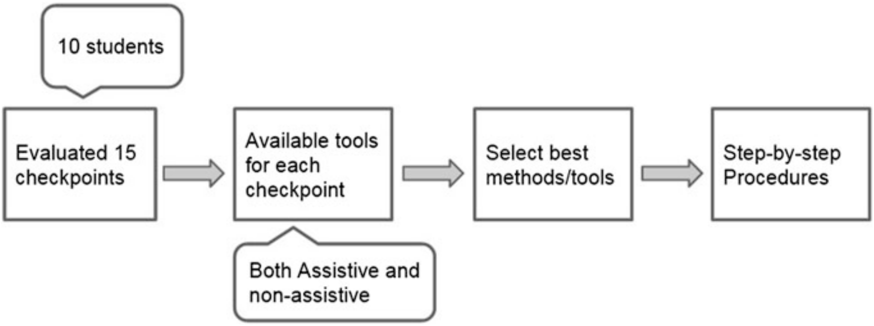


Fig. 1 First stage of manual creation and tool evaluation process

Table 2 Comparisons of four popular assistive technologies' based on format compatibility (yes or no) with different e-textbook formats

Software	Price	EPUB	HTML	PDF
JAWS	$895	Yes (JAW12+)	Yes	Yes (tagged files)
ZOOMTEXT	$595	No	Yes	Yes
Kurzweil 3000	$1495 (Pro, color)	Yes (ver. 13.09+)	Yes	Yes (pro/V5)
NVDA	Free	No (w/adobe digital editions)	Yes	Yes

availability of the software, functionality, pricing, popularity among users, and our subjective assessment of the reliability of the tools.

Non-assistive Technologies. For non-assistive tools, the group found 20 tools. For each checkpoint, we compared the proposed tools based on their functionality and price. Due to the page limitations of this paper, only one example is provided here for the color and contrast checkpoint (see Table 1 for a listing of tools and review criteria). After demonstrating each tool, the group discussed the advantages and disadvantages associated with each tool with the SMEs, and the most suitable tool was selected for each checkpoint. It should be noted that many of the checkpoints needed to be manually checked by the human evaluators.

Assistive Technologies. For checkpoints that required use of assistive technologies, four recommended software tools were compared (see Table 2): JAWS, ZoomText, Kurzweil 3000 and NVDA. These tools were candidates for comparison because they have been identified as the most popular tools that are used by students with disabilities.

JAWS (Job Access With Speech) is a screen reader software that converts text and components of the operating system into synthesized speech. JAWS can read text out loud from the computer screen, and can be used to browse the internet, read electronic books, and other materials. JAWS can also be used in telecommunicating and word processing [7].

ZoomText is mainly a screen magnification tool, but it has several other accessibility features. ZoomText allow you to zoom content from 1× to 36× magnification with eight zoom window types. It also includes features such as color, brightness, and contrast controls, visible pointers and cursors, dual monitor support and application-specific settings. Options for human-sound reading voices and automatic document reading are also available [8].

Kurzweil 3000 allows users to read electronic and printed materials. Kurzweil 3000 is an optical character recognition system that displays printed or electronic text on the computer screen. Text can be read by a speech synthesizer or by the user on the computer screen. Kurzweil 3000 has many options such as a dictionary and spell checker. It can also be used in web browsing with an installation of extension [9].

NVDA (Nonvisual Desktop Access) is an open-source, Windows screen reader that uses eSpeak speech synthesizer, SAPI 4 and SAPI 5 synthesizers. NVDA can be used for browsing the internet, telecommunications, word processing and reading electronic books and other materials [7].

2.2 Results

Non-assistive Technologies. The tools selected for non-assistive technologies are shown in Table 3. All of these tools/software are freely available to the general public. We focused primarily on these tools because free tools/software are more likely to be used by organizations such as educational institutions.

For each software, we analyzed the usability of the tool. For example, both CCA (Color Contrast Analyzer) and WebAIM contrast checker tools can be used across all e-textbook formats; however, CCA was selected as the best tool because it was easy to use. CCA uses an eyedropper tool that an evaluator can use to select a color from the electronic file and the tool will analyze the extracted color. In contrast, WebAIM requires that the evaluator enter the hex color code for specific color (for example, color white is represented as #ffffff). There were other free contrast checker tools that are available on the market, but most of them were not selected as potential candidates because they required special knowledge of color coding similar to WebAIM.

Assistive Technologies. The tools selected for assistive technologies are shown in Table 4.

Kurzweil 3000 was determined to be the most popular assistive software among all students due to its functionality and ease of use. Kurzweil, however, is not available to users for free. After comparing JAWS, ZoomText, Kurzweil and NVDA on their functions, we have decided that Kurzweil and NVDA were the best candidates to evaluate the e-textbooks.

Both Kurzweil and NVDA were tested for specific checkpoints with the different e-textbook formats. Kurzweil was found to be most appropriate for evaluating EPUB formats, and NVDA for HTML and PDF formats. Kurzweil was chosen for

Table 3 Selected non-assistive technology tools/methods to be used is accessibility evaluations for e-textbooks

Checkpoint	Suggested tools and methods		
	EPUB	HTML	PDF
Accessibility documentation	Manual check	Manual check	Manual check
Text access	Adobe digital editions with narrator (windows build in)	Google Chrome—select and speak[a]	Adobe Acrobat XI Pro
Text adjustment	Adobe digital editions	Google Chrome—care your eyes[b]	Adobe Acrobat XI Pro
Reading layout	Adobe digital editions with narrator (windows build in)	Google Chrome	Adobe Acrobat XI Pro
Reading order	N/A (not using assistive tech)	N/A (not using assistive tech)	Adobe Acrobat XI Pro and manual check
Table markup	N/A (not using assistive tech)	N/A (not using assistive tech)	Adobe Acrobat XI Pro and manual check
Structural markup	N/A (source code not available)	N/A (not using assistive tech)	Adobe Acrobat XI Pro
Hyperlinks	Adobe digital editions and manual check	Google Chrome and manual check	Adobe Acrobat XI Pro and manual check
Color and contrast	CCA[c] and manual check	CCA[c] and manual check	CCA[c] and manual check
Language	N/A	Google Chrome. manual check	Adobe Acrobat XI Pro
Images	Pagina EPUB checker and manual check	Google Chrome, W3C and manual check	Adobe Acrobat XI Pro and manual check
Multimedia	Manual check	Manual check	Manual check
Flickering	Adobe digital editions and manual check	Manual check	Adobe Acrobat XI Pro
STEM	Manual check	Manual check	Manual check
Interactive elements	Manual check	Manual check	Manual check

[a]Select and speak—text-to-speech extension for Google Chrome
[b]Care your eyes—webpage color modifier extension for Google Chrome
[c]Color contrast analyzer (CCA)—free accessibility tool to help determine color contrast ratio in all formats against WCAG 2.0 color contrast success criteria

EPUB evaluations due to its popularity among users and the fact that the other tools have issues with the EPUB reader that was employed for evaluation. The free EPUB reader that was used for non-assistive technology evaluations was Adobe Digital Editions and NVDA did not work well with it. Therefore, we decided to use Kurzweil to evaluate books in EPUB format only. NVDA was selected to evaluate HTML and PDF book formats. NVDA is a free tool that allows users to navigate

Table 4 Assistive technology tools/methods suggestions

Checkpoint	Suggested tools and methods		
	EPUB	HTML	PDF
Accessibility documentation	Manual check	Manual check	Manual check
Text access	Kurzweil 3000[a,b]	NVDA and Google Chrome	NVDA and Adobe Acrobat Pro XI
Text adjustment	Kurzweil 3000[a,b]	Google Chrome—care your eyes[b]	NVDA and Adobe Acrobat Pro XI
Reading layout	Kurzweil 3000[a,b]	NVDA and Google Chrome	NVDA and Adobe Acrobat Pro XI
Reading order	Kurzweil 3000[a,b]	NVDA and Google Chrome	NVDA and Adobe Acrobat Pro XI
Table markup	Kurzweil 3000[a,b]	NVDA and Google Chrome	NVDA and Adobe Acrobat Pro XI
Structural markup	N/A (source code not available)	NVDA and Google Chrome	NVDA and Adobe Acrobat Pro XI
Hyperlinks	Kurzweil 3000[a,b] and manual check	NVDA and Google Chrome and Manual check	NVDA and Adobe Acrobat Pro XI and manual check
Color and contrast	CCA[c] and manual check	CCA[c] and manual check	CCA[c] and manual check
Language	Kurzweil 3000[a,b]	NVDA and Google Chrome and manual check	NVDA and Adobe Acrobat Pro XI and manual check
Images	Kurzweil 3000[a,b] and manual check	NVDA and Google Chrome and manual check	NVDA and Adobe Acrobat Pro XI and manual check
Multimedia	Kurzweil 3000[a,b] and manual check	NVDA and Google Chrome and manual check	NVDA and Adobe Acrobat Pro XI and manual check
Flickering	Kurzweil 3000[a,b] and manual check	Google Chrome and manual check	Adobe Acrobat Pro XI and manual check
STEM	Kurzweil 3000[a,b] and manual check	NVDA and Google Chrome and manual check	NVDA and Adobe Acrobat Pro XI and manual check
Interactive elements	Kurzweil 3000[a,b] and manual check	NVDA and Google Chrome and manual check	NVDA and Adobe Acrobat Pro XI and manual check

[a]Select and speak—text-to-speech extension for Google Chrome
[b]Care your eyes—webpage color modifier extension for Google Chrome
[c]Color contrast analyzer (CCA)—free accessibility tool to help determine color contrast ratio in all formats against WCAG 2.0 color contrast success criteria

content with keyboard-only navigation. In addition to these tools, CCA (Color Contrast Analyzer) was selected to evaluate color and contrast across all formats, in both non-assistive technology and assistive technology evaluations, due to its ease of use.

3 Phase Two: Manual Creation

In Phase 2 of this project, we created a manual for each e-textbook format to illustrate how to use the tools selected in Phase 1, and the procedures for additional manual checks that need to be performed when conducting the accessibility evaluations.

3.1 Method

Two students (project leads) re-assessed the selected tools identified from Phase 1 and verified with the SMEs that the tools/methods to use for evaluations were indeed the ideal candidates. Subsequently, a step-by-step manual was created for each format of the e-textbooks, and pilot testing was conducted to make sure these manuals were comprehended by novice evaluators (see Fig. 2).

3.2 Results

We made minor adjustments relating to clarity to the manual for each e-textbook format based on the feedback from the novice evaluators. We also created an Excel checklist that provides a detailed breakdown of the SkillsCommons checklist and a recommended number of pages of the e-textbook to be checked based on the total number of pages for the book (Fig. 3). The final manuals consist of step-by-step guides on how to evaluate each checkpoint. Each checkpoint has its own section,

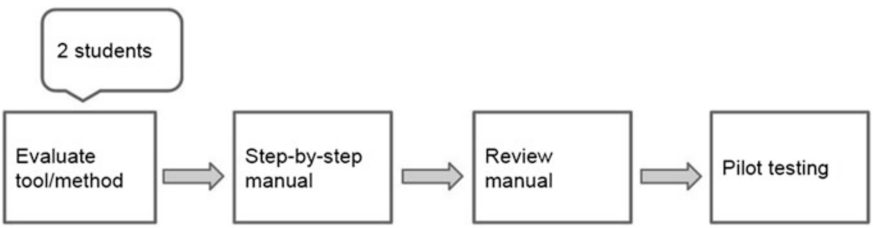


Fig. 2 Second stage of the manual creation and tool evaluation process

EPUB Accessibility Checklist

Content	Name of book	Format	OS Used	Total Number of Pa	Number of Chapters
		EPUB	Windows	300	

#	Check point	Criteria	Amount of Mater	Pass/Fail	Ranking 1-10	Additional Info	Detail
1	Acc. Documentation	A. URL to Formal Accessibility Policy					
1	Acc. Documentation	B. URL to Accessibility Statement					
1	Acc. Documentation	C. URL to Accessibility Evaluation Report					
2	Text Access	Text to Speech	30 pages				
3	Text Adjustment	A. Compatible (Size)	15 pages				
3	Text Adjustment	B. Adjust font and colors	15 pages				
4	Reading Layout	A. Reflow the text	15 pages				
4	Reading Layout	B. Page #s match printed material & reflow of	15 pages				
5	Reading Order	Digital resource layout	5 pages				
6	Structural Markup	A. Navigation text					
6	Structural Markup	B. Lists					
6	Structural Markup	C. eReader application					
7	Table Markup	Table Markup					
8	Hyperlinks	Hyperlinks (in-book)	30 links				
8	Hyperlinks	Hyperlink (live)	20 links				
		Hyperlink Functionality (Live)	20 links				
		Hyperlink Description (Live)	20 links				
9	Color and Contrast	A. Color redundancy	15 pages				
9	Color and Contrast	B. Contrast	30 pages				
		Contrast -Header					
		Contrast -Text					
		Contrast - Simple images					
10	Language	A. Markup					
10	Language	B. Passage Markup					
11	Images	A. Non-decorative	30 pages				
11	Images	B. Decorative	30 pages				
11	Images	C. Complex	30 pages				
12	Multimedia	A. Text Track					
12	Multimedia	B. Transcript					
12	Multimedia	C. Assistive Player					
13	Flickering	Flickering					
14	STEM	A. Markup (figures)	10 figures				
14	STEM	A. Markup (graphs)	10 graphs				
14	STEM	A. Markup (equation)	10 equations				
14	STEM	A. Markup (tables)	10 tables				
14	STEM	B. Notation Markup (figures)	10 figures				
14	STEM	B. Notation Markup (graphs)	10 graphs				
14	STEM	B. Notation Markup (equation)	10 equations				
14	STEM	B. Notation Markup (tables)	10 tables				
15	Interactive Elements	A. Keyboard					
15	Interactive Elements	B. Markup					
15	Interactive Elements	C. Text Prompts					

Fig. 3 Excel checklist

Fig. 4 Example of checkpoint description

10. Language

☑ The text of the digital resource includes markup that declares the language of the content in a manner that is compatible with assistive technology

☑ If the digital resource includes passages in a foreign language, these passages include markup that declares the language in a manner that is compatible with assistive technology

starting with a short description of what evaluators should be checking for (Fig. 4), and then followed by a step-by-step break down on how evaluators should evaluate the specific checkpoint (Fig. 5 for an example). Screenshots are provided for each step to help novice evaluators (see Figs. 6 and 7 for examples).

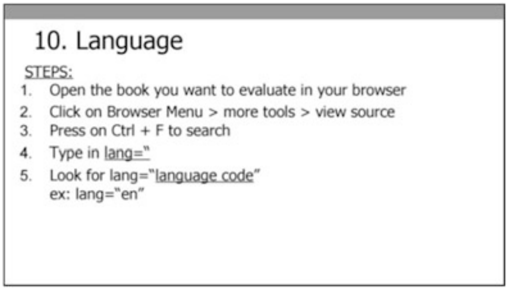


Fig. 5 Example of step-by-step breakdown for checkpoints

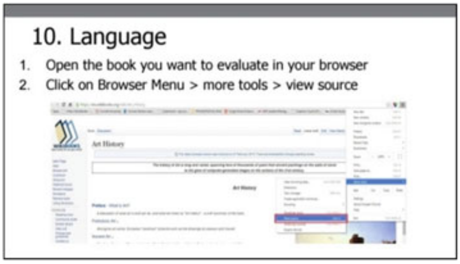


Fig. 6 Example of step-by-step breakdown with screenshots

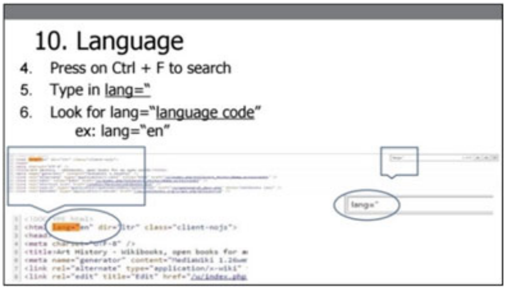


Fig. 7 Example of step-by-step breakdown with screenshots

4 Conclusion

We have successfully developed a step-by-step manual for three e-textbook formats, EPUB, HTML, and PDF. These manuals can be used by organizations to perform accessibility evaluations. They can also be used as prototypes for creating future manuals and training systems relating to accessibility evaluations.

Acknowledgments Special thanks to Accessible Instructional Materials Center and the Center for Usability in Design and Accessibility for supporting this project and all of our efforts. Partial funding for this project was provided by the State of California, the William and Flora Hewlett Foundation, and the Bill and Melinda Gates Foundation for the California Open Online Library for Education.

References

1. Hansen, D.: Knowledge transfer in online learning environments. J. Market. Educ. **30**(2), 93–105 (2008)
2. Mayadas, A., Bourne, J., Bacsich, P.: Online education today. Science **323**(5910), 85–89 (2009)
3. Caldwell, B.B., Vanderheiden, G.C.: Access to web content by those with disabilities and others operating under constrained conditions. Handbook Human Factors Web Design **2**, (2011)
4. Web Accessibility Initiative (WAI): Introduction to web accessibility. https://www.w3.org/WAI/intro/accessibility.php
5. Cool4Ed, about this project. http://www.cool4ed.org/about.html
6. SkillsCommons, Accessibility and SkillsCommons. http://support.taaccct.org/accessibility/
7. American Foundation for the Blind (AFB): Screen readers. http://www.afb.org/ProdBrowseCatResults.asp?CatID=49
8. American Foundation for the Blind (AFB): Screen magnification system. http://www.afb.org/ProdBrowseCatResults.asp?CatID=39
9. American Foundation for the Blind (AFB): Optical character recognition systems. http://www.afb.org/ProdBrowseCatResults.asp?CatID=38

Development of a Scoring System for Evaluating the Accessibility of eTextbooks

Mei Ling Chan, Yu Ting Sun, Andriana M. Tesoro and Kim-Phuong L. Vu

Abstract The Internet enables educators to provide free educational resources to many users, including those with disabilities. Existing regulations outline minimum accessibility requirements. However, we are not aware of any standard metric that would allow users to judge the relative degree of accessibility provided by e-learning materials, such as eTextbooks. The goal of the present study was to develop an accessibility scoring method that can be used to guide users' decision making when selecting eTextbooks. Using both non-assistive and assistive technologies, we analyzed a sample of 37 free access eTextbooks using 15 SkillsCommons accessibility checkpoints. We then worked with accessibility SMEs to determine severity weightings for each checkpoint in terms of its overall accessibility impact on users. Although, the scoring technique we developed needs further validation, it provides a starting point for accessibility researchers and post-secondary education stakeholders to quantify the level of accessibility provided by different eTextbooks.

Keywords Accessibility evaluation · Assessment · Assistive technology · eTextbook · Online resources · Scoring criteria · System development

M.L. Chan · Y.T. Sun · A.M. Tesoro · K.-P.L. Vu (✉)
California State University, Long Beach, 1250 Bellflower Blvd, Long Beach, CA 90840, USA
e-mail: kim.vu@csulb.edu

M.L. Chan
e-mail: email@meilingchan.com

Y.T. Sun
e-mail: debby130403@gmail.com

A.M. Tesoro
e-mail: ambtesoro392@gmail.com

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_29

1 Introduction

Forty-five percent of the world's population uses the Internet [1], and Americans, on average, spend 2.5 h per day online [2]. The wide use of the Internet has allowed the adoption of online learning materials as part of the curriculum for many post-secondary education systems. For example, in 2013, the California legislation enacted a bill that directed the California State University (CSU) system to establish a California Digital Open Source Library (CDOSL) with intention of making "higher education in California more affordable by providing free and low cost instructional materials" to faculty and students [3]. The California Open Online Library for Education (COOL4ed) is a project developed by CSU and CDOSL that was designed for faculty and students to find and adopt online course materials at little or no cost. Some of these eTextbooks incorporate advanced and interactive features, such as interactive learning diagrams and video tutorials. Thus, the adoption of these eTextbooks by faculty could lead to better learning experiences for students. If these eTextbooks are accessible, the books will also benefit students with disabilities by allowing them easier access to the content when using with assistive technologies.

Federal regulations provide minimum accessibility guidelines that need to be followed [4], but these regulations may not be sufficient in allowing students to maximally benefit from the content provided by the eTextbooks. In addition, there are no standard evaluation guidelines for educators and students to reference when deciding which eTextbooks to use or adopt. This project was conducted in support of the larger COOL4ed project with the goal of developing a quantitative scoring system that can be used by an organization or end users to determine the accessibility features of eTextbooks. Common validity issues in evaluating web resources' accessibility [5] were addressed through consultation with accessibility experts. Our metric for scoring eTextbooks can inform users in the selection of eTextbooks, and encourage publishers to evaluate their products with the goal of developing more accessible eTextbooks in the future.

2 Method

2.1 Evaluators

Six raters were recruited to perform accessibility assessments of a sample of eTextbooks posted on the COOL4ed website. Four raters were graduate students in the Human Factors Master's Program at CSU Long Beach, and had taken introductory human factors courses and were familiar with conducting usability evaluations. The other two raters were senior psychology students who were novices in human factors research. All raters received a training session that lasted 1-h. The training included a step-by-step walkthrough of an evaluation manual for one eTextbook format (see Sect. 2.3 below for details). After all raters completed the training eTextbook evaluation, a meeting was held to address any discrepancies observed in evaluation scores, and to discuss any technical issues found during individual evaluation process. The raters then use this knowledge to assess another training eTextbook, followed by another meeting to discuss discrepancies or issues. This procedure was repeated for four eTextbooks to help maximize their scoring reliability on subsequent eTextbook evaluations. The raters were also advised to strictly follow the evaluation manual.

2.2 eTextbooks

Thirty-seven free access eTextbooks from the COOL4ed website were selected for evaluation. All eTextbooks were published in at least one of the following document formats: EPUB, HTML, PDF, and Microsoft Word.

EPUB EPUB (Electronic Publication) is a free and open standardized eBook publishing format established in 1999 by the International Digital Publishing Forum (IDPF). The IDPF is the international trade and standards organization for the digital publishing industry [6]. EPUB became the official eBook format in 2007 [7]. IDPF released the newest version of EPUB, EPUB3.0.1, in 2014 [7].

HTML HTML (**H**yper **T**ext **M**arkup **L**anguage) is a markup language used for creating web documents. Computer scientist Tim Berners-Lee first developed HTML in 1992, and the World Wide Web Consortium (W3C) currently maintains and regulates HTML [8]. The newest version, HTML5, was released in 2014 [9].

PDF PDF (**P**ortable **D**ocument **F**ormat) is a file format that presents documents in a fixed layout across application software and operating systems [10]. Originally developed by computer scientist and co-founder of Adobe Systems Incorporated, John Warnock, in 1991. PDF became an open standard in 2008 and is now maintained by the International Organization for Standardization (ISO) [11]. The newest version, PDF 1.7, was released in 2011.

Microsoft Word Microsoft Word is a word processor developed by Microsoft Corporation in 1983 [12]. The newest version for Windows and Mac operating systems was released as part of Microsoft Office 2016 in 2015 [13]. Accessibility evaluation of Microsoft Word documents is not included in the present paper, though, because there was only one eTextbook from the sample that was produced in this format.

2.3 Accessibility Evaluation Manual

We used a step-by-step accessibility evaluation manual [14] to determine each eTextbook's accessibility level. The manual provided accessibility checks against the criteria of the 15 accessibility checkpoints developed by SkillsCommons, a free online library project managed by CSU and its Multimedia Educational Resources for Learning and Online Teaching (MERLOT) project.

SkillsCommons Accessibility Checkpoints The 15 SkillsCommons accessibility checkpoints (see Table 1) are criteria developed for producing accessible open educational resources [15]. The evaluation for accessibility checkpoints can be completed separately using non-assistive technologies (non-AT) and assistive technologies (AT).

Evaluation Tools The present study used evaluation tools that were listed in the accessibility evaluation manuals [14]. All evaluations were completed using computers with Windows operating systems.

2.4 Procedure

For training purpose, two eTextbooks were evaluated by multiple raters to determine inter-rater reliability. Because only a sample of pages was selected for evaluating each eTextbook, and each rater could select different pages to check, we found inter-rater reliability to be moderate to high (rs range 0.45–0.88, ps < 0.05). As a result, we trained the raters to a criterion. When a rater was able to produce a book evaluation, where when checked by the project lead, resulted in the same scoring for each checkpoint, the rater was considered trained. All raters were considered trained by the end of four book evaluations. Thus, the remaining eTextbooks was evaluated by a single rater using non-AT and AT procedures, and the project lead would verify the scores by revisiting a sample of the evaluated sections of the eTextbooks.

If an eTextbook was available in more than one document format, the format selected for evaluation was in the order of (1) EPUB, (2) HTML, and (3) PDF. Evaluations were performed individually at raters' own pace. Raters were asked to assess each eTextbook against each checkpoint criteria and provide scores and comments accordingly. An overall accessibility score for each eTextbook was complied at the end by the project lead.

Table 1 SkillsCommons 15 accessibility checkpoints

#	Checkpoint	Criteria
1	Accessibility documentation	• The online material has a formal accessibility policy and accessibility statement • An Accessibility Evaluation Report is available from an external organization
2	Text access	• The text of the digital resource is available to AT that allows the user to enable text-to-speech (TTS) functionality
3	Text adjustment	• Text is compatible with AT • The resource allows the user to adjust the font size and font/background color[a]
4	Reading layout	• Text is compatible with AT that allows the user to reflow the text by specifying the margins and line spacing[a] • If the digital resource is an electronic alternative to printed materials, the page numbers correspond to the printed material
5	Reading order	• The reading order for digital resource content logically corresponds to the visual layout of the page when rendered by AT
6	Structural markup/navigation	• Text includes: (1) markup (e.g. tags or styles) that allows for navigation by key structural elements (chapters, headings, pages) using AT[a]; and (2) markup for bullets and numbered lists that is compatible with AT[a] • If the text is delivered within an eBook reader application, a method is provided to allow users to bypass the reader interface and move directly to the text content that is compatible with AT
7	Tables	• Tables include markup (e.g. tags or styles) that identifies row and column headers in a manner that is compatible with AT[a]
8	Hyperlinks	• URLs (e.g. website or email addresses) within the text of the digital resource are rendered as active hyperlinks in a manner that allows them to be detected and activated with AT[a]
9	Color and contrast	• All information within the digital resource that is conveyed using color is also available in a manner that is compatible with AT • The visual presentation of text and images of text in the digital resource has a contrast ratio of at least 4.5:1
10	Language	• Text includes markup that declares the language of the content in a manner that is compatible with AT. This also applies to digital resources that include passages in a foreign language
11	Images	• Non-decorative images have AT-compatible alternative text[a] • Decorative images are marked with null alternate text or contain markup that allows them to be ignored by AT • Complex images, charts, and graphs have longer text descriptions that are compatible with AT[a]
12	Multimedia	• A synchronized text track (e.g. open or closed captions) is provided with all video content • A transcript is provided with all audio content • Audio/video content is delivered via a media player that is compatible with AT. This includes support for all criteria listed in checkpoint 15

(continued)

Table 1 (continued)

#	Checkpoint	Criteria
13	Flickering	• The digital resource content does not contain anything that flashes more than three times in any one-second period
14	STEM elements	• STEM content (e.g. Mathematics, Chemistry) is marked up in a manner that is compatible with AT[a]
15	Interactive elements	• Each interactive element (e.g. menu, hyperlink, button) and function (e.g. annotations) allows keyboard-only operation both with and without AT • Each interactive element conveys the element's name, type, and status to AT (e.g. "play, button, selected") • All instructions, prompts, and error messages necessary to complete forms are conveyed as text to AT[a]

[a]Or are rendered by an application such as a browser, media player, or reader that offers this functionality

2.5 Computing Accessibility Score

Each accessibility checkpoint was rated on a scale 1 to 10, with 10 being the highest accessibility rating. Using this scale and the common grading system, where 90–100 % = A, 80–89 % = B, and 70–79 % = C, an accessibility score of 7, or 70 %, is considered as the passing score. In other words, when a checkpoint has a score 7 or above, the eTextbook satisfies the accessibility criteria. Three different computations, composite score, average score, and weighted average score, were generated to represent the overall accessibility level of eTextbooks.

Composite Score This method combines the assigned scores from all measurable accessibility checkpoints (i.e. the number of checkpoints that we were able to evaluate using the evaluation tools) and divides the sum by the number of measurable checkpoints (see Eq. 1). For this method, if a checkpoint is measurable but such content is not presented in a specific eTextbook, that checkpoint will be automatically given a maximum accessibility score of 10 (i.e., the textbook showed no violation of the checkpoint).

$$\frac{\Sigma(Checkpoint\ Score)}{Number\ of\ Measurable\ Checkpoints} \tag{1}$$

Average Score This method for computing the accessibility scores is based on summing all scores on applicable checkpoints (i.e. the number of checkpoints that could be evaluated) and dividing it by the number of applicable checkpoints (see Eq. 2). When a checkpoint is not applicable to an eTextbook, for example, if the eTextbook does not contains any STEM content, the average score method simply takes that checkpoint out of consideration when computing the score.

$$\frac{Total\,Score\,Assigned}{Number\,of\,Applicable\,Checkpoints} \quad (2)$$

Weighted Average Score A weighted average score was also computed because not all of the checkpoints from the SkillsCommons checklist are equal contributors to the user's accessibility experience. For example, checkpoint 1 evaluates whether

Table 2 Assigned weighting for each checkpoint's criteria

#	Checkpoint	Criteria	Weight
1	Accessibility documentation	URL to formal accessibility policy URL to accessibility statement URL to accessibility evaluation report	0 0 0
2	Text access	Text to speech function	1
3	Text adjustment	Compatible (size and data loss) Adjust font/background color	1 0.5
4	Reading layout	Reflow the text Page numbers match printed material	1 0.3
5	Reading order	Digital resource layout	1
6	Structural markup/ navigation	Navigation text Lists eReader applications	1 0.7 1
7	Tables	Table markup	1
8	Hyperlinks	Hyperlinks (in-book) Hyperlink (live) functionality Hyperlink (live) description	0.4 1 0.5
9	Color and contrast	Color redundancy Contrast	0.7 0.5
10	Language	Markup Passage markup	0.3 0.3
11	Images	Non-Decorative Decorative Complex	1 0 1
12	Multimedia	Text track Transcript Assistive player	1 0 1
13	Flickering	Flickering	0.5
14	STEM elements	Markup (figures) Markup (graphs) Markup (equations) Markup (tables) Notation markup (figures) Notation markup (graphs) Notation markup (equations) Notation markup (tables)	1 1 1 1 1 1 1 1
15	Interactive elements	Keyboard Markup Text prompt	1 1 1

the organization provided accessibility documentation in its online materials. It is ideal that the publishers of eTextbooks provide accessibility statements about their products, but not having one does not negatively impact the users' experience with the book. After careful consultation with two different accessibility subject matter experts (SMEs), each accessibility checkpoint was given a weight from 0 to 1, with 1 being most important to the overall user experience when using an eTextbook. Table 2 shows the weighting for each criterion of accessibility checkpoints. As noted earlier, checkpoint 1 Accessibility Documentation has little or no relevance in contributing to the user experience and thus was given a weighting of 0. On the other hand, checkpoint 5 Reading Order was given a weighting of 1. That is, if an eTextbook's reading layout is not compatible with AT, such as screen readers, the material would not make much sense to the user with vision impairments, and would be very disruptive to the user's learning experience.

Weighted average scores can be calculated by dividing the sum of checkpoint scores by the number of applicable checkpoints (see Eq. 3). Each individual checkpoint score is computed using the criteria scores with their assigned weights (see Eq. 4).

$$\frac{\Sigma(\mathit{Checkpoint\ Score})}{\mathit{Number\ of\ Applicable\ Checkpoints}} \tag{3}$$

$$\frac{\Sigma(\mathit{Criterion\ Score} \times \mathit{Criterion\ Weight})}{\mathit{Number\ of\ Criteria}} \tag{4}$$

3 Results

Among the 37 eTextbooks, 7 EPUB, 22 HTML, and 8 PDF document formats were evaluated. PDF had the lowest passing rate for all three computation methods, with only a 12.5 % average passing rate. EPUB had the highest passing rate (86 %) using the composite score, but a 0 % passing rate using the average score. With the

Table 3 Composite accessibility scores for each document format

Document format	Book passing ratio (%)	Non-AT score[a]	AT score[b]	Total score[c]
EPUB	6/7 (86 %)	7.52[d]	6.80	14.32[d]
HTML	19/22 (86 %)	7.73[d]	7.85[c]	15.58[d]
PDF	2/8 (25 %)	6.80	6.18	12.98

[a]There were 11, 12, and 15 measurable checkpoints for EPUB, HTML, and PDF formats, respectively

[b]There were 12, 15, and 15 measurable checkpoints for EPUB, HTML, and PDF formats, respectively

[c]The sum of non-AT and AT scores

[d]Passed accessibility evaluation

Table 4 Average accessibility scores for each document format

Document format	Book passing ratio (%)	Non-AT score	AT score	Total score
EPUB	0/7 (0 %)	7.06[a]	6.16	13.22
HTML	13/22 (59 %)	7.13[a]	7.36[a]	14.49[a]
PDF	0/8 (0 %)	5.88	5.56	11.44

[a]Passed accessibility evaluation

Table 5 Weighted average accessibility scores for each document format

Document format	Book passing ratio (%)	Non-AT score	AT score	Total score
EPUB	3/7 (43 %)	7.26[a]	6.74	14.00[a]
HTML	15/22 (68 %)	7.19[a]	7.59[a]	14.78[a]
PDF	1/8 (12.5 %)	6.26	5.79	12.05

[a]Passed accessibility evaluation

weighted average score, 43 % of the EPUB books passed. HMTL was found to be the most accessible format, with a passing rate of 86, 59, and 68 % for composite, average, and weighted average scores, respectively. Tables 3, 4, and 5 show the composite, average, and weighted average scores for each document format respectively.

4 Discussion

We compared three scoring techniques for determining the overall accessibility level of eTextbooks: composite, average, and weighted average. The advantage of computing a composite score is the ease of calculation; however, scores were inflated because non-applicable checkpoints are automatically given scores of 10 given that the eTextbook did not violate those checkpoints. The average score method removed the score inflation problem associated with the composite score computation. However, because not all accessibility checkpoints are equally important for end users in terms of their accessibility experience, this metric may not be an accurate indicator of eTextbooks' accessibility level. We found that the weighted average score to be the best computation technique, as it provided a meaningful quantitative score to guide users' decision making when selecting eTextbooks based on accessibility by taking into account how accessibility would impact the user's experience.

Using the weighted average scores, we found that the HTML eTextbooks had the highest passing rate of 68 % compared to the EPUB and PDF. This is probably due to versatile nature of HTML, which allows Internet users to modify and customize the website and its content to fit their individual needs. In contrast, only 12.5 % of PDF eTextbooks passed the accessibility evaluation. EPUB eTextbooks had a 43 % passing rate, which is lower than our original expectation. One possible reason for the relatively low passing rate, though, is that EPUB is still a fledgling document format. IDPF only standardized EPUB as the official eBook format in 2007 [6], and it is only on its third version, which means EPUB is still in the growth development process. It took HTML more than 20 years to arrive at its current version [8], so it may take longer for EPUB reader software and other assistive technologies to work seamlessly with EPUB documents. Furthermore, since EPUB is a fairly new document format, there could be a lack of guidelines for programmers and eBook designers to follow when creating EPUB documents.

The current study is an on-going project and we are only in the preliminary stages. Further research is needed to validate and refine our scoring method. However, this project provides an initial guide for educators to evaluate the level of accessibility of each eTextbook in the COOL4ed collection. Once accessibility scores are compiled for all eTextbooks available on the COOL4ed website, an accessibility report for each book will be published on the website to help educators and students select eTextbooks. In future, we hope such evaluation practice would encourage publishers to develop more digital materials that are accessible to all users.

Acknowledgment The authors would like to thank the Accessible Instructional Materials Center and the Center for Usability in Design and Accessibility at CSULB for supporting this project and all of our efforts. We also thank the two accessibility SMEs who helped us generate the weights for the different checkpoints. Partial funding for this project was provided by the State of California, the William and Flora Hewlett Foundation, and the Bill and Melinda Gates Foundation for the California Open Online Library for Education.

References

1. Internet Live Stats.: Number of Internet Users. http://www.internetlivestats.com/internet-users/ (2015)
2. Statista, U.S.: Average Daily Media Use. http://www.statista.com/statistics/270781/average-daily-media-use-in-the-us/ (2014)
3. COOL4ed.: About the Project. http://www.cool4ed.org/about.html
4. United States Access Board.: Section 508 Standards. https://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-section-508-standards/section-508-standards
5. Brajnik, G.: Validity and reliability of web accessibility guidelines. In: 11th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 131–138. ACM, New York (2009)
6. International Digital Publishing Forum.: About Us. http://idpf.org/about-us
7. International Digital Publishing Forum.: EPUB. http://idpf.org/epub
8. World Wide Web Consortium (W3C).: Chapter 2—A History of HTML. https://www.w3.org/People/Raggett/book4/ch02.html
9. World Wide Web Consortium (W3C).: HTML5 is a W3C Recommendation, W3C News. https://www.w3.org/blog/news/archives/4167
10. Adobe Acrobat DC.: PDF files, Adobe Portable Document Format. https://acrobat.adobe.com/us/en/products/about-adobe-pdf.html

11. ISO 32000-1: 2008.: Document management, Portable document format, Part 1: PDF 1.7. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51502
12. Allen, R.A.: Chapter 12: microsoft in the 1980's. In: A History of the Personal Computer: The People and the Technology, 1st edn., pp. 12/25–12/26. Allan Publishing, Canada (2001)
13. Office Blogs.: Office 2016 for Mac is here! https://blogs.office.com/2015/07/09/office-2016-for-mac-is-here/
14. Sun, Y.T., Chan, M.L., Manabat, A.K.M., Vu, K.-P.L.: Accessibility evaluation: manual development and tool selection for evaluating accessibility of E-textbooks. In: Proceedings of the 7th AHFE International Conference (In press; this volume)
15. SkillsCommons.: Accessibility Requirement Guidelines. http://support.taaccct.org/ada-requirements/

Measuring the After-Effects of Disruption on Task Performance

Robert G. Abbott, Eric Moyer and Chris Forsythe

Abstract In many settings, multi-tasking and interruption are commonplace. Multi-tasking has been a popular subject of recent research, but a multitasking paradigm normally allows the subject some control over the timing of the task switch. In this paper we focus on interruptions—situations in which the subject has no control over the timing of task switches. We consider three types of task: verbal (reading comprehension), visual search, and monitoring/situation awareness. Using interruptions from 30 s to 2 min in duration, we found a significant effect in each case, but with different effect sizes. For the situation awareness task, we experimented with interruptions of varying duration and found a non-linear relation between the duration of the interruption and its after-effect on performance, which may correspond to a task-dependent interruption threshold, which is lower for more dynamic tasks.

Keywords Human factors · Multitasking · Interruption · Disruption

1 Introduction

Systems are becoming increasingly automated in the endless drive for higher productivity and lower cost. Increasing automation leads to increased dependency on the system, as the additional resources that were required for manual operations are not maintained. In extreme cases, it is virtually impossible to maintain operations in the face of a complete system failure. For example, air traffic control cannot be carried out after a failure of electrical power (including backup generators). Mission-critical systems are designed for high availability using techniques such as redundancy and extensive testing. Nevertheless, unlikely or unforeseen circumstances can result in an outage, and contingency plans must be developed. This requires understanding potential outages, including predicting the amount of time

R.G. Abbott (✉) · E. Moyer · C. Forsythe
Sandia National Laboratories, Albuqerque, NM, USA
e-mail: rgabbot@sandia.gov

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_30

that operations may be effected. Human operators and support personnel are a crucial factor in the system, and must be included in the analysis. This requires understanding how the cognitive processes that underlie task performance are impacted by disruption. In this study, we designed three tasks to exercise different aspects of cognition and measured various impacts of task disruption.

In a theoretical paper by Endsley and Jones [1], a framework was described that identified human decision making and cognitive performance parameters affected by disruptive events. These parameters explained the interface between the technological system impacted by the disruptive event and human operators, and the mechanisms by which disruptive events may have deleterious results that exceed those that might be expected solely on the basis of technological system components. The authors first note that following a disruptive event, personnel must first recognize that the event is distinct from normal operational problems (e.g., system crashes). The authors next describe four major effects of disruptive events on decision making and situation awareness: (1) disruptions affecting information processing (e.g., delayed or corrupted information); (2) disruptions that affect prioritization and attention (e.g., disruptions that shift attention away from other ongoing critical activities); (3) disruptions that affect confidence in information (e.g., disruptions that cast doubt on certain information causing undue effort to be expended to confirm its validity); and (4) disruptions that affect interpretation (e.g., disruptions that lend themselves to misinterpretation and consequently, prompt ineffective or counterproductive reactions).

Research undertaken to understand decision processes has generally focused on after-the-fact field studies. For example, in a study of disruptions in the availability of computing systems, it was found that levels of frustration varied with self-efficacy, with those having the greater sense of control over events experiencing the least frustration. In a study of emergency medical care, McDaniels et al. [2] found that "resilience" was a function of organizational factors that included: learning from past experiences; personnel staffing; personnel communication; and flexibility in addressing specific failures. From an analysis of emergency responses following the World Trade Center, Kendra and Wachtendorf [3] emphasized the critical role of creativity in adapting available resources to ongoing situational factors. While informative, these and other similar studies, provide limited insight into the relationship between parameters of human decision making and disruptions to technological systems. An understanding of these relationships is important in anticipating the magnitude of disruptive effects on technological systems, as well as understanding factors that might exacerbate or lessen the impact. Consequently, there is need for controlled laboratory studies that will allow a better understanding of these factors and their relationships to one another.

The objective of this study is to characterize the disruptive effects of loss of information technology. It is hypothesized that disruptions will result in performance decrements that go beyond that predicted by mere loss of computing and information systems (i.e., availability cost). Additionally, the current study allows for the development and refinement of paradigms and protocol for testing the effects of disruptions to information technology.

2 Methods

Participants were assigned to three-person teams and presented a series of tasks that exercised a variety of cognitive functions. In experimental conditions, task performance was interrupted for a pre-specified duration, during which subjects were required to perform a secondary task. In control conditions, subjects were allowed to complete tasks uninterrupted.

2.1 Subjects

Participants consisted of 21 employees of Sandia National Laboratories who responded to an announcement to the general workforce. Individuals had to meet the following criteria to be selected for the study: (1) 18–35 years of age; (2) fluent in English; and (3) possessed basic computer skills (e.g., Microsoft Office products and Internet web browser).

2.2 Procedure

Subjects were assigned to three-person teams for the study. There were seven teams of three subjects. In the consent form, it was explained that the study was being conducted with the objective of gaining a better understanding of the effects of interruptions on the performance of teams.

Each participant was assigned a computer and asked to do all work on the assigned computer and to not swap computers between members of the team. Subjects were also asked to not use any personal electronics (e.g., smart phone) to assist in performing tasks.

Tasks were presented in a fixed order, with experimental and control conditions counterbalanced. Each task consisted of a set of subtasks. Tasks and their subtasks were accessed through a game server. Once teams were allowed to work on a task, each of the task's constituent subtasks could be accessed by each member of the team through the game server. However, once one member of the team had accessed a subtask, no other team member could access that subtask, unless it was abandoned by the individual who had originally accessed the subtask. Only the participant with access to a given subtask could submit a response. For a given subtask, teams could submit only one response and there was no feedback provided with regard to the accuracy responses.

Warm-Up Task The warm-up task consisted of a set of puzzles that teams were presented to allow them to familiarize themselves with the game server and one another prior to the experimental tasks. For this task, subjects were presented nine separate puzzles that they could access through the game server. Below is an example of one of the puzzles:

> Assume the following: a pound of tea has twice as much caffeine as a pound of coffee. A pound of tea is enough to make 160 cups of tea. A pound of coffee is enough to make 40 cups. A 12-ounce can of cola has about one-fourth the caffeine as a cup of coffee. How much caffeine does one cup of tea have compared to one cup of coffee?

Teams were allowed 10 min to work on the puzzles.

Reading Comprehension and Retention Task For this task, subjects were provided 3 two-page documents that each provided a detailed description of a facility. Subjects were allowed 5 min to read the descriptions, after which those in the control condition were presented questions to answer based on the information within the descriptions. Each question appeared as a separate panel on a game board displayed through the game server. While answering the questions, the facility descriptions remained available to the subjects. Questions were constructed so that they were either specific to a facility (e.g., Where is the air traffic control tower at the airfield?) or required a comparison of the three different facilities (e.g., The north side of which facility is readily accessible from the road?) Teams were allowed as much time as needed to answer the questions.

For subjects in the experimental condition, the period they were allowed to read the descriptions prior to the questions was followed by a 5 min interruption. For the interruption, the computer monitors were turned off by the experimenter and subjects were instructed to direct their attention to completing a jigsaw puzzle on an adjacent table.

At the conclusion of the interruption, the monitors were returned and the questions made available, as in the control condition. Following the interruption, the display contents of computer monitors were identical to when the interruption began.

The task began with a practice. For the practice, subjects were presented three facility descriptions that were each approximately a half page in length and given a minute to read them prior to the questions. Then, subjects were allowed as much time as they needed to answer each question.

Visual Search and Spatial Memory Task For this task, subjects were presented photographs obtained from Google Earth Street View. Their task was to identify the location from which photographs were taken using the overhead imagery (e.g. satellite) view. This generally involved identifying distinct features within the photographs and matching those features to the corresponding features in the overhead images (Fig. 1). Locations were selected that had distinct features that would enable the location to be found in the imagery. Each set of photographs were from a specific city or region and subjects were told to restrict their search to this location. Subjects submitted the GPS coordinates for the location as their response.

Initially, subjects completed a practice round that consisted of three photos and were allowed as much time as needed to submit their responses. This allowed subjects to familiarize themselves with functions for zooming in and out, panning in different directions and identifying the GPS coordinates for a specific location.

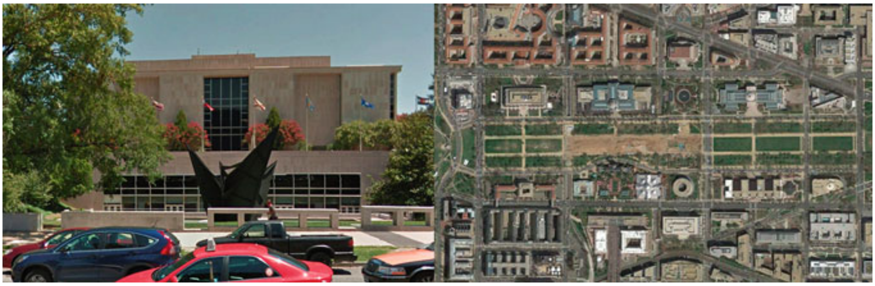


Fig. 1 In the visual search task, subjects were presented with street-level photograph (*left*) and asked to find the location on an overhead image (*right*) from which the photograph was taken. This example is the National Museum of American History in the vicinity of the National Mall in Washington, DC. *Photo Credit* Google Earth

For each experimental condition, subjects received nine photos from either of two regions. In both the experimental and control condition, subjects were first allowed 10 min to locate the photos. In the control condition, at the end of the ten minutes, subjects were told that there had been a system disruption and their work had been lost. They were then provided the same nine photos and instructed to find them a second time. The measure of interest was the difference in the time taken to find photos the first and second time. In the experimental condition, there was a 5 min interruption during which subjects were instructed to continue working on the jigsaw puzzle. After the interruption, subjects were asked to find the photos a second time.

Situation Awareness/Spatial Reasoning Task This task was a dynamic version of the visual search task. Subjects watched a video simulating an out-the-window view of an aircraft generated using the Play Tour function of Google Earth. While watching the video, the subject accessed a heading indicator (making it easier to project the search location from each previous location) and an overhead imagery application that allowed them to click on points to insert markers that were linked together to create a path. The task required the subjects to place a series of markers on the map that would chart the path of the aircraft. The parameters of the flyover video were tuned to make the task challenging but not impossible for most subjects. The flyover used an altitude of 500 m, a tilt angle of 60°, and a speed of 60 mph. Every 20 s the path turned approximately 15 or 30° to the left or right, or continued straight forward. Each subject on the team viewed a different flyover video from a different portion of the map, making this an individual task.

The task began with a practice session. Subjects were given the initial location of the aircraft on the map and then tracked the position of the airplane for 8 min.

In the control conditions, subjects performed tracking for rounds consisting of 8 min each. In the experimental condition, subjects were allowed to track the aircraft for 3 min, at which time the interruption occurred. The monitor was turned off and the subjects worked on the jigsaw puzzle. Each interruption was for one of three durations: 30 s, 60 s or 120 s. Following the interruption, subjects returned to the task for an additional 5 min. During the interruption, the simulated aircraft continued on its path, so it was necessary for subjects to reestablish the its location. Each team completed three rounds in the control condition and three rounds in the experimental condition, with one round for each duration of interruption.

3 Results

3.1 Reading Comprehension and Retention Task

For the data analysis, submissions were separated on the basis of which subject made the submission. Then, the time taken from the point that each challenge question was accessed until the answer was submitted was calculated. This allowed an average response time and the number of correct submissions to be determined for each subject. As shown in Fig. 2, on average, subjects submitted fewer correct responses for conditions involving an interrupt, as compared to conditions that had no interrupt. This difference was statistically significant: Paired t-test = 2.69; $p < 0.014$. On average, subjects required longer to answer questions in the interrupt condition than the no interrupt condition, however this difference was not statistically significant: Paired t-test = 1.17; NS (Fig. 3).

3.2 Visual Search and Spatial Memory Task

For the geospatial search task, photos were identified for which the same subject submitted a response during the initial block and again in the second block when

Fig. 2 In the situational awareness task, the subject viewed a simulated out-the-window view from an airplane (*left*) as it executed a series of turns, and charted the path of the airplane on a map (*right*). This example is from the vicinity of the National Mall in Washington, DC. *Photo Credit* Google Earth

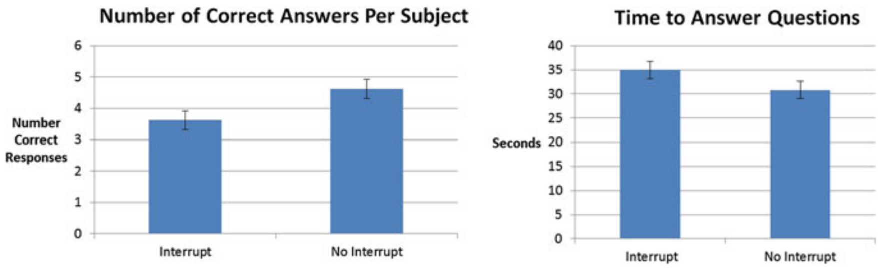


Fig. 3 Results from intelligence reports task averaged for each subjects showing accuracy of submissions and time for submissions

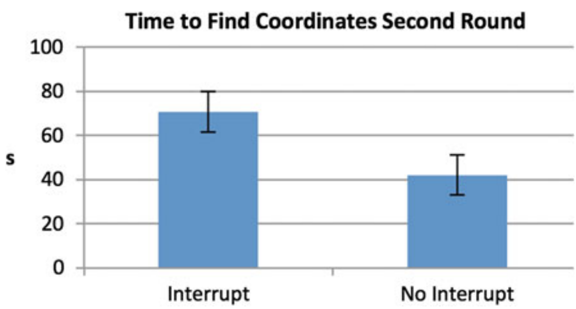
subjects were asked to repeat the task. The time for submissions was determined based on the total time that a photo was open. Thus, if a subject opened a photo and searched for it, but did not find it and abandoned the challenge, and later reopened the same challenge and submitted a correct response, the times for both instances were combined.

As shown in Fig. 4, subjects required more time to find the photos in the interrupt condition than in the no interrupt condition. This difference was statistically significant: t-test = 2.02; $p < 0.05$.

3.3 *Situation Awareness/Spatial Reasoning Task*

On the tracking task, longer interruptions resulted in higher tracking error after the end of the interruption. Figure 5 shows the mean tracking error over time for each interruption duration, with the highest (worst) 10 % samples discarded at each time step. The outliers were discarded because at any given time, a small percentage of subjects were unable or not trying to reacquire the track, so their tracking error came to dominate all the others. The mean tracking error for the control condition (no interruption) is the red curve. The area between the red curve and each other curve is the tracking error attributable to the interruption. The tracking error to the

Fig. 4 On average, subjects took significantly longer to find the location of a photograph on a map for the second time if there was an interruption of 5 min after finding the images for the first time



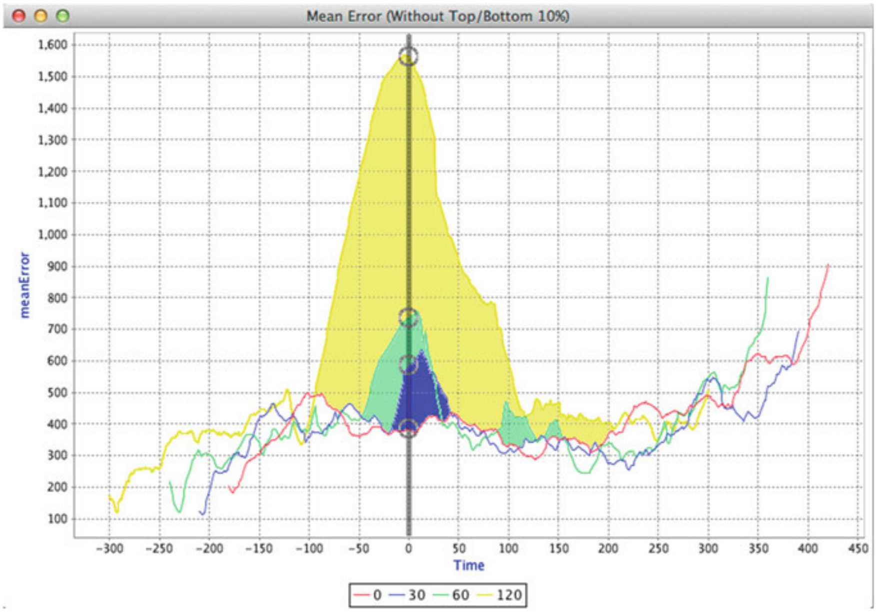


Fig. 5 Mean tracking error (*m*) as a function of time (*s*). Each curve corresponds to a different interruption duration

left of the vertical axis at t = 0 was beyond the subjects' control, as they had no access to the system during the interruption, whereas the area under the curves after this time reflects the time taken to re-acquire situational awareness.

After the longest interruption (120 s), the mean tracking error is degraded for approximately 180 s, which is longer than the interruption itself. However, the small magnitude of the effect after approximately 60 s implies that most subjects had recovered by this time. Shorter interruptions showed quicker recovery times. The shortest interruption, 30 s, had little effect.

4 Conclusions

These studies demonstrated that a system outage disrupts productivity not only while the computer is out of service, but for some time afterwards. The cognitive impacts depend on the type of task being performed. In all cases, disruption results in a loss of working memory. On returning to the task, subjects may initially be less productive than before the disruption as they regain context, as on our visual search task. They may also provide more incorrect responses, as on the reading comprehension and retention task. This increased error rate occurred despite the fact that subjects had access to the reading material after the interruption ended, and were not under time pressure. Possible explanations include impatience with repeatedly reviewing the same information motivating the subjects to answer a question despite being unsure of the answer, or the unconscious substitution of forgotten information with a plausible answer.

The situation awareness task has additional considerations. First, even a subject with perfect memory would show a decrease in performance after returning to the task, due to new information that was not observed during the interruption. Second, the nature of the task itself is changed by the disruption; what was a tracking/monitoring task becomes a search task until the track is re-acquired. It is possible that an individual could show above-average performance in one monitoring but not search, or vice versa. Moreover, the relative difficulty between monitoring and search could be altered dramatically by changing the task parameters, e.g. a slow-moving aircraft at an unknown location in a large area would be easy to monitor, but difficult to re-acquire once lost. The duration of the interruption had a nonlinear impact on the duration of the subsequent decline in performance. It appears there is a threshold or 'tipping point' below which situation awareness is not lost and performance resumes almost immediately when the system is available. Above this point, performance is strongly impacted. Once the individual loses situation awareness, their previous knowledge is invalidated and they must re-acquire the new state after the interruption, and it is questionable whether an even longer interruption would show a proportionally larger after-effect. This suggests a sigmoidal curve for the time to re-acquire situation awareness after an interruption, with the maximum slope corresponding to an 'interruption threshold' specific to the current task demands. For example, driving a car on a long, straight second of road has a larger interruption threshold than driving on a winding

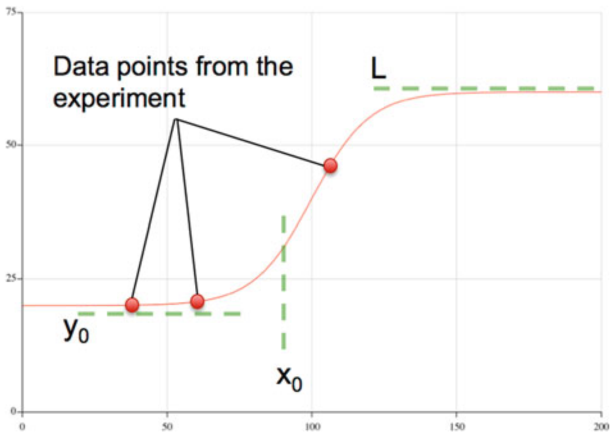


Fig. 6 Proposed formula for the median time to return to pre-disruption level of performance as a function of duration of disruption, and a plot of this formula using parameters selected to fit data collected from the Tracking task. Each of the three data points is the median of 63 trials (3 trials for each of 3 subjects on 7 teams)

mountain road, because the situation is more dynamic. Further experimentation is needed to explore this hypothesis (Fig. 6).

$$t = y_0 + \frac{L - y_0}{1 + e^{-k(X - X_0)}}$$

Term	Description	Tracking task data fit
t	Recovery time	20–60 s
y_0	Minimum recovery time	20 s
L	Maximum recovery time	60 s
k	Maximum slope	0.1
x_0	Interruption to achieve 50 % effect	100 s

On each of our tasks, the measured cognitive impact was on the order of seconds, rather than minutes. This may or may not be large compared to the time required to repair the computer system, depending on the severity of the disruption. Therefore both system modeling and cognitive modeling are necessary components of an overall model of the impact of disruption.

References

1. Endsley, M.R., Jones, D.G.: Disruptions, interruptions and information attack: impact on situation awareness and decision making. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 45, no. 2, pp. 63–67. SAGE Publications (2001)
2. McDaniels, T., Chang, S., Cole, D., Mikawoz, J., Longstaff, H.: Fostering resilience to extreme events within infrastructure systems: characterizing decision contexts for mitigation and adaptation. Glob. Environ. Change **18**(2), 310–318 (2008)
3. Kendra, J., Wachtendorf, T.: Creativity in emergency response to the world trade center disaster. Beyond Sept. 11th Account Post-Dis. Res. 121–146 (2003)

Complexity and Reliability as Basic Quantitative Characteristics of Computer Based Tasks

Inna Bedny, Gregory Bedny and Waldemar Karwowski

Abstract This work is dedicated to the quantitative methods of task analysis that allow to evaluate task complexity and reliability of task performance. Specific attention is paid to reliability and complexity assessment of computer based task and the correlation between these vital characteristics of human performance. The qualitative and quantitative description of computer based task performance is utilized as an example. In our work we utilize methods of task analysis that were developed in the framework of systemic-structural activity theory (SSAT). Description of activity structure during task performance and its relationship to the structure of computer interface is analyzed. Reliability and task complexity assessment calls for the creation of various models of activity during interaction of user with computer. Such approach is different from cognitive approach which considers cognition only as a process and ignores the concept of cognitive structure.

Keywords Complexity · Reliability · Precision · Errors · Failure

1 Introduction

In this paper the main focus is on the study of the relationship between reliability and complexity in the performance of computerized tasks. We will first consider the difference between such concepts as reliability and precision of task performance. The main concept for probability is failure, for precision it is error. These two basic concepts are tightly interconnected but they are not the same. Precision refers to the

I. Bedny (✉) · G. Bedny
Evolute, Louisville, KY, USA
e-mail: innayarosh@gmail.com

G. Bedny
e-mail: gbedny@optonline.net

W. Karwowski
University of Central Florida, Orlando, FL, USA
e-mail: wkar@ucf.edu

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_31

accuracy with which a goal of task is achieved. Reliability refers to failure in task performance, and how probability of failure in task performance can change over time. In systemic-structural activity theory (SSAT) ergonomists and psychologists distinguish between user errors and failures based on the criticality of the errors for the system [2, 3]. Operator's actions that negatively affect functioning of the system without shutting system down, or decreasing some criteria of performance that still can be farther corrected are considered as errors. Operator's actions that lead to the errors that render incapability to the system functioning, or shutting it down, or causes inability to achieve a required goal are considered as operator's or user failure. Specialist can talk about operator's and user errors or failures only when she/he discovers erroneous cognitive or behavior actions during task performance. In this abbreviated publication we focus on task complexity evaluation. There are interconnections between reliability and task complexity. Cognitive psychology treats cognition and activity in general as a process, making it difficult to study activity from a systemic-structural perspective. In contrast to cognitive psychology SSAT considers human activity not only as a process, but also as a structure that unfolded in time. This structure includes cognitive and behavioral components that can be performed in sequence or simultaneously, and can have various probabilities. Cognitive and behavioral actions and their components such as operations are used as units of analysis. In this work we utilize morphological analysis of computer based task. It includes algorithmic description of human activity and its time structure. Morphological analysis preceded quantitative analysis of activity during task execution. Basic quantitative methods of task analysis in SSAT are complexity and reliability of task performance [1]. Relationship between complexity and reliability of computer based task is discussed.

2 Complexity Evaluation of Human-Computer Interaction Task

As object of study we have selected a real production human-computer interaction task "receiving the orders". The considered task is performed just one time at the beginning of a shift. The purpose of this task is to receive the file containing orders into the local computer system from another distant computer. This is the key task the successful completion of which at the beginning of the shift allows creating work for 50–60 employees for the whole 8 h shift. Without successful completion of this task 50–60 employees with hourly pay cannot start their work. Unsuccessful completion of this task leads to loss of about 1 h for these. Loss of an hour is due to the fact that re-creation and resending of "the orders" file from computer that located in another state and in a different time zone required coordination of efforts with a number of specialists. This causes significant financial losses. Such negative event should be considered as a failure. The failures were more frequent on Mondays after 2 days off. Receiving the orders task is the first task performed by operator in the early morning shift. At the same time, it is a critically important task. Functional state of an operator is not yet at an optimal level. Optimal state is achieved only after some warm-up period. This factor can have negative effect on the pace of task performance and on the probability of emergence of erroneous actions.

All of these factors have been considered in our study and particularly when assessing the pace of task performance. Our study includes a qualitative analysis, preliminary algorithmic analysis of task performance, and evaluation of its reliability. The obtained data are further accompanied by a more detailed algorithmic description of the task in combination with analysis of the time structure of activity. It was necessary for farther evaluation of task complexity. At the final stage we have analyzed the relationship between such characteristics of task as complexity and its reliability.

Let us to present the very abbreviated qualitative description of the task. The operator checks if the file that contains the orders is on the list. If the file has been received the operator checks the date stamp on the file to make sure it is a new order file, not an old one. If the date stamp on the file is from today, then the task "receive the orders" is completed and the file can be processed using the software specifically design for this purpose, and the orders for current shift are distributed among all workers. If the file name is not on the list or does not have a current date on its date stamp, then the operator should restart the communication software that facilitates the file transfer process. The qualitative analysis of the task showed that the operator could not understand why the same actions that in most cases where given the desire result in some cases lead to the failure. The operator blamed computer specialists and insisted that computer system functioned incorrectly. Therefore, the purpose of this study was to determine the cause of the failure and find out how complexity and reliability influence each other in this task.

3 Algorithmic Description of the Existing Task and Time Structure Analysis

Contemporary tasks that combine cognitive and motor components have extremely variable structure. Outside of SSAT there are no methods of description and analysis for such tasks. Some authors even reject a possibility of analysis and description of such tasks. For example, it was suggested that the constraint-based approach to task analysis instead of the instruction-based approach. If we assume that an operator strictly follows normative prescriptions this is the instruction based or normative approach to the task analysis. The purpose of this approach is to specify all steps to achieve a required goal. The constraint-based approach presents only the requirements on what should be achieved and constraints on actions that should be used by operators. An operator is given no restriction on how the task should be accomplished. Consequently, an operator independently decide how to accomplish the task within the given constraints. Such approach contradicts with the basic principles of task analysis and design, with productivity and safety requirements. Constraint-based approach shows that there are no methods of task analysis and description of variable human activity in cognitive psychology. In contrast, SSAT offers the powerful method of analysis and description of extremely variable human activity during performance of any tasks. This method is known as algorithmic analysis of task performance. It is an important stage of task analysis in SSAT. The purpose of algorithmic description of activity is subdividing activity into qualitatively distinct psychological units and determining their logical organization. At this stage of analysis specialist utilize hierarchically organized units of analysis that includes: members of algorithm, cognitive and behavioral actions and their components-operations. Usually, the number of cognitive or behavioral actions in one member of algorithm is restricted by the capacity of short-term memory. Members of the algorithm are operators and logical conditions. Operators represent actions that transform objects, energy, and information. Logical conditions are members of the algorithm that describe cognitive decision-making actions. Each members of the algorithm are designated by standardize symbols. For example, standardized symbols which describe members of the algorithm are presented in the left column of the table. There are operators that are involved in receiving information, keeping information in memory, performing motor actions, (O^{α}; O^{μ}; O^{ε};) etc. Decision making or logical conditions are depicted by symbols such as $l \uparrow$ with a corresponding number of outputs. The arrow with the numbers depicts possible transitions to other member of the algorithm with the same numbers.

Symbolic standardized description in the left column is combined description of elements of activity by psychological units of analysis (see Table 1). The second column depicts combined description the same elements of activity utilizing common language or technological units of analysis. The third column utilizes detailed description using psychological units of analysis. MTM-1 system has been used in the third column to describe motions that are included into motor actions. Cognitive actions are also described in a standardize manner. Thus executive and cognitive components are described in totally new manner that has never been used outside of SSAT. The right column shows time that is required for performance of corresponding elements of activity. So, Table 1 depicts not only algorithmic description but also time structure of activity during task performance. Table 1 demonstrates that it is possible to describe extremely variable human activity that includes cognitive and behavioral components. This is a very informative model of activity during task performance.

Table 1 Algorithmic description of the existing method of task performance (fragment)

Members of the algorithm: psychological units of analysis	Description of the members of the algorithm (technological units of analysis)	Psychological units of analysis (detailed method of description)	Time (s)
O_1^{ε}	Type user name, press "*Enter*" Type password, press "*Enter*"	(1) Type ID (seven signs); (2) Move finger and press "*Enter*"—(AP1); (3) Type password (ten signs); (4) Move finger and press "*Enter*"—AP1	8.98
O_2^{ε}	Type command "*ls -l*" and press "*Enter*".	(1) Move finger and press "*l*" (AP1); (2) Type "*s -l*" (three signs); (3) Move finger and press "*Enter*" (AP1)	2.4
O_3^{α}	Check to see if the file "*orders*" is on the list (distance of top-down eye movement is 4 cm)	Simultaneous perceptual action combined with scanning operation of target item; eye movement; time for scanning is calculated using coefficient 1.1 (EF + ET + EF)	1.24
$l_1 \overset{1}{\uparrow}$	If the file is on the list go to O_4^{α}. ($P_1 = 0.8$). If the name of the file is absent from the list go to O_5^{ε} ($P_2 = 0.2$). P = 1	Simple decision making action from two alternatives at sensory-perceptual level ("Yes"-"No" type decision)	0.33
$O_4^{\alpha\mu}$	Check to see if the date stamp of the file "*orders*" has current date (move eyes 3 cm, read the date and compared with the current date that is retrieved from memory). P = 0.8	Simultaneous perceptual action combined with mnemonic operation (ET + EF)	0.39
$l_2 \overset{2}{\uparrow}$	If the file has today's date then received file is the expected file, go to O_8^{ε} ($P_1 = 0.76$). If the file has the old date, then go to O_5^{ε}. ($P_2 = 0.04$). P = 0.8	Simple decision making action from two alternatives at sensory-perceptual level ("Yes"-"No" type decision)	0.26
$\overset{1}{\downarrow} O_5^{\varepsilon}$	Type "*restore_communication*" and press "*Enter*". Repeat if necessary. P = 0.2 + 0.04 = 0.24	Type "*restore_communication*" (twenty signs); move finger and press "*Enter*" (AP1)	2.1
$O_6^{\alpha w}$	Waite until initial screen comes up. P = 0.24	Waiting period requires minimum level of attention concentration	1.92
$^*\boldsymbol{O_7^{\alpha w}}$		Waiting period includes	(1) 0.43

(continued)

Table 1 (continued)

Members of the algorithm: psychological units of analysis	Description of the members of the algorithm (technological units of analysis)	Psychological units of analysis (detailed method of description)	Time (s)
(1) $(O_7^{\alpha w})$ (2) $(O_7^{cal\mu w})$ (3) $(O_7^{\varepsilon w})$ (4) $({}_{ch}O_7^{\alpha w} + l_7^{\mu w})$ These members of sub-algorithm are repeated 3 times during waiting period	Make a note of the time of completion of $O_6^{\alpha w}$. Waite for ≈7 min The most preferable strategy of performance includes (1) move the site to the right bottom corner of the screen and read the time; (2) add 7 min and remember this time until the time is written down; (3) take a pen, move hand with pen to the paper, wright down the target time and put the pen back; (4) checking actions that compare current time with the target time (the checking actions can be repeated 2–3 times); the rest of the time is pure waiting period. P = 0.24	(1) Simultaneous perceptual action (ET + *EF*); (2) arithmetic calculation combined with mnemonic operation ($O_7^{cal\mu w}$) = 2 s; (3) (R30B + G1A) + (M30B + 2 s) + (M13A + RL1) = 3.33 s; (4) the first perceptual action (ET + *EF*) = 0.54 s; the second perceptual action, including mnemonic operation and comparing numbers (ET + *EF* + *EF*) = 1.04 s; decision-making based on visual information (*EF*) = 0.5 s	(2) 1.58 (3) 2.64 (4) 1.65 $\sum$ 6.3
${}_1O_2^{\varepsilon}$	Type command "*ls -l* " and press "*Enter*" (the same as O_2^{ε}) P = 0.24	Move finger and press "*l*" (AP1); type "*s -l*" (three signs); move finger and press Enter (AP1)	0.576
${}_1O_3^{\alpha}$	Check if the file "*order*" is on the list (distance of top-down eye movement is 4 cm) (the same as O_3^{α}). P = 0.24	Simultaneous perceptual action combined with scanning operation for a target item; eye movement; time for scanning is calculated using coefficient 1.1 (EF + ET + EF)	0.3
${}_1l_1\overset{1}{\uparrow}_1$	If the file is on the list go to ${}_1O_4^{\alpha}$. (P_1 = 0.12). If the name of the file is absent from the list go to ${}_1O_5^{\varepsilon}$ (the same as l_1). (P_2 = 0.12). P = 0.12 + 0.12 = 0.24	Simple decision making action from two alternatives at sensory-perceptual level ("Yes"-"No" type decision)	0.08
${}_1O_4^{\alpha\mu}$	Check if the date stamp of the file "*orders*" has the current date (move eyes 3 cm, read the date and compared with the current date that is retrieved from memory), the same as $O_4^{\alpha\mu}$. P = 0.12	Simultaneous perceptual action combined with mnemonic operation (ET + EF)	0.06
${}_1l_2\overset{2}{\uparrow}_2$	If the file has today's date then received file is the expected file, go to O_8^{ε} (P_1 = 0.06); If the file has the		0.04

(continued)

Table 1 (continued)

Members of the algorithm: psychological units of analysis	Description of the members of the algorithm (technological units of analysis)	Psychological units of analysis (detailed method of description)	Time (s)
	old date, then go to ${}_1O_5^{\varepsilon}$ (the same as l_2) ($P_2 = 0.06$). P = 0.06 + 0.06 = 0.12	Simple decision-making action from two alternatives at sensory-perceptual level ("Yes"-"No" type decision)	
${}_1\overset{1}{\downarrow}{}_1O_5^{\varepsilon}$	Type "*restore_communication*" and press "*Enter*". Repeat if necessary (the same as O_5^{ε}). P = 0.12 + 0.06 = 0.18	Type "*restore_communication*" (twenty signs); move finger and press "*Enter*" (AP1)	1.6
${}_1O_6^{\alpha w}$	Waite until initial screen comes back (the same as $O_6^{\alpha w}$). P = 0.18	Waiting period requires the minimal level of attention concentration	1.44
${}_1O_7^{\alpha w}$ (1) ($O_7^{\alpha w}$) (2) ($O_7^{cal\mu w}$) (3) ($O_7^{\varepsilon w}$) (4) ($O_7^{\alpha w} + l_7^{\mu w}$) These members of sub-algorithm are repeated 3 times during waiting period	Make a note of the time of completion of $O_7^{\alpha w}$. Waite for ≈7 min. The most preferable strategy of performance includes (1) move the site to the right bottom corner of the screen and read the time; (2) add 7 min and remember this time until the time is written down; (3) take a pen, move hand with pen to the paper, wright down the target time and put the pen back; (4) checking actions that compare current time with the target time (the checking actions can be repeated 2–3 times); the rest of the time is pure waiting period. P = 0.18	Waiting period includes (1) Simultaneous perceptual action (ET + *EF*); (2) arithmetic calculation combined with mnemonic operation ($O_7^{cal\mu w}$) = 2 s; (3) (R30B + G1A) + (M30B + 2 s) + (M13A + RL1) = 3.33 s; (4) The first perceptual action (ET + *EF*) = 0.54 s; the second perceptual action, including mnemonic operation and comparing numbers (ET + *EF* + *EF*) = 1.04 s; decision-making based on visual information (*EF*) = 0.5 s	(1) 0.32 (2) 1.19 (3) 1.98 (4) 1.23 Σ 4.72
${}_2O_2^{\varepsilon}$	Type command "*ls -l* " and press "*Enter*" (the same as O_2^{ε}). P = 0.18	Move finger and press "*l*" (AP1); type "*s -l*" (three signs); move finger to corresponding key and press "*Enter*" (AP1)	0.43
${}_2O_3^{\alpha}$		Simultaneous perceptual action combined with scanning operation of target item; eye movement;	0.22

(continued)

Table 1 (continued)

Members of the algorithm: psychological units of analysis	Description of the members of the algorithm (technological units of analysis)	Psychological units of analysis (detailed method of description)	Time (s)
	Check to see if the file "*orders*" is on the list (distance from the screen 45 cm; distance of top-down eye movement is 4 cm) (the same as O_3^{α}). P = 0.18	time for scanning is calculated using coefficient 1.1 (EF +ET + EF)	
${}_{2}l_{1}\overset{1}{\uparrow}_{2}$	If the file is on the list go to ${}_{2}O_4^{\alpha\mu}$; ($P_1 = 0.09$). If the name of the file is absent from the list go to O_9^{ε} (the same as l_1), ($P_2 = 0.09$). P = 0.18	Simple decision making action from two alternatives at sensory-perceptual level ("Yes"-"No" type decision)	0.059
${}_{2}O_4^{\alpha\mu}$	Check to see if the date stamp of the file "*orders*" has the current date (move eyes 3 cm, read the date and compared with the current date that retrieved from memory), the same as $O_4^{\alpha\mu}$. (P = 0.09)	Simultaneous perceptual action combined with mnemonic operation (ET + EF)	0.044
${}_{2}l_{2}\overset{2}{\uparrow}_{2}$	If the file has today's date then received file is the expected file, go to O_8^{ε} ($P_1 = 0.045$); If the file has the old date, then go to O_9^{ε} (the same as l_2), ($P_2 = 0.045$). P = 0.09	Simple decision making action from two alternatives at sensory-perceptual level ("Yes"-"No" type decision)	0.03
$\overset{2}{\downarrow}_{2}\overset{1}{\downarrow}_{2}\overset{2}{\downarrow}O_8^{\varepsilon}$	Type command "*interface_orders*". The end of the considered task (**the goal is achieved**) and the beginning of the following task. (P = 0.76 + 0.06 + 0.045) = 0.865	Type command with 16 characters; move finger and press "*Enter*" (AP1)	6.31
${}_{2}\overset{1}{\downarrow}_{2}\overset{2}{\downarrow}O_9^{\varepsilon}$	Call computer specialist (**Failure**). This member of algorithm is excluded from analysis. P = 0.09 + 0.045 = 0.135	Consists of the sequence of verbal and motor actions. The sequence and number of actions are not strictly defined	– –

Total performance time of the whole task is 39.829 s

4 Evaluation of Complexity of the Existing Method of Task Performance

Based on the analysis of Table 1 we can choose developed in SSAT measures of task complexity evaluation. The first measure determines duration of activity during task performance. Time for the algorithm execution (total time of task performance) is determined according to Formula 1:

$$\mathrm{T} = \sum \mathrm{P_i t_i}. \tag{1}$$

where P_i—probability of the i-th members of the algorithm, t_i—performance time of the i-th member of algorithm. Total performance time for the considered task is T = 39.829 s. At the next step it is necessary to determine duration of separate components of activity that are classified as follows: perceptual components; decision-making components (logical conditions); components that require to retain current information in working memory; executive components of activity; all cognitive components of activity. Duration of perceptual components of activity is determined using Formula 2

$$\mathrm{T}_\alpha = \sum \mathrm{P}^\alpha \mathrm{t}^\alpha. \tag{2}$$

where P^α—probability of afferent or perceptual components of activity, t^α—performance time of afferent components of activity. The right column of Table 1 already has performance time of member of the algorithm taking into account probability of each member of this algorithm and its components. Therefore, farther we can determine duration of considered elements of activity simply by summarizing their performance times. So, duration of perceptual components of activity is determined utilizing Formula 3:

$$\begin{aligned} \mathrm{T}_\alpha = {} & \mathrm{T}(O_3^\alpha) + \mathrm{T}(O_4^{\alpha\mu}) + \mathrm{T}(O_6^{\alpha w}) + \sum \mathrm{T}({}^*O_7^{\alpha w}) + \mathrm{T}({}_1O_3^\alpha) + \mathrm{T}({}_1O_4^{\alpha\mu}) \\ & + \mathrm{T}({}_1O_6^{\alpha w}) + \sum \mathrm{T}({}_1^*O_7^{\alpha w}) + ({}_2O_3^\alpha) + \mathrm{T}({}_2O_4^{\alpha\mu}) = 8.544\,\mathrm{s}, \end{aligned} \tag{3}$$

where $\sum \mathrm{T}({}^*O_7^{\alpha w})$ and $\sum \mathrm{T}({}_1^*O_7^{\alpha w})$ are summarized time of performing perceptual actions during waiting period.

Duration of included in this task thinking components of activity is determined as:

$$\mathrm{T_{th}} = \mathrm{T}\left(O_7^{\mathrm{cal}\mu w}\right) + \mathrm{T}({}_1O_7^{\mathrm{cal}\mu w}) = 2.77\,\mathrm{s}. \tag{4}$$

Duration of decision-making components (logical conditions) in task performance is:

$$Lg = (l_1) + (l_2) + (l_7^{\mu w}) + ({}_1l_1) + ({}_1l_2) + ({}_{\mathbf{1}}l_7^{\mu w}) + ({}_2l_1) + ({}_2l_2) = 1.499\,\mathrm{s} \quad (5)$$

Duration of retaining current information in working memory is determined according to the following formula:

$$\begin{aligned} \mathrm{T_{wm}} = \mathrm{T}(O_4^{\alpha\mu}) + \mathrm{T}\left(O_7^{\mathrm{cal}\mu w} + l_7^{\mu w}\right) + \mathrm{T}({}_1O_4^{\alpha\mu}) \\ + \ \mathrm{T}({}_{\mathbf{1}}O_7^{\mathrm{cal}\mu w} + {}_{\mathbf{1}}l_7^{\mu w}) + \ \mathrm{T}({}_2O_4^{\alpha\mu}) = 3.964\,\mathrm{s} \end{aligned} \quad (6)$$

Total duration of cognitive components in task performance is calculated as follows:

$$\mathrm{T_{cog}} = \mathrm{T}_\alpha + Lg + \mathrm{T_{th}} + (T_{wm}) = 12.813\,\mathrm{s} \quad (7)$$

Duration of executive components of activity (performance time of efferent operators):

$$\begin{aligned} \mathrm{T_{ex}} = \mathrm{T}(O_1^{\varepsilon}) + \ \mathrm{T}(O_2^{\varepsilon}) + \mathrm{T}(O_5^{\varepsilon}) + \mathrm{T}(O_7^{\varepsilon w}) + \mathrm{T}({}_1O_2^{\varepsilon}) + \mathrm{T}({}_1O_5^{\varepsilon}) \\ + \mathrm{T}({}_{\mathbf{1}}O_7^{\varepsilon w}) + \mathrm{T}({}_2O_2^{\varepsilon}) + \mathrm{T}(O_8^{\varepsilon}) = 27.016\,\mathrm{s}. \end{aligned} \quad (8)$$

And T_{cog} + T_{ex} = 39.829 s is the execution time of the whole task consists of these two components. Some elements of activity are counted in several formulas. For example, performance time of such members of algorithm as $O_7^{\mathrm{cal}\mu w}$ and ${}_{\mathbf{1}}O_7^{\mathrm{cal}\mu w}$ is counted as in T_{th} and in T_{wm}. These members of the algorithm include different cognitive operations such as arithmetic calculation (thinking component) and mnemonic component. Similarly, there are several other members of the algorithm that include different cognitive operations. Obtained data allowed to calculate other measures of complexity.

The first one of them is N_α. This measure presents fraction of perceptual components of activity in the performance of the entire task.

$$\mathrm{N}_\alpha = \mathrm{T}_\alpha/\mathrm{T} = 0.215. \quad (9)$$

Fraction of thinking components of activity in the performance of the entire task is calculated as follows:

$$\mathrm{N_{th}} = \mathrm{T_{th}}/\mathrm{T} = 0.07 \quad (10)$$

Fraction of decision-making components of activity in the performance of the entire task is calculated utilizing Formula 11:

$$\mathrm{N_j} = \mathrm{Lg}/\mathrm{T} = 0.0376. \quad (11)$$

The level of working memory workload during the task performance is:

$$N_{wm} = T_{wm}/T = 0.0995. \quad (12)$$

The fraction of external behavioral (executive) components of activity in the entire task is:

$$Nex = Tex/T = 0.678 \quad (13)$$

The next measure determines fraction of cognitive components of activity in the performance of the entire task:

$$Ncog = Tcog/T = 0.322. \quad (14)$$

In the considered task an operator cannot predict the success and failure of some of her/his actions. The same cognitive and behavioral actions to restore communication can result in success or failure. Such uncertainty produces emotional tension and increased level of concentration of attention or level of mental efforts. In normal conditions measures that are associated with attention concentration during performance of computer based tasks are related to the third category of complexity. However, the period of time when operator attempts to restore communication for the second time produces emotion tension because a possibility of failure sharply increases. To carry out the necessary calculations it is required to determine time that is needed for achieving the goal or encounter failure after performing ${}_1l_2$. This is the period of time when an operator attempts to restarting communication for the second time. We calculate this time that is required for cognitive and motor components of activity separately because belongs to the fourth category of complexity.

Performance time for perceptual components that are of the fourth category of complexity are calculated as:

$$T_{\alpha(4)} = T({}_1O_6^{\alpha w}) + \sum T({}_1^{*}O_7^{\alpha w}) + ({}_2O_3^{\alpha}) + T({}_2O_4^{\alpha\mu}) = 3.434\,s. \quad (15)$$

At the next step time for decision-making (logical conditions) of the fourth category of complexity is calculated as:

$$Lg_{(4)} = ({}_1l_2) + ({}_{\mathbf{1}}l_7^{\mu w}) + ({}_2l_1) + ({}_2l_2) = 0.429\,s. \quad (16)$$

Finally, time for thinking components that are related to the fourth category is defined:

$$T_{th(4)} = T({}_{\mathbf{1}}O_7^{cal\mu w}) = 1.19\,s. \quad (17)$$

Time for mnemonic operations that are performed simultaneously with other cognitive components is not considered in our calculations.

Time for all cognitive components of activity that are related to the fourth category of complexity is calculated according Formula 18:

$$T_{cog(4)} = T_{\alpha(4)} + Lg_{(4)} + T_{th(4)} = 5.053\,\text{s}. \tag{18}$$

Time for mnemonic operations that are performed simultaneously with other cognitive components is not included separately.

Time for executive (motor) components of activity that are related to the fourth category of complexity are determined similarly:

$$T_{ex(4)} = T({}_1O_5^{\varepsilon}) + T({}_1O_7^{\varepsilon w}) + T({}_2O_2^{\varepsilon}) + T(O_8^{\varepsilon}) = 10.98\,\text{s}. \tag{19}$$

Next, we determine the ratio of time spent performing cognitive component related to the fourth category of complexity to the time for all cognitive components of the task:

$$T_{cog(4)}/T_{cog} = 5.053/12.813 = 0.394. \tag{20}$$

Finally, we determine the ratio of performance time of executive (motor) component related to the fourth category of complexity to the time for all motor components of the task:

$$T_{ex(4)}/T_{ex} = 0.406. \tag{21}$$

The last two formulas demonstrate the following. There is an order scale of complexity with 5 levels of attention concentration. The simplest one is related to the first category and the most complex is of the fifth category of complexity. The last two formulas demonstrate that according to this criterion fraction of cognitive components related to the forth category is 0.36, and fraction of the third category of complexity is 0.64.

Fraction of forth category of complexity of motor component is 0.39 and fraction of the third category is 0.61. Thus there is the sufficiently high level of mental efforts during the task performance according to this criterion. Specifically, it is true for the executive or motor components of activity. This is due to the fact that for the cognitive components minimum level of concentration of attention is of the third category and maximum level is of the fifth category. In contrast, motor activity in normal conditions for various manual work requires the first level of attention concentration for the simplest type of motor activity, and the most complex motor activity is related to the third category. Motor activity of a user during interaction with computer, when he/she utilizes the keyboard, usually is of the third category of complexity. However, because of stress factor and derived from it emotional tension significant part of motor activity is transformed into fourth category. Therefore, interaction of user with the keyboard in stressful conditions requires high level of mental efforts.

5 Conclusion

Complexity and reliability of task performance can influence each other. An increase in equipment precision can often be accompanied by increasing the complexity of the task. Task in more precise systems often become more complex and its performance less reliable. Probability of operator's erroneous actions might also increase. Thus, complexity and reliability of task performance are interdependent characteristics. Increasing of reliability of task performance can be achieved by introducing new cognitive and behavioral actions. This new components of activity can be a source of errors or failures.

When considering the reliability issues it is of particular importance to take into account operator efficiently during the entire shift, time limit, stressful situations, etc. In other words, operators have to work reliably in the conditions on which the system is designed. For example, an operator can demonstrate reliable performance in the beginning of the shift and not so reliable towards the end of the shift. Increasing the complexity of the task is one of the most important sources of mental fatigue. This factor is especially important to take into account at the end of the shift. Therefore, the more complex the task is the less reliable its performance can be at the end of the shift. Analysis of complexity measures allows to detect potentially critical points in task performance which can be a source of such errors. All this suggests that the evaluation of the reliability and complexity of the task in some cases should be carried out together.

References

1. Bedny, G., Karwowski, W., Bedny, I.: Applying Systemic-Structural Activity Theory to Design of Human-Computer Interaction System. CRC Press, Boca Raton (2015)
2. Bedny, G.: Application of Systemic-Structural Activity Theory to Design and Training. CRC Press, Boca Raton (2015)
3. Kotik, M.A., Yemelyanov, A.M.: Nature of Human-Operator's Errors. Transport Publisher, Moscow (1993)

Windshield Frame Shape and Awareness of the External World While Driving an Automobile

Yusaku Takeda, Koji Iwase, Toshihiro Hara, Atsuhide Kishi, Kazuo Nishikawa, Richard Veale, Masatoshi Yoshida, Tadashi Isa and Takahide Nouzawa

Abstract The vehicle windshield is supported and framed by the hood, roof, and pillars, which occlude the driver's view of the outside. It has been previously shown that awareness of the external world changes according to differences in windshield shape. This directly affects the drivability of a vehicle. Thus, the windshield shape must be designed by considering driver's visual performance so that it can be balanced with other performance measures such as weight and roominess to design the optimal cockpit. Visual performance during driving is affected by (1) bottom-up attention and (2) top-down attention, and (3) selection between them. This study

Y. Takeda (✉) · K. Iwase · T. Hara · A. Kishi · K. Nishikawa · T. Nouzawa
Technical Research Center, Mazda Motor Corporation,
3-1, Shinchi, Fuchu-cho, Aki-gun, Hiroshima 730-8670, Japan
e-mail: takeda.yus@mazda.co.jp

K. Iwase
e-mail: iwase.k@mazda.co.jp

T. Hara
e-mail: hara.tos@mazda.co.jp

A. Kishi
e-mail: kishi.a@mazda.co.jp

K. Nishikawa
e-mail: nishikawa.ka@mazda.co.jp

T. Nouzawa
e-mail: nouzawa.t@mazda.co.jp

R. Veale · M. Yoshida · T. Isa
Department of Developmental Physiology, National Institute for Physiological Sciences, Division of Behavioral Development,
38 Nishigonaka Myodaiji, Okazaki, Aichi 444-8585, Japan
e-mail: richard@nips.ac.jp

M. Yoshida
e-mail: myoshi@nips.ac.jp

T. Isa
e-mail: tisa@nips.ac.jp

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_32

focuses on Itti and Koch's visual saliency that attracts bottom-up attention as a visual scene changes in shape and color around front windshield frame during driving. This paper aims to quantify the relationships between drivers' gaze movements and visual saliency.

Keywords External awareness · Visual shapes and properties of the windshield frame · Visual saliency while driving a vehicle

1 Introduction

We look through the windshield when driving a vehicle. The windshield is supported and framed by the hood, roof, and pillars, which occlude the driver's view of the outside. It has been previously shown that recognition of objects outside the car changes according to differences in windshield shape, and affects drivability. Thus, the relationship between windshield shape and driver visual performance must be quantified so that it can be balanced with other performance measures such as weight and roominess to design the optimal cockpit. Visual performance during driving consists of role-sharing between (1) bottom-up attention and (2) top-down attention, and the processes selecting between them.

Many psychological and computational studies on bottom-up attention have been reported. For example, Treisman and Gelade [1] proposed the feature integration theory of visual attention. They explained psychologically that visual stimuli are simultaneously processed in independent feature channels (e.g., intensity, color, orientation and motion) and are finally integrated into a 'master map'. Furthermore, Koch and Ullman [2] proposed a 'saliency map' that expresses how bottom-up attention is calculated computationally. Since then, several researchers have measured gaze movements and quantified the relationship between gaze movements and visual stimulus characteristics quantitatively. Furthermore, Itti and Koch [3] proposed a computational algorithm to compute the saliency map while subjects look at a still image or video. This algorithm estimates the time needed to find a target in a visual search paradigm by assuming that the subject will look at items in order of how salient they are, while inhibiting return to places already looked. This saliency map model fits the empirical results collected in psychophysical studies with regard to the looking behavior of subjects. On the neural side, Yoshida and colleagues [4, 5] have investigated how visual feature channels are re-weighted after certain types of brain damage, and what that implies about how and where the saliency map is actually implemented in the brain. Furthermore, Yoshida and colleagues [6] have developed neural network models to more accurately understand the physiological mechanisms responsible for saliency map computation in humans. For example, they have developed a model of the superior colliculus (a midbrain area responsible for visual processing and eye movements) based on the saliency map model which predicts how electrical microstimulation applied to the superior colliculus will affect the targets of looking behavior.

In parallel to the academic research, there is a large body of research in the auto-motive field regarding the mechanisms behind human visual awareness. Tanaka and colleagues [7] have proposed a method to quantitatively evaluate the best interior design of a car during high-speed driving by measuring the changes in eye position and head/neck movement based on seat posture combined with measurements of torque in the cervical joint. Similarly, Takeda and colleagues [8, 9] have demonstrated that placing the A pillar and instrument panel in appropriate locations has a positive effect on the driver's sense of safety with respect to the world outside the car. They furthermore demonstrated experimentally that the choice of good locations and shapes for the pillar and instrument panel is related to the sense of distance felt by the driver.

Finally, there is a large and venerable body of research in psychology relating to the mechanisms behind looking behavior which are highly relevant to our current study of driver's awareness of the outside world [10, 11]. For example, Yarbus [10] famously showed that the pattern of looking behavior of a subject to the exact same visual scene changes based on the subject's goal (for example, "tell me what is happening" versus "tell me how many people are in the room"). On a related note, Laretzaki [11] importantly showed that looking behavior changes based on subjects' psychological states. He showed that as subjects become more distressed and uneasy, the distribution of the places they look likewise becomes wider. This thread of research has highlighted the importance of a good automobile interior design and layout to increase the driver's sense of safety. However, until now, there has been little research to elucidate *why* looking behavior is influenced by these seemingly aesthetic factors, and this has had a negative effect on the design of car interiors. The current research is a step towards quantifying the effect of interior design/layout using computational models of visual attention and human behavioral research.

As a first step, this study focuses on saliency characteristics that attract bottom-up attention in relation to visual stimulus changes in the shape and color around front windshield frame. This paper aims to quantify the relationships between drivers' gaze movements and saliency characteristics of the front windshield and external world as viewed through the front windshield.

In this report, we measured changes in gaze movement that result from manipulation of the shape and color of the front windshield, using an experimental apparatus we designed in-house. This apparatus displays movies in which a 3-dimensional front windshield design is superimposed on the front view of car driving scene. We quantified the relationship between the measured gaze trajectories and the main bottom-up attention factors (color, intensity, orientation, and motion) using a saliency map model.

2 Analysis of Visual Recognition in Driving

2.1 Experimental Equipment

Figure 1 depicts an experimental system measuring visual awareness when driving a vehicle. The system is composed of a gaze measurement device, which can measure position of both the subject's eyes, and a visual display in which a movie of the view through the front windshield is displayed to the subject. The gaze measurement uses an eye mark recorder (NAC Image Technology Co., Ltd.; Glass-wear formula eye camera; Measurement range: horizontal direction ±40°, vertical direction ±20°; eye movements resolution: horizontal direction 0.1°, vertical direction 0.4°; pupil diameter resolution: 0.02 mm; sampling: 240 Hz). This records the subject's voice in synchronization with measured gaze direction.

The experimental setup furthermore contains a mockup of the driver's seat of an automobile, including acceleration and brake pedals, and footrest. The display shows a test movie of the front view through the windshield during driving, using a Tft active matrix display (TOSHIBA CORPORATION Co., Ltd.; Display resolution: 3840 × 2160 (4K); Display size: 58 in.; Display devise: active-matrix) as shown in Fig. 1. The measurements from the eye recorders are synchronized with the visual display.

2.2 Experimental Methods

Figure 2 shows the experimental method for displaying the visual stimulus (a movie through the front windshield) to the subject and measuring his/her eye movements. As shown in Fig. 2a, a movie previously recorded in a car driving on a public

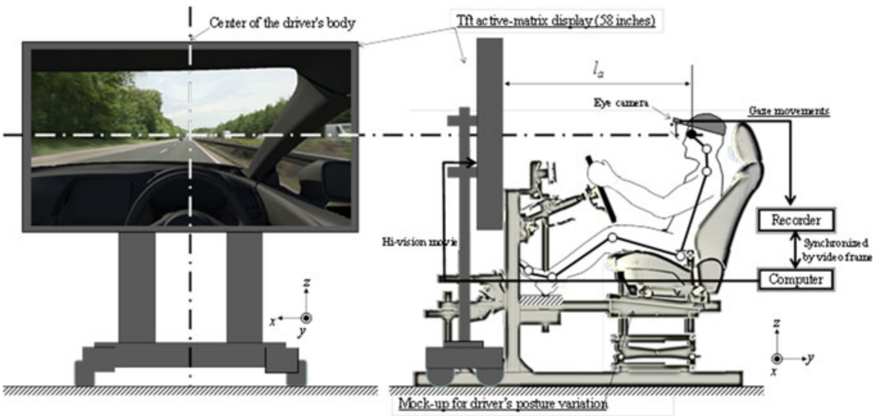


Fig. 1 Experimental system for measuring visual awareness while driving a vehicle

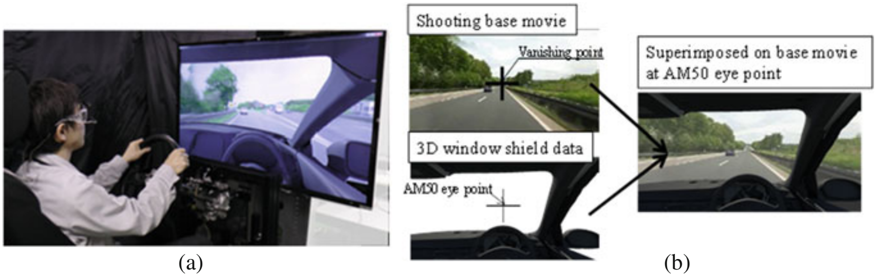


Fig. 2 Experimental design: the subject watches the displayed video of driving scenery, and the movie vanishing point and CAD data is used to superimpose the base movie based on AM50 (U.S. male 50th percentile) data. **a** A subject driving the virtual simulator. **b** How to make a presentation movie

highway is displayed to the subject at a distance of 1 m on the display. The viewpoint of the visual scenery displayed is that of an average 50th percentile male driver from the U.S., as estimated from CAD data of the car model [12] and the vanishing point of the video taken (Fig. 2b). The driver's seat mockup from Fig. 1 is adjusted to match the subject's viewpoint so that the vanishing point is appropriate for the driver. To give the subject an even stronger sense of "driving", the subject is told to turn the steering wheel to match the perceived direction of travel of the vehicle.

The experiments are comprised of 4 conditions, being 4 different 60 fps high definition videos of 30 s each. These are shown to the subjects in 3 trials (Total: 90 s), during which the subject's visual behavior is measured. The experimental conditions are designed to investigate the differences in salience based on the A pillar, front header, and instrument panel shape. Table 1 shows that the experimental conditions (A)–(D) are manipulations of the windshield frame shape and color.

Table 1 Experimental conditions

(A) No windshield	(B) Black reverse trapezoidal frame
	
(C) Black rectangular frame	(D) White rectangular frame
	

Experimental conditions (*A*)–(*D*) are manipulations of the windshield frame shape and color. The *shaded region* represents portions where the driving video is displayed. (*B*)–(*D*) have the superimposed windshield frame

2.3 Gaze Movements in Relation to Windshield Frame Properties

Figures 3, 4, 5, 6 and 7 show the measured gaze behavior of 4 subjects. To mimic natural driving conditions, the heads of the subjects were not fixed and gaze behavior thus represents combined head, eye, and trunk movements. An important caveat is that the subjects are employees of MAZDA corporation and thus experienced drivers.

Figures 3, 4, 5 and 6 show the per-subject eye movements for the 4 corresponding experimental conditions (A)–(D). The figures show 480 by 720 pixel still frames of the viewed scenery. The blue lines represent changes in eye position during the corresponding 30-s trial. This data is quantified later, in Fig. 7.

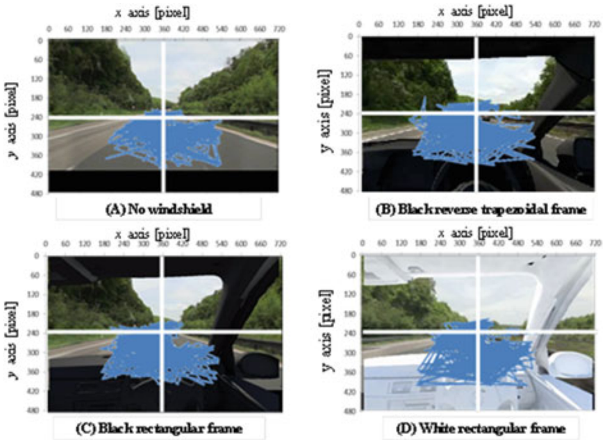


Fig. 3 Subject A

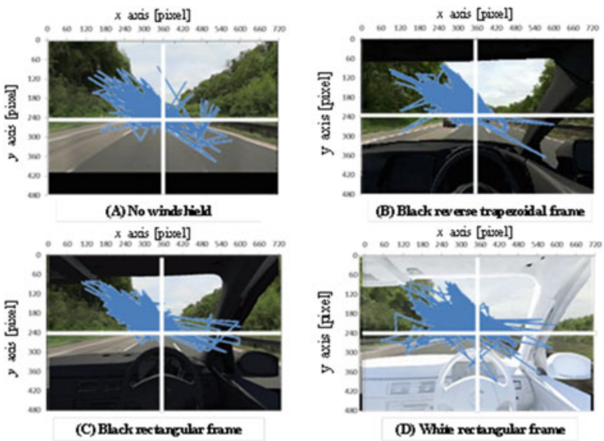


Fig. 4 Subject B

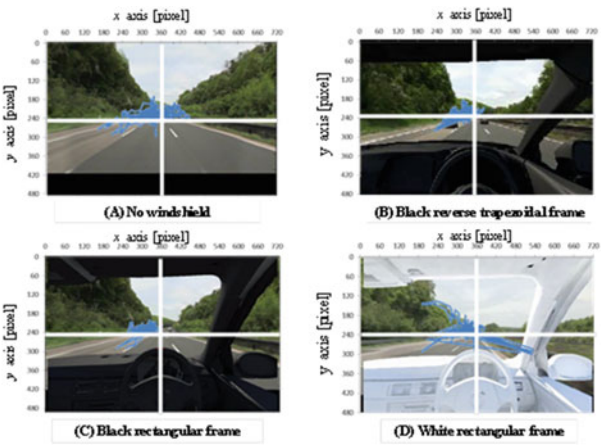


Fig. 5 Subject C

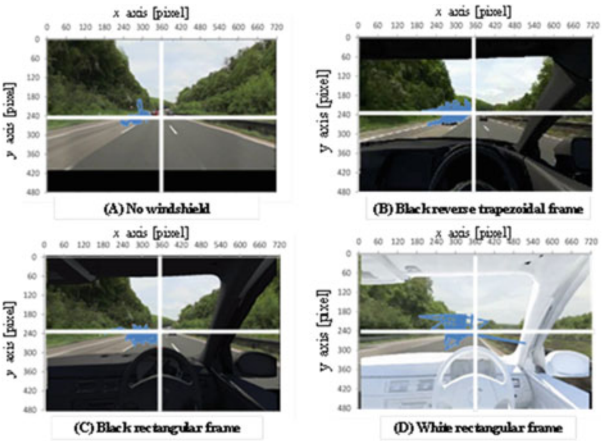
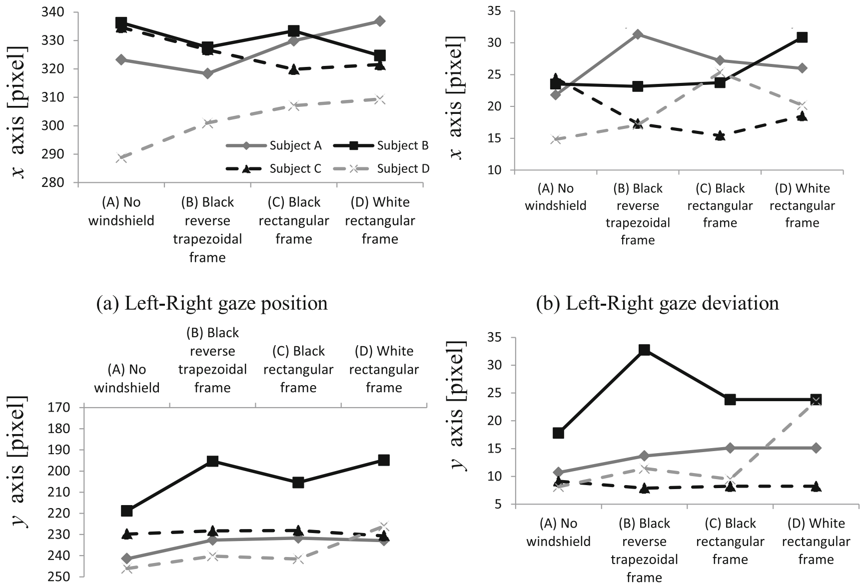


Fig. 6 Subject D

In with-frame experimental conditions (B)–(D), due to the existence of the windshield frame, attention is more easily pulled to the upper regions and attention is more responsive to movement in the outside world. In contrast in the no-frame condition (A), the subject's visual regard is centered and distributed around the lower regions of the visual scene. In conditions (B) and (C), the looking distribution expands to be distributed higher than in the opposite condition. In experimental subject A and B, we can see a tendency for gaze to respond to the A pillar in the reverse trapezoid condition. By comparing conditions (C) and (D), we can see that the black colored windshield frame allows gaze to be directed to important points more readily, and gaze is concentrated more in the upper portion of the frames. As



(a) Left-Right gaze position (b) Left-Right gaze deviation

(c) Up-Down gaze position (d) Up-Down gaze deviation

Fig. 7 Analysis of each subject's eye movement behavior: gaze position moves up in frame conditions, and the left-right distribution is suppressed. Up-down distribution is likewise suppressed. **a** Left-Right gaze position. **b** Left-Right gaze deviation. **c** Up-Down gaze position. **d** Up-Down gaze deviation

in comparison between the no-frame condition and the reverse trapezoid, we can see a tendency for the subject's gaze to respond to the black A pillar.

Figure 7 shows the distribution of eye movements around the average position, and the standard deviation, for each subject. Subfigure (a) shows the mean left-right distribution of eye position for each subject, in each condition, and (b) shows the standard deviation. Note that the center of the image is pixel 360 (since it is 720 pixels wide), thus the center is slightly biased to the left of the image. Comparing conditions (A) and (B), there is some bias to be pushed away from the A pillar on the right side of the frame. Furthermore, comparing conditions (B) and (C), the tendency of eye position to be pushed away from the A pillar is obvious when one observes that the eye position approaches the pillar side on the right more in the rectangular condition than in the reverse trapezoid. Furthermore, comparing conditions (C) and (D), the less luminant the interior/frame materials show the greatest separation of center of gaze from the A pillar. Figure (b) shows the left-right standard deviation. The large these values, the more subjects tended to distribute gaze along the left-right dimension. From these results, we can see that the windshield frame's existence or non-existence allows us to separate the results into two groups, with different levels of left-right gaze movement. The data naturally separates into two groups, the rectangular condition (with fewer left-right eye movements than the reverse trapezoid condition), and the condition with more eye movements. Furthermore, the less luminant of the interior material, the fewer the left-right eye movements, and vice versa.

Figure 7c shows the center of gaze distribution on the vertical axis. In this case, the vanishing point was at pixel 240. We can see from the figure that all subjects had their gaze center shift upward with the addition of the window frame. Furthermore, in the rectangular condition gaze is directed more towards the vanishing point, and if the interior is more luminant, the center of gaze is shifted upwards. Figure (d) shows the variance of up-down gaze position. The existence of the windshield frame seems to increase gaze shifts with a vertical component. We can separate the data into two groups based on the variance of gaze position in the vertical direction, and the interior luminance.

From these results, we can see that conditions including the windshield frame result in the subjects looking most at the most important region for driving (straight ahead). Furthermore, the shape of the frame (the angle of the A-pillar) has an influence on the vertical distribution of gaze and its likelihood to become fixed on the vanishing point. In the next section, we will evaluate these 3 points with relation to the saliency map model of bottom-up visual attention.

3 Numerical Analysis of Visual Awareness While Driving

3.1 Evaluation Procedure

We use the bottom-up visual attention model called the "saliency map" to analyze the visual stimuli presented to the experimental subjects. We importantly analyze the relationship of the main important factors of hue (color), luminance, orientation, and motion. Figure 8 introduces the basic concepts between saliency map computations.

In the left of the three visual stimulus examples shown, one element "pops out" due to difference along at least one of these visual feature channels. However, in the rightmost, it does not pop out along any single feature dimension and so is more difficult to locate. In the leftmost image, the element differs by its color, in the middle example, by its orientation. In Triesman's account [1], these channels would be computed each in parallel, and then combined into a single "pop out" map, as shown at the bottom of the figure. In the rightmost example in the figure, the feature search is not straightforward, as the search for the target that meets all criteria (e.g. a conjunctive search of both vertical orientation and blue color) is made difficult by the existence of elements that match in one feature channel but not the other. This was a psychological theory, but has been made quantitative by the development of computational models such as the saliency map. In this paper, we use a saliency map with strengthened color channels due to the blindsight studies of Yoshida et al. [4], and investigate the relationship between subject's eye movements in the previous section and the output of a saliency map model when it is run with the driving visual stimuli as input.

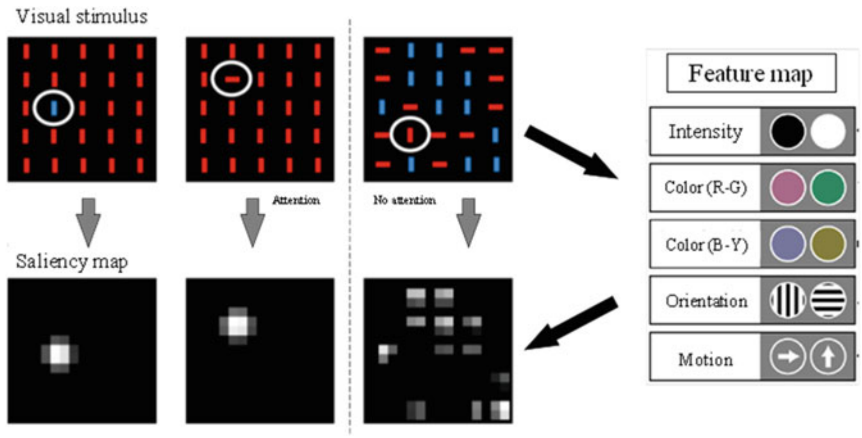


Fig. 8 Explanation of the saliency map [4–6]. Examples of 3 visual stimuli conditions of search tasks for which the saliency map was developed, and the corresponding saliency map output for each

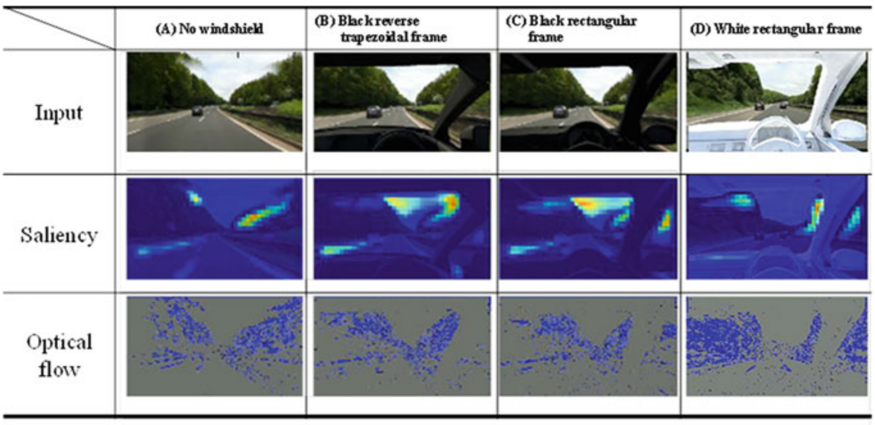


Fig. 9 Visual saliency and optic flow computed for sample frames

3.2 *Calculation Results of Visual Saliency*

Figure 9 shows the saliency map output for representative frames of experimental conditions (A)–(D).

If one compares condition (A) with e.g. (B) and (C), then we can see that the existence of the windshield frame causes the upper part of the road to become more salient. Furthermore, in cases without the frame, it becomes more salient due to the addition of the optic flow. If we compare (B) and (C), then we can see that in the reverse trapezoid condition, the salience above the pillar is strong, whereas in the more rectangular frame, the salience is pulled to the side window above the side mirror. This may be related to the increased density of optic flow, as shown in the third row of Fig. 9.

Finally, if we compare (C) and (D), we can see that the luminance of the frame material is also important. Specifically, lower luminance of the material makes higher salience in the road and external world, whereas higher luminance increases the relative salience of the instrument panel and interior. Low luminance frame material seem to also block some of the optic flow due to the A pillar.

3.3 Discussion

Figure 10 shows the result of different visual feature channels (color, direction, orientation, and motion) computed for the same visual stimulus frames shown in Fig. 9. In condition (A) the saliency is related to all channels, including luminance, color, motion, and orientation, However, in conditions (B) and (C), the intensity channel is stronger in the upper portions of the frame, and thus the effect of intensity is strong. Furthermore, the contribution of the orientation channel is strong along the parts of the windshield frame that intersect the outside image.

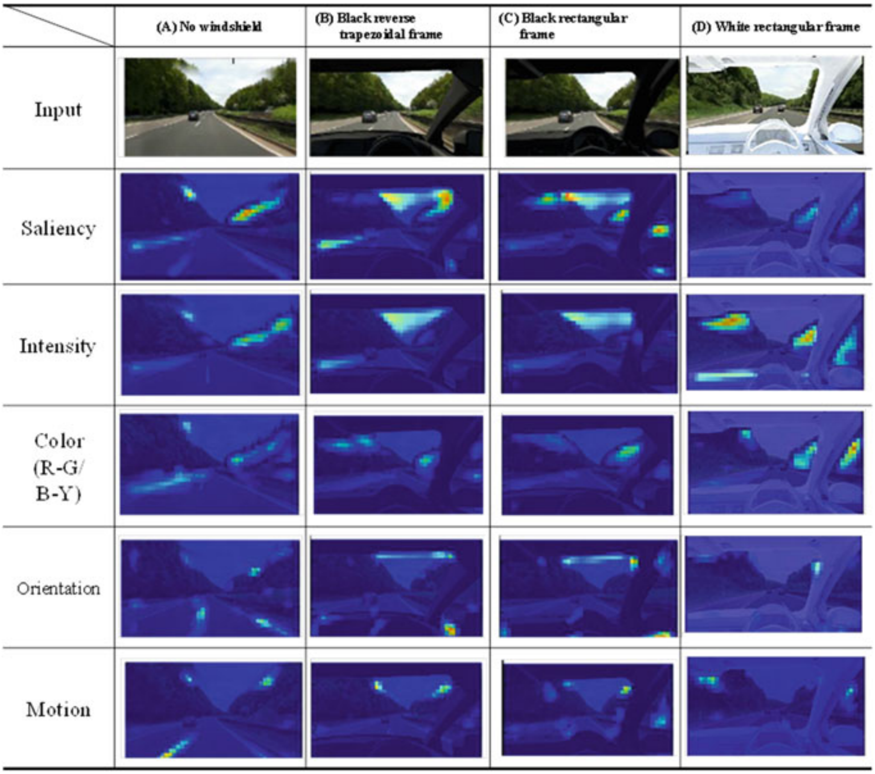


Fig. 10 Comparison of independent feature channels

Furthermore, it is dependent on the color of the trees to the left and right. In this way, the shape and distribution of visual saliency in the image changes depending on the windshield frame. In condition (D), due to the high luminance of the interior material and windshield frame, saliency is focused around the trees outside. This contrasts with the conditions in which the windshield frame/interior is of a darker material. Furthermore, it is separated on both color and luminance feature channels.

For example, if one compares between condition (A) with e.g. (B) and (C) in Figs. 3, 4, 5, 6 and 7, then we can see that the existence of the windshield frame causes to gaze the upper part of the outside. We can quantitatively estimate that the tendency, which the intensity channel is stronger in the upper portions of the frame, influences the gaze movements. Furthermore, if one compares between condition (B) and (C) in Figs. 3, 4, 5, 6 and 7, we can see the tendency of eye position to be approached the pillar side on the right more. In addition, we can quantitatively estimate that the tendency, which the salience is pulled to the side window above the side mirror, influences these gaze movements.

Finally, Fig. 11 shows that the example of analyzing specifically to extract scene. Figure 11a directly compares conditions with or without a windshield frame. Depending on the frame shape, motion and color salience around the A-pillar are dispersed, and the gaze of the driver can be forced to be fixed nearer to the vertical meridian. Figure 11b shows the effect of windshield frame material properties (such as luminance and color) on the driver's attention. Changing the luminance causes the salience to be transferred to other channels, spatially de-correlating the saliency and reducing undesirable "hot-spots". In this way, by having a windshield frame, the luminance of regions unnecessary to driving is suppressed, and the "color", "motion", and "luminance" of the outside world is artificially increased. By manipulating these factors correctly, the goal is to focus driver's attention on visual regions most important to driving and increase the distribution of his gaze in desirable ways.

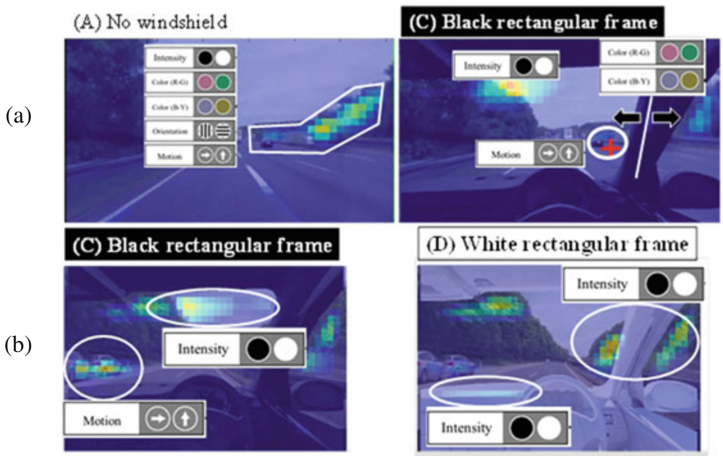


Fig. 11 Distribution of visual saliency channels in response to scene

4 Conclusions

We conclude that drivers can control a vehicle on a highway by gazing at the front center in the distance due to four functions of the windshield frame that affect bottom-up attention.

(1) The front windshield frame stabilizes gaze movements by suppressing unnecessary visual intensity of the outside view.
(2) The front windshield frame restricts the driver's gaze point to the most important region (straight ahead) by artificially de-correlating the spatial distribution of color, motion, and intensity salience of the outside view. This makes it harder for an exciting but irrelevant stimulus to draw the driver's attention.
(3) The visual properties of the windshield frame (such as the color of the material) have an effect on the size and shape of the region to which the driver's attention is drawn. This is because the luminance saliency of the windshield frame affects the saliency of surrounding visual features, and thus contributes to decorrelation of the different saliency channels.
(4) The shape of the windshield frame, especially the inclination of the pillars, make it easier for drivers to set their gaze point along the vertical axis by de-correlating color and motion saliency.

Acknowledgments This research was supported by "Center of KANSEI Innovation Nurturing Mental Wealth." COI STREAM (Center of Innovation Science and Technology based Radical Innovation and Entrepreneurship Program) by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) started in fiscal 2013.

References

1. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cogn. Psychol. **12**(1), 97–136 (1980)
2. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. **4**(4), 219–227 (1985)
3. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision. Res. **40**(10–12), 1489–1506 (2000)
4. Yoshida, M., Itti, L., Berg, D.J., Ikeda, T., Kato, R., Takaura, K., White, B.J., Munoz, D.P., Isa, T.: Residual attention guidance in blindsight monkeys watching complex natural scenes. Curr. Biol. **22**(15), 1429–1434 (2012)
5. Yoshida, M.: Application of saliency map to the analysis of visual search. Jpn. Neural Netw. Soc. **21**(1), 3–12 (2014). (in Japanese)
6. Yoshida, M., Veale, R.: Saliency-guided neural prosthesis for visual attention: design and simulation. Neurosci. Res. **78**, 90–94 (2014)
7. Tanaka, Y., Rakumatsu, T., Horiue, M., Miyazaki, T., Nishikawa, K., Nouzawa, T., Tsuji, T.: An evaluation of automotive interior packages based on human ocular and joint motor properties. Trans. Soc. Instr. Control Eng. **46**(2), 123–129 (2010). (in Japanese)

8. Takeda, Y., Daikokuya, Y., Murase, K., Ohtsubo, T., Hashimoto, S., Fushimi, A., Tamitani, K.: Analysis of automotive interior structure considering relationships human visual direction perception and visual space KANSEI. Trans. Jpn. Soc. Mech. Eng. Ser. C **79**(807), 4039–4046 (2013). (in Japanese)
9. Takeda, Y., Murase, K., Kajikawa, H., Ohtsubo, T., Hashimoto, S., Ioka, T., Tamitani, K.: An evaluation of automotive interior shapes based on human egocentric direction. In: Proceedings of the 6th International Workshop on Image Media Quality and Its Applications (IMQA2013), pp. 86–91 (2013)
10. Yarbus, A.L.: Eye Movements during Perception of Complex Objects, Eye Movements and Vision. Plenum Press, New York (1967)
11. Laretzaki, G., Plainis, S., Vrettos, I., Chrisoulakis, A., Pallikaris, I., Bitsios, P.: Threat and trait anxiety affect stability of gaze fixation. J. Psychopharmacol. **24**(5), 667–675 (2011)
12. A2Mac1 Automotive Bench Marking: http://www.a2mac1.net, Website version: 4.1.1895, Accessed 2013.1 (2013)

Managing Mediated Interruptions in Manufacturing: Selected Strategies Used for Coping with Cognitive Load

Jessica Lindblom and Jonna Gündert

Abstract Interruption research can provide human factors and applied ergonomics with an enhanced understanding of how to notify assembly workers in manufacturing. The paper investigates and analyzes what happens in the transition phase when resuming to the primary task; to understand what kind of support assembly workers may need during this critical and cognitively demanding phase—so that the interval between the interrupted and the primary tasks can be shortened to increase efficiency, during mediated interruptions. Subjects were interrupted during primary assembly tasks via a mobile device which delivered various notifications. We focused on the selected cognitive strategies applied when decreasing the subjects' experienced cognitive load as they resumed to their primary task. Based on the obtained results, some recommendations from a distributed cognition perspective are provided when analyzing "cognitive workscapes."

Keywords Manufacturing · Mediated interruptions · Manual assembly · Cognitive load · Distributed cognition · DiCoT · Cognitive strategies

1 Introduction

It is widely acknowledged that various kinds of interruptions affect work performance, error handling and cognitive load. Interruption research has generally increased in recent years, resulting in a large and growing body of literature in the area. Consequently, the definitions of relevant concepts, theoretical and methodological approaches, and explanations of the obtained results show great variety in both detail and scope. Briefly stated, an interruption is characterized as any event, interruption task, that either breaks a person's attention from their current activity,

J. Lindblom (✉) · J. Gündert
University of Skövde, Box 408, 541 28 Skövde, Sweden
e-mail: jessica.lindblom@his.se

J. Gündert
e-mail: jessica.lindblom@his.se

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_33

primary task, and requests or forces the user to focus on a new, interrupted task [1]. Switching between tasks, even when done in a controlled manner, incurs a cost in terms of time and performance on both tasks. The study of interruptions can easily be applied to the manufacturing domain, which quite surprisingly has not been studied extensively from an interruption perspective (but see work conducted by [2–4]). In order to optimize work performance in manufacturing, we believe it is of major importance to take into consideration the ways information interrupts and notifies assembly workers.

The aim of this paper is to investigate and analyze what happens in the transition phase when assembly workers resume to their primary tasks, and what kind of support the workers may need during this critical and cognitively demanding phase—so that the interval between the interrupted and the primary tasks can be shortened, thereby increasing efficiency. We focused on examining so-called *mediated interruptions*, as found in McFarlane's and Latorella's [5] seminal work of interruption coordination methods. The study was carried out in a in a simulated manufacturing environment. In order to offer a holistic perspective, the theoretical framework of Distributed Cognition (DCog) [6] was used as the foundation. In order to analyze how assembly workers handle interruptions and to determine what can be done to facilitate the resumption of the primary task after being interrupted with the secondary task; we applied the Distributed Cognition for Teamwork (DiCoT) methodology [7]. The purpose of this work is to identify and provide insights to the strategies employed when reducing the levels of cognitive load as the assembly workers return to the primary tasks. Based on our results, we have drawn up recommendations for reducing the cognitive load as the assembly workers resume their primary tasks.

2 Background

2.1 Interruption Research

An interruption is characterized by Corraggio as any event that either breaks a human's attention from their current activity (primary task) and requests or forces the human to focus on a new task (interruption task) [1]. It should be noted that the primary and the interruption task may vary in complexity, level of severity, and urgency. Some interruption tasks may be critical; requiring an immediate response, whereas others can be withheld before being attended. Figure 1 illustrates the interruption process including the task that is interrupted (primary task) and the task that is interrupting (interruption task), as well as the interruption and resumption lags that inherently follow in the transitions between the two types of tasks.

In many cases interruptions can incur a high cost for the human being in question, because switching between tasks, even when controlled, demands a price of time and performance from both tasks [9]. The human performance loss can be

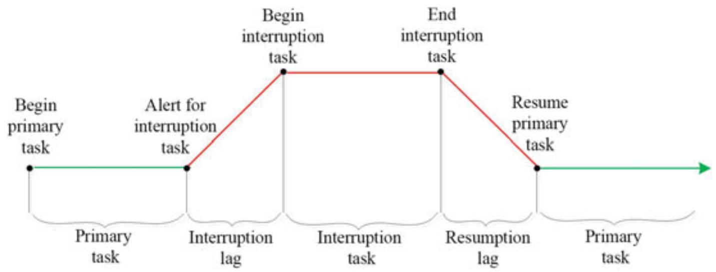


Fig. 1 The process of interruption and resumption involving a primary and an interruption task (modified from Trafton et al. [8], p. 585. © Rebecca Andreasson)

decreased through proper preparation before the task switching, as well as through offering easy ways of "entering" the new task. Similarly, when switching back to the primary task after finishing the interruption task, it is optimal if the primary task can be left in such a state that it can be easily resumed, since then the performance cost of task switching is greatly reduced [10]. Accordingly, McFarlane and Latorella [5] argue that the negative effect of interruptions can be minimized by taking into consideration the demands of the primary task and the interruption task, and the requirements of the person.

There are several proposed ways for managing interruptions. McFarlane [10], for example, suggests a taxonomy of interruption coordination methods that focuses on two main factors: the type/priority of the task being interrupted, and the type and/or priority of interruptions. McFarlane [10] presents four forms of interruption coordination methods, and these are referred to as: (1) *negotiated*, (2) *mediated*, (3) *scheduled* and (4) *immediate* interruption coordination methods. In this paper, we focus mainly on mediated interruptions, which is briefly characterized as having a notification system behaving as a human secretary that judges whether and when interruptions are appropriate, but is also the most complicated to implement in a computer based notification system. Some research exists on how notification systems may respond to the context and user activity in order to display notifications at natural breakpoints in a human's activity [11], instead of sending notifications at any random time, which incurs a lower cost in both human errors and switching times. Bailey's and Konstan's [12] results reveal that interruptions presented during more natural breakpoints has a significantly a lesser impact on the user's anxiety and annoyance, as opposed to being interrupted in the middle of a task. In summary, the best possible option is to present interruptions between well-defined and clear tasks; and the second best option is to direct interruptions at the end of a sub-task [13].

Trafton et al. [8] examined how subjects behaved when resuming a computer task after being interrupted, whether they had or did not have access to various kinds of environmental cues to aid them. Their results show that salient environmental cues can assist the subjects in their reorientation effort when resuming the primary task after an interruption. Briefly stated, environmental cues have a positive impact on the resumption of the primary task.

It is widely acknowledged that interruptions are common in human contexts, and that they either cannot or should not be fully reduced. It is therefore of major importance to consider the nature of human cognitive abilities that have not altered much during evolution, e.g. humans still have limited attention span and short-term memory, although the cognitive demands on human performance has increased in modern working life. McFarlane [14] points out that "*interruption of people is problematic because people have cognitive limitations that restrict their ability to work during interruptions*" (s. 295), and interruptions should not be considered as an isolated phenomena, given they occur in a context, which make it important to decide proper selection of breakpoints, i.e. when and how to interrupt.

2.2 The Theoretical Framework of DCog and the DiCoT Methodology

DCog was introduced by Hutchins [6] in response to more individual models and theories of human cognition. From a DCog perspective, human cognition is fundamentally distributed in the socio-technical environment that we inhabit, and it takes a systems perspective. It discards the idea that the human mind and the social and material environment should be separated, and cognition should instead be considered as a process, rather than as something that is contained inside the mind of the individual. Hence, DCog views cognition as distributed in a complex socio-technical environment and cognition is considered as *creation*, *transformation*, and *propagation* of *representational states* within a socio-technical system [6].

The systems level view makes DCog a fruitful approach for studies of complex socio-technical domains. The framework differs from other cognitive approaches, by its commitment to two theoretical principles [15]. The first principle concerns the boundaries of the unit of analysis for cognition, which is defined by the functional relationship between the different entities of the cognitive system. The second principle concerns the range of processes that is considered to be cognitive in nature. In the DCog view, cognitive processes are seen as coordination and interaction between internal processes, as well as manipulation of external objects and the propagation of representations across the system's entities. When these principles are applied to the observation of human activity in situ, three kinds of distributed cognitive processes becomes observable [15]: (1) across the members of a group, (2) between human internal mechanism (e.g., memory, attention) and external structures (e.g., tools, artifacts, computer systems), and (3) distributed over time.

Different kinds of representations are central to the unit of analysis in DCog, and Hollan et al. [15] argue that representations should not only be seen as tokens that refer to something other than themselves but that they also are manipulated by humans as being physical properties. Humans shift from attending to the representation to attending to the thing represented, which produce cognitive outcomes that could not have been achieved if representations were always seen as representing something else. That is, to study external, material and social structures reveals properties about the internal, mental structures that become observable. Thus, the underlying principle in DCog is that cognition is an *emergent* phenomenon resulting from the *interactions* between the different entities in the functional system, implying that the whole is more than the sum of the individual parts.

However, DCog has been criticized; and two posed forms of criticism regard the DCog view of the very nature of cognitive phenomena, and its utility as an analytic tool since it lacks interlinked concepts that can be easily used to identify specific aspects out of the collected data [16]. Consequently, the DCog approach has been used as a base for the construction of methods, such as the DiCoT methodology [7], which has been proven to facilitate the learning of applying the DCog framework [17].

Briefly stated, DiCoT draws on ideas from Contextual Design [18], but re-orients them towards the governing principles that are central to DCog, resulting in several themes, models and principles that are at the focus of DiCoT. The models include the *information flow model*, *physical model*, *artefact model*, *social model*, and *evolutionary model* that can be used, together with the accompanying principles, to investigate and analyze how and why the work environment have developed over time, and to reason about re-design [7].

3 Method and Procedure

3.1 Setting and Research Design

We used a simulated manufacturing facility in our investigation and analysis of what happens when resuming to the primary task in mediated interruptions; to understand what kind of support assembly workers may need during this critical and cognitively demanding phase. The setting and the scenario resembled an earlier study performed by Kolbeinsson and Lindblom [4], which consisted of a lab-based experiment that simulated tasks and environments typically found in manual assembly and in manufacturing (see Fig. 2a). Kolbeinsson's and Lindblom's study was based on the works of Iqbal and Bailey [11], and Adamczyk and Bailey [13], on selecting breakpoints. The overall research design was an adaptation of McFarlane's [10, 14], prior to contrived studies of interruption coordination methods, but here modified to resemble an authentic manual assembly tasks (see [19] for some motivations for using a simulation).

The simulation consisted of an assembly (primary) task and an interruption task. The *primary task* involved mounting two different kinds of front wheels onto a pedal car, with the extra difficulty that each wheel required a specific setup of tire

Fig. 2 The simulated assembly line to the *left* (**a**), and a closer look of a delivery note to the *right* (**b**)

hardness (soft, medium, or hard) and a specific colour coded bolt (green, blue, yellow, or red), and the tightening of the bolts included using a cordless drill. The delivery note (see Fig. 2b) was placed at the back of each pedal car displaying abbreviations of which kind of tire hardness and colored bolt that should be used on each side, e.g. "LF S" represented "Left side, front wheel, soft tire" and "RF Yellow" represented "Right side, yellow bolt". The mounting of each wheel was categorized as being a distinct sub-task. The *mediated interruption*, which was sent from a computer-based notification system, occurred either between the mounting of the two front wheels (sub-task) or between two pedal cars (task). The *interruption task* consisted of receiving a notification as a text message on a mobile device, then walking to the stationary computer, looking for the right kind of information in form of letters, "decode" them into colored symbols, and finally sending the correct information to the person who was named in the text message. The *breakpoints* were selected from Iqbal's and Bailey's [11] prior findings, which revealed that breakpoints at the end of a task or sub-task cost less in terms of time and error, than the cost from random interruption points [4].

In order to resemble an authentic assembly task, the subjects were forced to manage a standard assembly time, called takt time, which was set to 40 s, with subjects getting a visual indicator when the time was up. This created a sense of time pressure and illustrated the time limits the subjects were expected to meet when no interruptions were sent. Quantitative and qualitative data were collected via observations, video recordings, and semi-structured interviews. The subjects received auditory and tactile notifications as vibrations from a mobile device. During the whole study, loud noise recorded from a real manufacturing facility was played in the background.

3.2 Subjects and Procedure

21 subjects were recruited in total, nine were female and twelve were male, and the majority of them (57 %) were between the ages 20–25. Only one subject had prior experience of manual assembly. Each subject received a practice period until the subject performed the primary and interruption tasks without procedural errors at least three times in a row. All three treatments: "assembly only", "between wheels", and "between cars", were tested in a single condition with a random order of breakpoints. "Assembly only" was being used to keep subjects from predicting when they would be interrupted (these obtained results are reported in another publication). A colleague sent notifications according to a pre-defined random list, with each subject receiving six of each: "no interruption", "between cars", and "between wheels", resulting in each subject assembling at least 18 pedal cars. After participating in the experiment, subjects were interviewed regarding their experiences of handling the interruptions and what tentative cognitive strategies used when resuming to the primary task. On average, one session lasted for approximately 60 min for each subject, including the whole procedure from entering the lab to finishing the interview.

4 Analysis and Results

4.1 The Physical/Artefact Model and the Information Flow Model

The combination of the physical layout model and the artefact model displays how the spatial arrangement was organized, and when taken together, depicts the factors that influences the work performance within the unit of analysis, at a physical level (see Fig. 3). It is of importance from a DCog perspective to represent what factors that have an impact on the subject's "cognitive workspace", and how the physical arrangement hinders or supports the performance of the primary and interrupted tasks. That is, in DCog the environment plays a significant role in describing and analyzing cognition, particularly how these tools and artefacts are designed to boost cognitive processes. Figure 3 shows the physical layout, artefacts, and tools available in the simulated assembly setup.

The combined physical/artefact model provides the general constraints that partly determines and affects in what ways the information flow is created, transformed, and propagated within the "cognitive workspace". The information flow model turns the focus of attention to the communication between peers and artefacts/tools, their individual roles, and the sequences of events over time, which defines the coordinating mechanisms of the system. That is, identifying communication pathways among the entities and the key properties of the of information flow [7]. However, we experienced that the prescribed way of doing a flow model

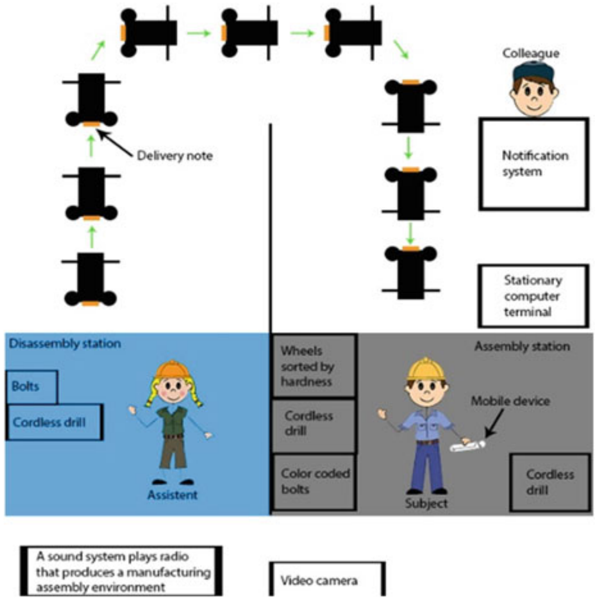


Fig. 3 The combined physical/artefact model of the simulated assembly line, which represent the "cognitive workspace"

in the DiCoT methodology was too limited, since it was on a more general level, explicitly omitting the transformations of the representational states in the different entities of the system during the performance of the primary and interrupted tasks. Consequently, we developed a more detailed way to display the information flow (see Table 1).

As displayed in Table 1, the subjects go through various kinds of representational states, which are transformed and propagated in the information flow within the "cognitive workspace". We have also included physical and manual actions that are represented as *actions-perception couplings*, implying that cognition is embodied (e.g. [5, 20]). Furthermore, the analysis of the above models and the observations done, revealed that there are two separate information flows occurring, which are not connected and thus interfering with each other. On the one hand, there is an information flow for the primary task, and on the other hand, there is another information flow for the interrupted task. Both of these information flows require adequate coordination of both "mental" and "physical" aspects in order to create a functional "cognitive workspace" through the use of available tools and artefacts. In other words, the attention, concentration and performance of the tasks follow two different information flows, which mean that they are strongly attached to respectively task so partly allocating attention to the 'other' task simultaneously is hindered. Summing up, the three modified models provided a solid base to understand and visualize what was happening in the "cognitive workspace" during primary and interruption tasks.

Table 1 The information flow of the primary tasks, the interrupted tasks, and the resumption of the primary task

Activity	Representational formats via the information is propagated	Artefacts and tools	Physical and cognitive actions
Primary task			
1. Read and interpret the instructions on the delivery note	Written abbreviations	Delivery note	The written external representation has to be translated to Swedish and create an internal spatial representation "in the head"
2. Go and grab the right kind of wheels and bolts	Color coding, spatial information	Wheels and bolts	Moving to the place were artefacts and tools are situated while remembering the information on the delivery note "in the head"
3. Place the first wheel, using the correct washer and bolt at the correct side on the pedal car	Embodied action-perception loops of arms, hands, and fingers	"Hands", pedal car, wheel, bolt, and washer	Motoric coordination, concentration, attention
4. Initially fastening the washer and bolt on the axle manually	Embodied action-perception loops of arms, hands, and fingers	"Hands", pedal car, wheel, bolt, and washer	Fine-tuned motoric coordination, concentration
5. Retrieve the cordless drill and properly fastening the bolt	Embodied action-perception loops of arms, hands, and fingers and using the cordless drill	"Hands", pedal car, wheel, bolt, and cordless drill	Fine-tuned motoric coordination, concentration, embodied experience when the bolt is fastened properly
Notification			
6. Notification arrives	Sound and tactile vibration	Mobile device	Alert the shift of attention, from performing the primary task to the interrupted task via hearing and touch
Interruption task			
7. Pick up the mobile device from the pocket	Action-perceptual couplings	Mobile device	Interruption in the current workflow, manually pick up the mobile device
8. Go to the stationary computer and pick off the gloves in order to use the	Action-perceptual couplings	Gloves, mobile device	Forced to leave the workstation, pick off the gloves to tap on the touch function on the

(continued)

Table 1 (continued)

Activity	Representational formats via the information is propagated	Artefacts and tools	Physical and cognitive actions
touch function on the mobile device			mobile device, while remembering the status of the primary task "in the head"
9. Read, interpret and understand the text message sent	Written text	Mobile device	Interpret and understand the information content in the message
10. Find the correct information on the graphical user interface (GUI)	A lot of written text displayed on the screen. Information "stored in the head"	Mobile device, stationary computer terminal	Identify the correct information through mapping of the information content in the artefacts
11. Detect the correct "error code"	A lot of written text displayed on the screen. Information "stored in the head"	Stationary computer terminal	Have to identify the correct "error code", focus of visual attention
12. Fill in the right symbolic sequence (red triangles, yellow circles, etc.), and find the correct receiver of the message	Written text, and graphical symbols of different colors. Information "stored in the head". Action-perception couplings	Mobile device, Stationary computer terminal	Re-represent the written text to graphical symbols of different colors, and manually touch the screen with a finger when filling in the correct symbolic sequence, concentration, short-term memory, attention. Transforming and mapping internal representations to external representations
13. Send the message, pick the mobile device back into pocket, pick up the gloves and return to the workstation	Embodied action-perception loops of arms, hands, and fingers. Information "stored in the head"	Mobile device, stationary computer terminal	Manually send the message via touch screen, put the mobile device back in the pocket, and remember to put on the gloves. Walk back to workstation
Resumption phase			
14. Identify what side of the pedal car that already is mounted	Scaffolds present in the "cognitive workspace" by looking at the pedal car	Pedal car	Trying to remember what have been done and what to do next "in the head", creating an awareness of the current situation
15. Look at the delivery note again	Written abbreviations	Delivery note	Retrieving and remembering the correct

(continued)

Table 1 (continued)

Activity	Representational formats via the information is propagated	Artefacts and tools	Physical and cognitive actions
			information of the process status of the primary task, and what kind of tire and bolt needed
Resuming to primary task			
16. Resume from the beginning (stages 1–5)	See stages 1–5	See stages 1–5	See stages 1–5

4.2 Subjects' Experiences and Strategies Used

The analysis of the collected data from the semi-structured interviews revealed that more than 90 % of the subjects reported that they frequently experienced they had to check the delivery notes although no notifications were sent, and this behavior was also confirmed from the video-recordings and the observations. This implies that the amount of information on the delivery notes was too large to keep "in the head" (limited short-term memory), independently of whether subjects had been or had not been were interrupted. The delivery notes were used as external memory aids, i.e. functioning as *scaffolds*, by the subjects so they could re-orientate themselves in order to resume to the primary task. One explanation for this way of acting was that the information on the delivery note in the primary task was too complex and cognitively demanding since the subjects could not keep every piece of information (tire hardness, colored bolt, and the spatial instructions) "in the head" simultaneously. It was revealed that 81 % of the subjects did not remember how to resume to the primary task when they had finished the interruption task. Furthermore, the design and layout of the interruption task differed significantly from the primary task, since the interruption task used different kinds of representational formats to those used in the primary task. It should be noted, however, that the representational designs were inspired by poorly designed GUIs used in manual assembly of many international manufacturing organisations.

Additionally, the mix of different representation formats between the primary and interrupted tasks resulted in subjects experiencing heavy cognitive load. This means that it is not enough to display the correct information, since the way the information is represented has a severe impact on how the subjects deal with it. That is, as a designer one cannot assume, even when the correct information has been presented—that everyone will have understood it. In addition, 67 % of the subjects reported that the interruption task's cognitive demands negatively affected their general cognitive ability to resume the primary task. The major reason addressed was that the interruption task contained several sequences that differed significantly from the manual assembly task. According to the subjects, the big difference between the sequence of events of the primary task compared to the interrupted tasks led to the use of slightly different cognitive abilities which made remembering the current status of the primary task very difficult.

As mentioned above, the delivery notes were frequently used as external memory aids, and approximately 50 % of the subjects used rehearsal as the strategy to remember significant information displayed on the delivery notes when retrieving the correct wheel (hardness) and the bolt (color). Hence, this strategy aligns with the results obtained by Monk et al. [21]. Another strategy used was trying to follow as certain action pattern, i.e. creating an assembly procedure that was rather "fixed", e.g., always start mounting on the left side, or strictly following the order sequence on the delivery notes. Additionally, some subjects rearranged the physical space, e.g., by moving the pedal cars closer to the wheels and bolts to shorten the distance between them.

Each subject, except for one, mentioned that they found it much harder to resume mounting the second front wheel (sub-task), than to start with another pedal car (task). This implies that it is better to be notified "between cars" than "between wheels" since the breakpoint does not occur during a sequence of several activities being performed on top of each other as in the manual assembly of a wheel. Instead, it occurs at a more "natural" breakpoint of activity (finished assembly of one pedal car). In other words, the more "natural" and "obvious" the breakpoint is from a hierarchical task perspective, the better the timing to be notified is. Several subjects mentioned the feeling of frustration when being interrupted when they had created a good assembly pace, and almost everybody mentioned that they wanted to finish what they were doing for the moment, hence, postponing the shift to the interruption task.

5 Discussion and Conclusions

The aim of this paper was to investigate and analyze what happens in the transition phase when resuming to the primary task, and what kind of support assembly workers may need during the cognitively demanding critical phase so that the interval between the interrupted and the primary tasks becomes shorter and more efficient. This study has revealed that the use of mediated interruptions is a viable interruption coordination method when proper selection of breakpoints is performed in a simulation of manual assembly in manufacturing. However, several issues affect the subjects' cognitive load when they resume to the primary task from the interrupted task. The physical/artifact model and the modified information flow model from the DiCoT [7] methodology, provided visualization that helped identify reasons for the heavy cognitive load. The major reasons for the occurrence of heavy cognitive load when resuming to the primary task are: (1) several different sequences of events, (2) the use of different representations formats, and (3) the combination of different sequences of events with different representational formats that requires several kinds of cognitive and motor abilities. The cognitive strategies identified for reducing cognitive load are the use of environmental cues, memory aids via rehearsal, development of "fixed" action patterns, and reorganization of the physical space.

5.1 Recommendations

The purpose of this work was to provide insights concerning cognitive strategies used to reduce the level of cognitive load when resuming to the primary task. Although our study is limited to a simulation and not conducted in a real assembly context, and rather few subjects participated with limited experience of manual assembly, we still hope that our results provide some insights. Based on the obtained results, the following recommendations have been developed for reducing the assembly workers' cognitive load

- Realizing that manual assembly is a form of cognition that emerges in a "cognitive workspace", allowing assembly workers to use strategies to reorganize the physical space around them
- Providing visual memory aids of relevant information that are designed with clarity, and then placing them at a central and accessible spot so they are available and within reach
- Avoiding a mix of different representational formats as well as too many transformations between representational formats both within and between the primary and interrupted tasks
- Identifying relevant breakpoints for sending notifications, preferably aligned with "natural" interruptions in the assembly sequence of events
- Minimizing the complexity and cognitive demands of the interruption task
- Offering scaffolds that support the resumption of the primary task so that the interval between the tasks becomes shorter and more efficient.

5.2 Concluding Remarks

We emphasize the importance of studying and providing insights of how internal and external processes affect and interact with each other during interruptions. The DiCoT methodology provides a tentative approach, although we argue that the information flow model needs further development to focus also on the transformation of representational states between the primary and interrupted tasks. Our work investigated only mediated interruptions, but the other forms of interruption coordination methods should be included in future work. We argue for the necessity for not only considering the physical aspects of manual assembly work in manufacturing, but also reducing the cognitive load of the workers since there are huge costs associated with neglecting cognitive aspects within manufacturing. To conclude, we address the need for creating mutual relationships between humans and technology, thus designing for the emergence of “cognitive workspaces” in manufacturing.

Acknowledgments The authors wish to thank Ari Kolbeinsson for designing and running the experimental setup, Rebecca Andreasson for the use of Fig. 1, the participants in the study as well as the Sense&React Consortium for their input and support. The Sense&React project is funded by the EU grant FP7-314350.

References

1. Coraggio, L.: Deleterious effects of intermittent interruptions on the task performance of knowledge workers. Unpublished PhD thesis, University of Arizona, USA (1990)
2. Andreasson, R.: Interruptions in manufacturing from a distributed cognition perspective. Master thesis, School of Informatics, University of Skövde, Sweden (2014)
3. Kolbeinsson, A., Thorvald, P., Lindblom, J.: Context aware interruptions: existing research and required research. In: Stanney, K., Hale, K.S. (eds.) Advances in Cognitive Engineering and Neuroergonomics, vol. 11, pp. 260–271 (2014)
4. Kolbeinsson, A., Lindblom, J.: Mind the body: how embodied cognition matters in manufacturing. Procedia Manuf. **3**, 5184–5191 (2015)
5. McFarlane, D.C., Latorella, K.A.: The scope and importance of human interruption in human-computer interaction design. Hum. Comput. Inter. **17**(1), 1–61 (2002)
6. Hutchins, E.: Cognition in the wild. MIT Press, Cambridge (1995)
7. Blandford, A., Furniss, D.: DiCoT: a methodology for applying distributed cognition to the design of team working systems. In: Gilroy, S., Harrison, M. (eds.) IDSV, pp. 26–38. Springer, Berlin (2006)
8. Trafton, J.G., Altmann, E.M., Brock, D.P.: Huh, what was I doing? How people use environmental cues after an interruption. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **49**(3), 468–472 (2005)
9. Spiekermann, S., Romanow, A.: Attention and interruption management for systems design—a research overview. Working paper, Institute of Information Systems at Humboldt University Berlin (2008)
10. McFarlane, D.C.: Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. Human Comput. Inter. **17**(1), 63–139 (2002)
11. Iqbal, S.T., Bailey, B.P.: Effects of intelligent notification management on users and their tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 93–102. ACM (2008)
12. Bailey, B.P., Konstan, J.A.: On the need for attention aware systems: measuring effects of interruption on task performance, error rate, and affective state. Comput. Hum. Behav. **22**(4), 685–708 (2006)
13. Adamczyk, P.D., Bailey, B.P.: If not now when? The effects of interruptions at different moments within task execution. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 271–278 (2004)
14. McFarlane, D.C.: Coordinating the interruption of people in human-computer interaction. In: Sasse, M.A., Johnson, C. (eds.) INTERACT99, pp. 295–303. IOS Press, Edinburgh (1999)
15. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput. Hum. Inter. **7**(2) (2000)
16. Rogers, Y.: HCI Theory: Classical, Modern, and Contemporary. Morgan & Claypool Publishers, San Rafael (2012)

17. Berndt, E., Furniss, D., Blandford, A.: Learning contextual inquiry and distributed cognition. Cogn. Technol. Work, 1–19 (2014)
18. Beyer, H., Holtzblatt, K.: Contextual Design. Morgan Kaufmann, San Francisco (1998)
19. Sanderson, P.M., Grundgeiger, T.: How do interruptions affect clinician performance in healthcare? Int. J. Hum. Comput. Stud. **79**, 85–96 (2015)
20. Lindblom, J.: Embodied social cognition. Springer, Berlin (2015)
21. Monk, C.A., Trafton, J.G., Boehm-Davis, D.A.: The effect of interruption duration and demand on resuming suspended goals. J. Exp. Psychol.: Appl. **14**(4), 299–313 (2008)

The Design Features of Pictorial Face Design for Facilitating Exertion Perception

Ding Hau Huang, Wen Ko Chiou, Bi Hui Chen and Yi-Lang Chen

Abstract The purpose of this study is to determine how many of face features and their combinations could be better comprehensive for different level of physical exertion. Thirty-four healthy volunteers, including 18 males and 16 females were recruited from a university population to participate in the study. There were four types of stimuli outline faces including one holistic condition and three partial face conditions were examined. The sequence of displaying the four types of stimuli faces was from single feature to multi-feature, the order was mouth only condition, eyebrows with eyes condition, eyebrows with mouth condition, and holistic face condition. The main finding of this study is that the different combination of face feature stimuli would influence on perceiving the facial expression of effort. If the key features of facial expression were shown no matter other features were reduction, participants could still correctly understand the information.

Keywords Perceived exertion · Facial perception · Pictorial design

D.H. Huang (✉)
Department of Industrial Design, Ming Chi University of Technology,
New Taipei City, Taiwan
e-mail: hau@mail.mcut.edu.tw

W.K. Chiou
Graduate Institute of Business and Management, Chung Gung University,
Tao Yuan, Taiwan

B.H. Chen
Department of Business Administration, Chihlee Institute of Technology,
New Taipei City, Taiwan

Y.-L. Chen
Department of Industrial Engineering and Management,
Ming Chi University of Technology, New Taipei City, Taiwan

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_34

1 Introduction

Researchers suggested that pictures could facilitate comprehension [1–3]. Comprehension is the process of interpreting the meaning of words or pictures to understand their collective meaning. Pictorial symbol or graphic is type of picture can common be useful in conveying important messages. Well-designed symbols have the ability to communicate large amounts of information at a glance and in facilitating comprehension. Pictorial symbol can also be useful in conveying information to persons who cannot read a printed verbal message, because of vision problems (e.g., older adults), lower-level verbal skills, or inadequate knowledge of the language being used in the descript their own stats (e.g., children) [4, 5]. Many researchers began to modify or develop more appropriate scales using meaningful pictorial symbols, especially for children [6–8]. In order to have a better understanding and assess the ability of children to rate how difficult or easy the effort was perceived to be.

Some scales that specifically consist of face pictorial symbols were developed to help assessment of many kinds of physiological and psychological statuses such as emotion, pain and physical exertion. For instant, pictorial face scales are used in the assessment of emotional states. The scales present different numbers of pictorial faces to measure the varying levels of intensity of emotional responses. One of most used facial emotion scale is the Facial Affective Scale (FAS), FAS comprises nine drawings of children's faces whose expressions vary according to the level of discomfort. The original order of the nine faces ranges from a smiling face to a frowning one with eyes closed, wailing and mouth turned down (which shows the highest level of discomfort). The expressive elements of eyebrows, eyelids, tears and smile respected the degree of emotional intensity. Face scales are also well used in assessment of intensity of pain. Facial expression drawings are a popular method of pain severity assessment in pediatric populations. Faces scales use a series of facial expressions to illustrate a spectrum of pain intensity. Numerous face-based rating scales are available such as the Wong-Baker FACES Pain Rating Scale (WBS) for children [9] and Faces Pain Scale (FPS) for adult [10]. Otherwise, the pictorial rating of perceived exertion (RPE) scale that focuses on the facial expression of effort was developed for evaluating physical tasks [11]. All of these pictorial face scales using pictorial faces to convey different information of intensity to help user understand their internal feeling of their body. However, most of previous studies only evaluated the effectiveness and reliability of these scales, few studies discuss how the features of faces been designed.

Complexity is key element of pictorial comprehension. Complexity is conventionally defined as the level of detail or intricacy contained within a picture [12]. For symbol design in general, pictorial symbols comprising simple forms are preferred [13]. Symbols should be elegant, simple, in their design [14]. The symbol should design as simple as possible to let user understand. It is similar in the case of instructional image, minimizes or eliminates unnecessary information reducing extraneous cognitive load [15, 16]. Readence and Moore [17] suggest that simple drawings are most effective in facilitating comprehension. The advantage of simple drawings over more complex pictures may be due to their minimizing distracting details. Although simplicity is desirable, it may not always be possible to develop an understandable symbol. Sometimes certain cortical details are necessary to convey the concept adequately [18].

The purpose of this study is to determine how many of face features and their combinations could be better comprehensive for different level of physical exertion.

2 Method

2.1 Participants

Thirty-four healthy volunteers, including 18 males and 16 females were recruited from a university population to participate in the study.

2.2 Stimuli

There were four types of stimuli outline faces including one holistic condition and three partial face condition. All the stimuli faces were modified from the Facial pictorial RPE scale [11]. The four types were a mouth only condition (one features), eyebrows with eyes condition (two features), eyebrows with mouth condition (two features), and a holistic face condition (four features). The stimuli faces have six level of exertion expression including extremely easy, easy, somehow easy, somehow hard, hard, and extremely hard. The standard exertion value of the six stimuli faces of each condition was reference by Huang and Chiou [11] define as 0, 2, 4, 6, 8, 10 Facial pictorial RPE scale. Figure 1 shows all the stimuli faces, each rows show the different condition of faces and the columns show the level of exertion intensity from easy to hard, from left side to right side.

2.3 Procedures

The sequence of displaying the four types of stimuli faces was from single feature to multi-feature, the order was mouth only condition, eyebrows with eyes condition, eyebrows with mouth condition, and holistic face condition. Three different 3 min video films unrelated with this study was performed in order to avoid the confounding effect of memory between each type of stimuli condition. Each stimuli face was projected in a randomized sequence on a large white screen using a slide

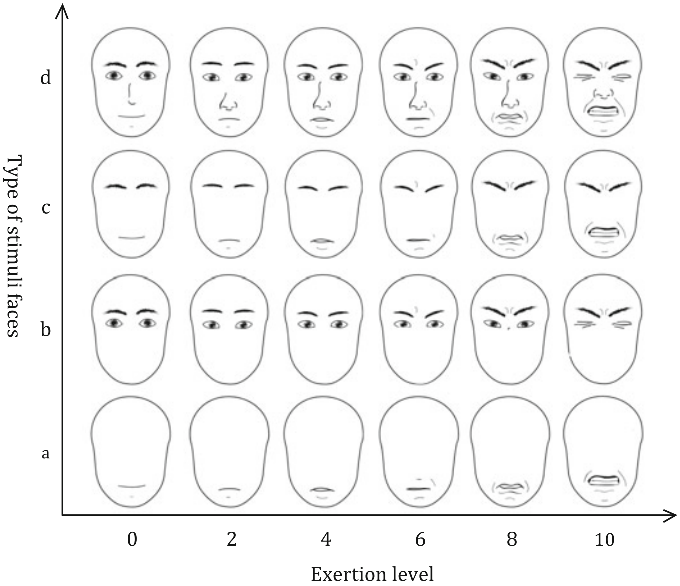


Fig. 1 Six level of exertion expression of four types of stimuli faces

projector. Presentation time was approximately 5 s of each face, but this was modified as needed to give all subjects ample time to complete the questions.

Each participant was asked to rate, map and sort the exertion level when they see the stimuli face. The exertion level was defined as the subjective intensity of effort, strain, discomfort, and/or fatigue experienced during physical activity. The rating is about that when participants see the stimuli faces, if the feelings of exertion lies somewhere between Extremely Easy (0) and Extremely Hard (10), then assign a number between 0 and 10. The mapping is similar as rating question but using RPE scale to map the appropriate number on the scale. The RPE scale was modified facial pictorial RPE scale [11] that removed the pictorial faces and only showed the worlds and umbers. The sorting is that participants were asked to sort the six stimuli faces of each type of condition from easy to hard.

2.4 Data Analysis

Two kinds of errors were measured. One is absolute error from rating and mapping, the measurement error between participants given value and standard value. The other is error rate, which is percentage of number of error, for rating and mapping the error was counted if the difference between standard value and participant's given value was more than or equal to 2. For sorting, the error was number of wrong sequence.

Analysis of variance (ANOVA) and student t test was conducted on these measures. Significance was set at a p value of 0.05 (two-sided) for all analyses. All analyses were conducted using the Statistical Package for Social Sciences Version 19 (SPSS Inc., Chicago, IL, USA).

3 Results

The one-way ANOVA analysis revealed that there was significant difference between these four types of stimuli faces for mean absolute error of rating ($F_{3, 132} = 15.68$, $p < 0.01$) and mean absolute error of mapping ($F_{3, 132} = 7.27$, $p < 0.01$). Figure 2 shows the mean absolute error of rating and mapping for the four types of stimuli faces. The face type of eyebrow with mouth and holistic face has lower rating and mapping error than the type of mouth only and eyebrow with eye. However, post hoc analysis revealed that both errors were no difference between the type of mouth only and eyebrow with eye, and no difference between type of eyebrow with mouth and holistic face. The independent one sample t test revealed that there was significant difference between rating error and mapping error on mouth only and eyebrow with eyes condition however there was no difference between rating error and mapping error on eyebrow with mouth and holistic face condition.

Figure 3 shows the error rate of sorting exertion faces was much lower than rating and mapping. The one-way ANOVA analysis revealed that there was significant difference between these four types of stimuli faces for rating error rate ($F_{3, 132} = 10.24$, $p < 0.01$), mapping error rate ($F_{3, 132} = 5.20$, $p < 0.01$) and

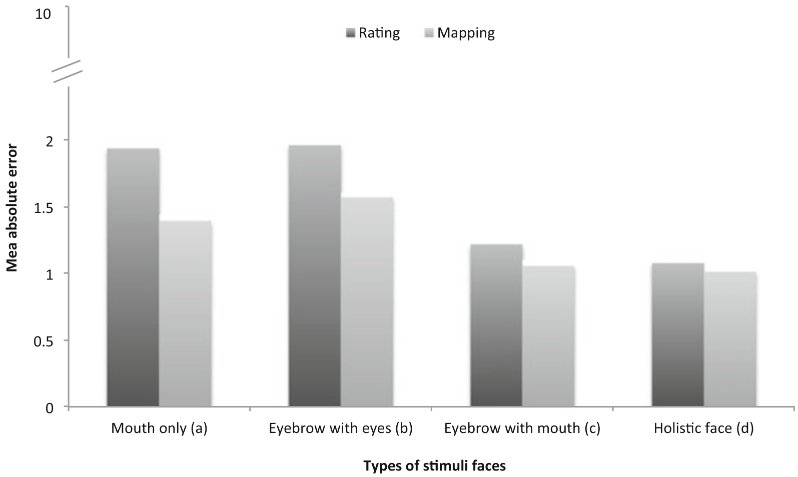


Fig. 2 Mean absolute error of rating and mapping of four types of stimuli faces

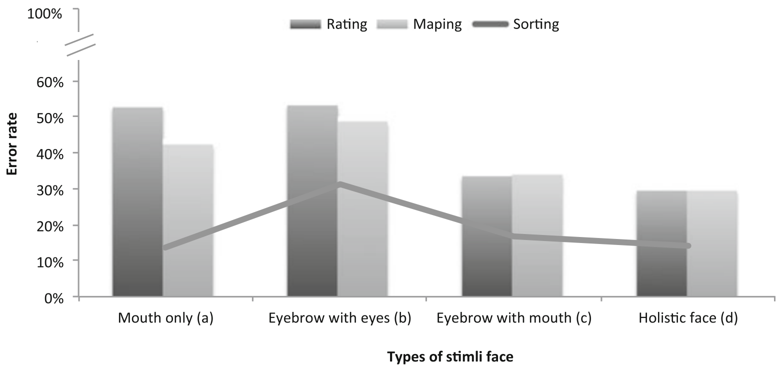


Fig. 3 Error rate of rating, mapping and sorting exertion faces

sorting error rate (F 3, 132 = 5.61, $p < 0.01$). Post hoc analysis revealed that the face type of c and d has significant lower rating error rate than the type a and b. The mapping error rate was significant between face type b and d. The sorting error rate was significant between face type a–c and face type d.

4 Discussion

The main finding of this study is that the different combination of face feature stimuli would influence on perceiving the facial expression of effort. If the key features of facial expression were shown, no matter other features were reduction, participants could still correctly understand the information.

Both results of error value and error rate indicated that eyebrow with mouth condition had the similar effect as holistic faces when rating, mapping and sorting the exertion faces. Although only two features were shown that participants still can perceive the different level of exertion as holistic face. This could be explained that eyes and nose may not the critical region that conveys exertion information. Oppositely the eyebrow and moth may be the key feature for perceiving facial expression of effort. Some studies have showed that facial muscles activity such as frowning and jaw clenching provided a sensitive index of the degree of exerted effort. de Morree and Marcora [19] demonstrated that facial muscle activity associated with frowning was indicative of effort during physical tasks. In their study, the effect of workload on EMG activity of the corrugator muscles increased concomitantly with overall RPE. Huang et al. [20] indicated that the jaw clenching facial expression could be considered an important factor for improving the reporting of perceived effort during high-intensity exercise. Therefore, the combination of eyebrow and mouth was good for the reporting of perceived effort.

However, eyebrows and mouth were important factors for perceived exertion but the combination with other features should be warranted. Mouth only condition has lowest sorting error rates but higher error value. This means that participant can distinguish the different intensity of the six levels by the face with individual mouth feature but it's not sufficient to identify the exactly exertion value. Although eyebrows with eyes condition also reduced to two features, the error value and error rates were much higher than eyebrows with mouth. The eyebrows with eyes have highest sorting error, this mans the change percentage of eyebrows and eyes were too slight to be identified the difference.

Recent finding also indicated that the holistic face with four single features has lower rating, mapping and sorting error, however to reduce different number and different part of face features result in different effect. Eyebrow with mouth condition was more simply design than holistic face but still has sufficient information for perceiving exertion, which fit the pictorial design principle. This maybe explained that human working memory and cognitive load has limited capability, researchers suggest that well-designed materials can effectively minimize extraneous cognitive demands placed on individuals, making working memory resources more available to better process content-related information [21]. Otherwise, simple pictures without distracting, irrelevant details used with easy to read captions would minimize these problems for everyone and especially for people with low reading skills.

References

1. Carney, R.N., Levin, J.R.: Pictorial illustrations still improve students' learning from text. Educ. Psychol. Rev. **14**, 5–26 (2002)
2. Filippatou, D., Pumfrey, P.D.: Pictures, titles, reading accuracy and reading comprehension: a research review (1973-95). Educ. Res. **38**, 259–291 (1996)
3. Houts, P.S., Doak, C.C., Doak, L.G., Loscalzo, M.J.: The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ. Couns. **61**, 173–190 (2006)
4. Kalsher, M.J., Wogalter, M.S., Racicot, B.M.: Pharmaceutical container labels: enhancing preference perceptions with alternative designs and pictorials. Int. J. Ind. Ergon. **18**, 83–90 (1996)
5. Wogalter, M.S., Silver, N.C.: Warning signal words: connoted strength and understandability by children, elders, and non-native English speakers. Ergonomics **38**, 2188–2206 (1995)
6. Parfitt, G., Shepherd, P., Eston, R.: Reliability of effort production using the children's CALER and BABE perceived exertion scales. J. Exerc. Sci. Fit. **5**, 49–55 (2007)
7. Robertson, R.J., Goss, F.L., Andreacci, J.L., Dube, J.J., Rutkowski, J.J., Frazee, K.M., Aaron, D.J., Metz, K.F., Kowallis, R.A., Snee, B.M.: Validation of the children's OMNI-resistance exercise scale of perceived exertion. Med. Sci. Sports Exerc. **37**, 819 (2005)
8. Yelling, M., Lamb, K., Swaine, I.: Validity of a pictorial perceived exertion scale for effort estimation and effort production during stepping exercise in adolescent children. Eur. Phys. Educ. Rev. **8**, 157 (2002)
9. Wong, D.L., Baker, C.: Pain in children: comparison of assessment scales. Pediatr. Nurs. **14**, 9–17 (1988)

10. Stuppy, D.J.: The Faces Pain Scale: reliability and validity with mature adults. Appl. Nurs. Res. **11**, 84–89 (1998)
11. Huang, D.H., Chiou, W.K.: Validation of a facial pictorial rating of perceived exertion scale for evaluating physical tasks. J. Ind. Prod. Eng. **30**, 125–131 (2013)
12. Snodgrass, J.G., Vanderwart, M.: A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J. Exp. Psychol. Hum. Learn. **6**, 174 (1980)
13. Mullet, K., Sano, D.: Designing visual interfaces: communication oriented techniques. SunSoft Press, Englewood Cliffs (1995)
14. Goonetilleke, R.S., Martins Shih, H., Kai On, H., Fritsch, J.: Effects of training and representational characteristics in icon design. Int. J. Hum. Comput. Stud. **55**, 741–760 (2001)
15. Mayer, R.E.: Multimedia learning. Psychology of Learning and Motivation, vol. 41, pp. 85–139. Academic Press (2002)
16. Sweller, J., Van Merrienboer, J.J., Paas, F.G.: Cognitive architecture and instructional design. Educ. Psychol. Rev. **10**, 251–296 (1998)
17. Readence, J.E., Moore, D.W.: A meta-analytic review of the effect of adjunct pictures on reading comprehension. Psychol. Sch. **18**, 218–224 (1981)
18. Woyalter, M.S., Silver, N.C., Leonard, S.D.: Warning symbols. In: Wogalter, M.S. (ed.) Handbook of warnings, pp. 159–176. Lawrence Erlbaum Associates, Mahwah (2006)
19. de Morree, H.M., Marcora, S.M.: The face of effort: frowning muscle activity reflects effort during a physical task. Biol. Psychol. **85**, 377–382 (2010)
20. Huang, D.H., Chou, S.W., Chen, Y.L., Chiou, W.K.: Frowning and jaw clenching muscle activity reflects the perception of effort during incremental workload cycling. J. Sports Sci. Med. **13**, 921 (2014)
21. Wilson, E.A.H., Wolf, M.S.: Working memory and the design of health materials: a cognitive factors perspective. Patient Educ. Couns. **74**, 318–322 (2009)

Design an Interactive Game App of Horticultural Therapy for Older Adults

Pin-Yi Lai and Chien-Hsu Chen

Abstract Horticultural therapy is a non-pharmacologic intervention that has significant benefits of older adults. The aim of this paper is to design an interactive game app of horticultural therapy for aged people. The game is designed by following design methods. First, the interview is performed to realize the procedures and the requirements of horticultural therapy. Then, paper prototypes are made to test the app is appropriate for the elderly. The result is expected to benefit the elderly's physiological conditions and psychological state; therefore, it supports not only the elderly who suffering from the disease such as Dementia but also the healthy aged people who pursuing quality of life (QOL).

Keywords Interactive game · Horticultural therapy · Older adults

1 Introduction

As aging society looms, more and more elderly people facing a high risk of suffering from dementia. Physical aging results in the decline of cognitive abilities and also causes diverse physiological and psychological problems. According to the 2015 Dementia Report of World Health Organization, 47.5 million people worldwide have dementia and there are 7.7 million new cases every year [1]. It is a serious issue in developed countries nowadays. However, medical treatment can only keep the symptoms of Dementia from getting worse [2]. As a result, it is very important to promote the cognitive of the elderly and prevent them from Dementia. In addition, help the older adults with Dementia with another non-pharmacologic treatment.

There are emerging non-pharmacologic interventions (NPIs) currently which are used to address the unmet needs of the agitated individual and to avoid the

P.-Y. Lai (✉) · C.-H. Chen
Department of Industrial Design, National Cheng Kung University,
No. 1, University Road, Tainan City 701, Taiwan
e-mail: pinyi1019@gmail.com

C.-H. Chen
e-mail: chenhsu@mail.ncku.edu.tw

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_35

drawbacks of pharmacological interventions such as medication interactions and side effects. NPIs include music therapy, memories therapy, pet therapy and horticultural therapy [3].

Horticultural therapy is a popular intervention of non-pharmacologic interventions. It is a low-cost therapy that has significant beneficial results [4, 5]. More and more hospitals, nursing homes and community centers begin to import such courses. Through horticultural therapy, participants can obtain social, emotional, physical, cognitive and other kinds of benefits [6, 7]. It is a way to support not only the elderly who suffering from the disease such as Dementia but also the healthy aged people who pursuing quality of life (QOL).

American Horticultural Therapy Association (AHTA) defines horticultural therapy as the engagement of a person in gardening and plant-based activities, facilitated by a trained therapist, to achieve specific therapeutic treatment goals. Hence, horticultural therapy is a goal-oriented treatment just like doctors sign prescription according to the symptom of patients. Horticultural therapist analysis the status of the participants and consider climatic and environmental factors. Then, choose the appropriate course to carry out and evaluate the effects in the end [8]. Horticultural therapy can be divided into several stages that take many weeks. Participants need to be patient and persistent throughout the whole treatment.

The course of horticultural therapy can be divided into two types. One is one time workshop like embossing flowers, leaves rubbing craft, flower arrangement and other creative courses. And the other is long-term cultivation like growing vegetables, fruit and flowers. Horticultural therapists often choose one time workshop for aged people on account of their physical conditions. Many old adults are unable to take care of themselves due to the chronic disease like dementia, depression and diabetes. They need their family or nursing assistant accompany all day long. If therapist arrange long term cultivation courses for the elderly might increase the burden of family member or nursing assistance. Nevertheless, the whole growth process of long-term cultivation is essence of horticultural therapy. The courses not only let older adults express their emotions but also stimulate their cognitive abilities.

Therefore, the aims of this paper is to design an interactive game app of horticultural therapy for aged people in order to shorten the time limit of growing plants. Let the elderly experience the growth processes of plant and know how to raise plant. In addition, turning plants into art crafts let the value of plants last longer. Moreover, more plant species which are difficult to raise or are danger to the elderly can be added into this app making it more diverse, safer and easier to operate.

2 Design and Implementation

This paper is a qualitative research. The design follows the methods below (Fig. 1).

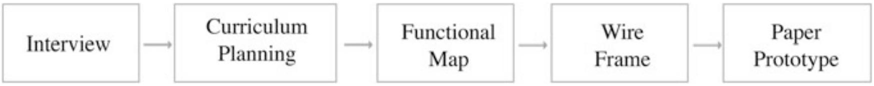


Fig. 1 The design method flow chart

2.1 Interview

First of all, A Semi-structured Interview was performed. We interviewed a horticultural therapist with qualified certification. We prepared several questions that can be divided into three parts—before treatment, during treatment and after treatment. In addition, the difference between the older adults and young people participated in horticultural therapy was compared. Eventually, key elements of horticultural therapy courses was found to design the interactive game app.

After the interviews, we found that therapists usually learn the physical and mental condition of the elderly through the nursing assistance of day care center. In the beginning of the therapy, therapy tells a story about plants and chatting with older adults to understand their conditions. Therapists support them respectively on account of their different problem. After the course, therapists let them share their crafts to gain their self-confidence and interact with each other. The difference between aged people and young people is their declined cognitive abilities and poor physical conditions. Therefore, therapist usually arranges one time workshop for the elderly. This kind of course can be finished within one class. Therapists also make the course simpler to prevent older adults from frustration. In addition therapists add the concept of reminiscence therapy to the course in order to evoke their memories. It also benefits their cognitive abilities too.

2.2 Curriculum Planning

The following steps of horticultural therapy operation process were concluded by the previous literature and the conclusion of the interview (Fig. 2).

During the course of horticultural therapy, horticultural therapist evaluate different condition of the elderly and set the goals which can be divided into four classes—physical function, cognitive function, emotional adjustment and social skills. This paper aims to achieve the fundamental goal of which is to promote cognitive ability. In addition, other additional benefits are also in our concern.

After setting the goal, we started to arrange the horticultural therapy courses. From what has been discussed previously, lots of aged people cannot participate in long term cultivation due to their physical and mental limit. Consequently, we planed to shorten the process of growing plants and present it on mobile devices. In order to choose which parts of growing plants are beneficial to cognitive, we analysis each step of cultivation and figure out what kind of benefits of cognitive

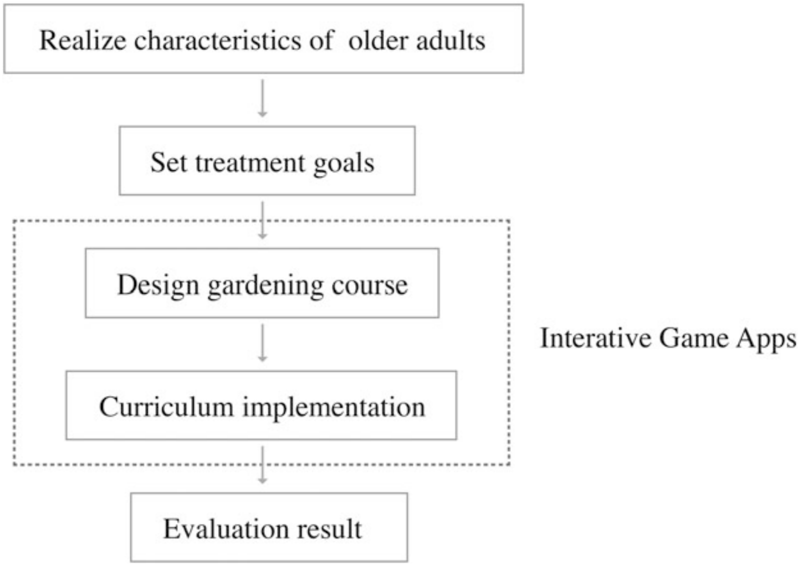


Fig. 2 Horticultural therapy operation process

Table 1 The cognitive benefits from the steps of growing plants

Steps	Visuospatial	Attention	Logical reasoning	Memory
Selecting seeds	●	●	●	
Digging	●			
Sowing	●		●	
Watering	●	●		
Applying fertilizer	●			
Pest control	●	●		
Harvest	●			●

ability is concluded. In the interactive game app, each step can train the visuospatial ability and some of the steps can also train attention, logical reasoning and memory ability (Table 1).

Lastly, we added the art course that can create crafts by plants like embossing flowers, leaves rubbing craft and flower arrangement. This curriculum is beneficial to not only visuospatial ability but also originality. As the elderly finish the craft on mobile devices, therapist can print it on a card that converting virtual achievements to real works (Table 2).

Table 2 The cognitive benefits from the steps of horticultural creative works

Steps	Visuospatial	Attention	Logical reasoning	Originality
Selecting paper	●			●
Placing flowers	●	●	●	●
Entering the words	●		●	
Spraying water	●	●		
Covering paper	●			
Ironing flowers	●	●		

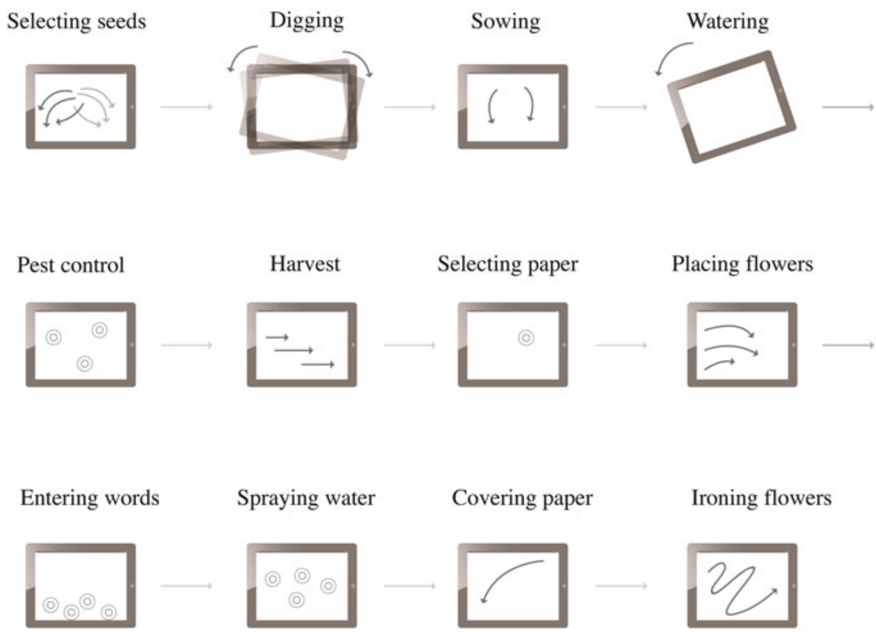


Fig. 3 Gesture on mobile device

2.3 *Prototype*

After planning the course, we selected the steps that is beneficial to cognitive ability and delete the useless one. We arranged the growing plants process into a row and add the mobile device's shape and the gestures to simulate finger motion (Fig. 3).

After drawing the gestures of interface, we started to draw wireframe, using grid and frame to simulate the rough interface and then make paper prototype to test whether the process goes smoothly. Each interface of the task is clear and simple. There are less than 5 objects in one frame to prevent the elderly from confusion (Figs. 4 and 5).

Fig. 4 Paper prototype of growing plants

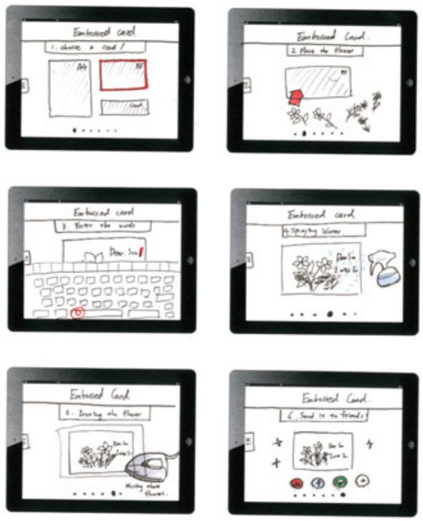


Fig. 5 Paper prototype of making embossed card

3 Discussion

The usability test and mini–mental state examination will be performed after graphic user interface is done.

The result is expected to benefit the elderly's physiological conditions and psychological state; therefore, it supports not only the elderly who suffering from the disease such as Dementia but also the healthy aged people who pursuing quality of life (QOL).

The aged people are not familiar with mobile devices, with lead by therapist and social workers, it will be not only much easier than most of traditional therapy, but also an interesting experience for them. There are many pros we expect in this app: (1) Let the aged people can experience the whole process of growing plants in a short period of time. (2) Enrich the content of horticultural therapy. (3) Safer than other therapy courses.

There are also few cons. Horticultural therapy focus on the experience of five senses including sense of visual, hearing, taste, touch and smell, with this app game, only sense of visual and hearing will be included.

Therefore, we suggest therapist to use some plants to let older adults can touch real leaf texture and smell fragrance of a flower. Also, we can arrange a class to teach them the basic control tutorial and knowledge of how to growing plants before playing this app game.

References

1. Dementia. http://www.who.int/mediacentre/factsheets/fs362/en/
2. Yasukawa, M.: Horticultural therapy horticultural therapy for the cognitive functioning of elderly people elderly people with dementia. In: International Handbook of Occupational Therapy Interventions, pp. 431–444. Springer, New York (2009)
3. Janzen, S., Zecevic, A., Kloseck, M., Orange, J.: Managing agitation using nonpharmacological interventions for seniors with dementia. Am. J. Alzheimer's Dis. other Dementias **28**, 524–532 (2013)
4. Jarrott, S., Gigliotti, C.: Comparing responses to horticultural-based and traditional activities in dementia care programs. Am. J. Alzheimer's Dis. other Dementias **25**, 657–665 (2010)
5. Kortum, P., Peres, S.: The relationship between system effectiveness and subjective usability scores using the system usability scale. Int. J. Hum. Comput. Inter. **30**, 575–584 (2014)
6. Wang, D., MacMillan, T.: The benefits of gardening for older adults: a systematic review of the literature. Activities Adapt. Aging **37**, 153–181 (2013)
7. Van Den Berg, A., Custers, M.: Gardening promotes neuroendocrine and affective restoration from stress. J. Health Psychol. **16**, 3–11 (2010)
8. DiNardo, M.F.: Horticultural therapy: bringing new growth to people with disabilities. J. Extension (2007)

Visual Psychophysical Thresholds for Perceiving Objects and Letters on Monocular Head-Up Displays in Indoor and Outdoor Lighting Conditions

Breanne K. Hawes, Tad T. Brunyé and Brian P. Westgate

Abstract Monocular and binocular head-up displays (HUD) can enhance situational awareness by providing hands-free, real-time information to users on the move. These displays hold the potential for enhancing human experience in many activities, including ambulatory first responders or military personnel. The present study involved a visual psychophysics assessment of three commercially available HUD systems: Vuzix M2000AR, Epson BT-200, and Google Glass. Testing involved 36 participants viewing 112 trials of shape and letter stimuli, presented using the Ascending Methods of Limits psychophysics approach. Half of the trials were completed indoors and half completed outdoors for each HUD. Results demonstrated that participants were able to reliably perceive smaller stimulus sizes with the Epson and Google devices, relative to the Vuzix. This was especially the case in outdoor environments. Results demonstrate the importance of identifying perceptual thresholds for reliably perceiving and interpreting visual stimuli, with large implications for conveying information to the HUD user. Findings of this study demonstrate important practical considerations for selecting commercially-available HUD systems, with particular emphasis on understanding system-specific resolution in tandem with the inherent perceptual capabilities and limitations of human users.

Keywords Head-up display · Cognition · Psychophysics

B.K. Hawes (✉) · T.T. Brunyé · B.P. Westgate
Cognitive Science Team, U.S. Army Natick Soldier Research, Development and Engineering Center,
15 General Greene Ave, Natick, MA 01760, USA
e-mail: Breanne.k.Hawes.civ@mail.mil

T.T. Brunyé
e-mail: Thaddeus.t.Brunye.civ@mail.mil

B.P. Westgate
e-mail: Brian.p.Westgate.civ@mail.mil

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_36

1 Introduction

Tactical monocular and binocular head-up displays hold great potential for enhancing situational awareness by providing hands-free, just-in-time information to users on the move. Helmet- and eyewear-mounted head-up displays are designed to present information in the forward field of view via semi-transparent optical lenses that are positioned in front of one or both eyes. This type of display system affords several potential advantages, including hands-free viewing while maintaining head-up posture and sustained awareness (see McCann et al. [1]).

The present effort involved a visual psychophysics assessment of three head-up display systems:

1. The Vuzix M2000AR system uses a monocular polymer waveguide optical lens and is mounted to the standard issue Army Advanced Combat Helmet, immediately forward of the left or right eye (see Fig. 1a). Resolution is 1280 × 720.
2. The Epson BT-200 system is commercially available and uses a binocular poly-silicon TFT active matrix embedded in a transparent plastic lens. The system is donned like conventional eyewear (see Fig. 1b). Resolution is 960 × 540.
3. The Google Glass system was commercially available (discontinued January 2015) and uses a monocular Himax HX7309 LCoS display embedded in a transparent plastic lens. Like the Epson, the system is donned like conventional eyewear (see Fig. 1c). Resolution is 640 × 360.

2 Approach

Our approach was inspired by visual psychophysics methods developed in the cognitive and perceptual sciences. Visual psychophysics involves evaluating perceptual experiences as a function of systematically varying the properties of a visual stimulus along particular dimensions such as size, brightness, contrast, or motion.

The present work used the *ascending method of limits* which involves successively increasing a visual stimulus property until a person is not only aware of its presence but also its identity [2]. The present evaluation examined identification

Fig. 1 The three head-up displays tested

accuracy as a function of ascending stimulus size (in pixels^2). This method allows us to define detection and identification thresholds for varied types of stimuli, affording an analysis of the realm of possible device applications given its inherent capabilities and limitations across varied lighting conditions (indoor, outdoor).

3 Methodology and Data Collection

3.1 Stimuli

Stimuli were developed using both letters and shapes, allowing us to evaluate the device's ability to present both verbal (e.g., communications) and spatial (e.g., maps/schematics) information to a user. Letter stimuli included all 26 letters of the ISO basic Latin Alphabet, presented in uppercase. Shape stimuli included 5 shapes: circle, ellipse, rectangle, triangle, and square. All stimuli were presented as white against a black background.

3.2 Software

We developed a custom software package to present ascending sequences of letter and shape stimuli using the Microsoft C# programming language. Letter stimuli ascended in size across 29 size increments, from a font size of 1 pt (1 px^2) to 140 pt (5066 px^2) at 100 px/in., in increments of 5 pt, using Microsoft standard Segoe UI font. To increase the number of unique shape items and thus the generalizability of our data, the five shapes were presented in three clock-wise angular rotations, 0°, 45°, and 90°, for a total of 15 shapes. Note that there were only 12 *unique* shape orientations given that rotation doesn't influence the appearance of a circle, and 45° rotation doesn't influence the appearance of a square. As with letters, shape stimuli ascended in size across 29 size increments, closely matched to letter stimuli in pixel area, with average pixel areas (across all 15 shapes and orientations) from 1 px^2 to 5194 px^2.

3.3 Volunteers and Procedure

Thirty six enlisted Soldiers (5 female, 31 male) with normal or corrected to normal visual acuity (confirmed via Snellen eye chart), volunteered for this evaluation (see Table 1 for volunteer demographics). Each volunteer completed a total of 112 trials. Half of the trials (56) were completed in an indoor environment, and half (56) in an outdoor environment. Within each environment, two blocks of trials were

Table 1 Volunteer demographics, including age, visual acuity, and color vision

	Average (SD)	Minimum	Maximum
Age	20.0 (3.5)	19	34
Visual acuity	20/24.9 (6.0)	20/40	20/20

completed, one block with 26 letter stimuli, and the other with 30 shape stimuli (each shape and rotation repeated twice). The order of trials within a block was randomized for each volunteer. Furthermore, whether volunteers performed the task indoors or outdoors first, and whether they viewed the letter or shape block first, were balanced across volunteers.

For the Vuzix HMD, volunteers donned the Advanced Combat Helmet equipped with the monocular head-up display, and the display position (distance from eye, viewing height and angle) was adjusted until each volunteer was able to comfortably and fully view a desktop image projected by a Samsung Series 7 tablet. The tablet operated at 1280 $\times$ 720 resolution, connected to the device via HDMI cable.

For the Epson BT-200 and Google Glass, volunteers donned the system and adjusted vertical and forward/backward positions until comfortable and stimuli were visible. Custom software was adapted for the two devices, using the Android operating system. In both cases, the software was run locally from the device. Like the Vuzix system, the Google Glass was monocular, though the Epson BT-200 used a stereoscopic display.

Ambient lighting conditions were recorded using a calibrated hand-held photometer (EXTECH Instruments, Nashua, NH) placed on a flat horizontal surface immediately in front of each volunteer; focal lighting was recorded by placing the photometer immediately in front of each volunteer's eyes, ahead of the head-up device, oriented away from the participant. The indoor environment was artificially illuminated with fluorescent ceiling lighting, and participants looked straight forward during the task. The outdoor environment was naturally illuminated with indirect (overhead or toward volunteer's back) sunlight that varied from non-occluded to partially occluded by cloud coverage.

Participants were divided into three groups, one for each device, and each containing 12 participants. Within a single trial, volunteers would view an ascending stimulus proceeding to get larger at 1000 ms per increment. They were instructed to press a hand-held response button at the first moment they could accurately identify the letter or shape; upon button press, the incrementing paused and the participant verbally confirmed stimulus identity. Verbal responses were recorded by the experimenter to check for accuracy. For each trial, an output file recorded: participant number, trial number, letter or shape type, and stop frame (increment number).

4 Results

Our analyses are aimed at answering three specific questions. First, for each device we identify psychophysical thresholds for the accurate identification of visual stimuli in indoor and outdoor conditions (*Psychophysical Thresholds*). Second, we ask whether psychophysical thresholds vary for each device (*Device Comparisons*). Finally, we develop quantitative guidelines and limitations for the presentation of verbal and visual information on the display (*Development Recommendations*).

4.1 *Psychophysical Thresholds*

For each trial we coded volunteer responses as a 0 (button press but incorrect response), or a 1 for a correct detect. To define psychophysical thresholds for the identification of our stimuli we calculated cumulative probability distributions for each stimulus size, for both letters and shapes. We then plotted Weibull functions to characterize each participant's detection and predict the requisite stimulus size for 90 % detection rates. The formula for the cumulative distribution function of the Weibull is below, where x is the value at which the function is to be calculated, k is the shape parameter of the distribution, and λ is the scale parameter of the distribution:

$$F(x; k, \lambda) = 1 - \mathrm{e}^{-(x/\lambda)^k} \quad (1)$$

The Weibull functions relating stimulus size and cumulative detection probability for the *letter* detection task, for both indoor and outdoor conditions, are plotted in Figs. 2a–c. Figure 2a depicts the function for the Vuzix device, Fig. 2b for the Epson device, and Fig. 2c for the Google Glass device.

The Weibull functions predict 90 % detection rates at the letter stimulus sizes (in font size and pixels^2) indicated in Table 2. For reference, font sizes are converted into pixels^2 (excluding white space) in Table 3.

The Weibull functions relating stimulus size and cumulative detection probability for the *shape* detection task, for both indoor and outdoor conditions, are plotted in Figs. 3a–c. Figure 3a depicts the function for the Vuzix device, Fig. 3b for the Epson device, and Fig. 3c for the Google Glass device.

The Weibull functions predict 90 % detection rates at the shape stimulus sizes (in font size and pixels^2) indicated in Table 4.

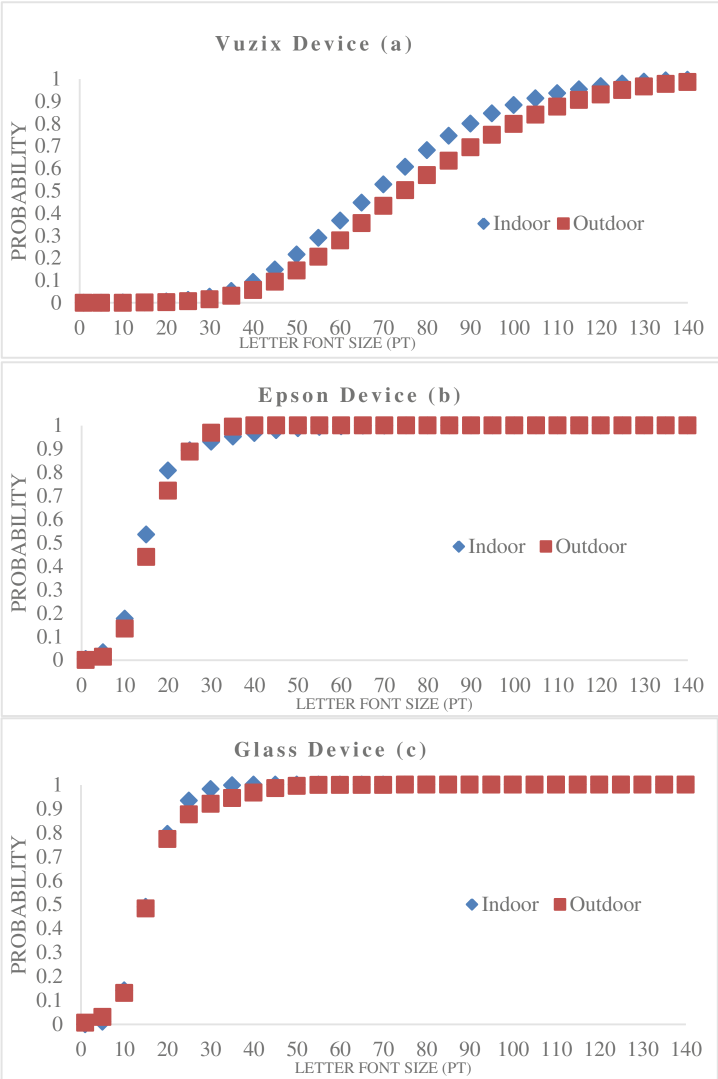


Fig. 2 The Weibull functions relating stimulus size and cumulative detection probability for the *letter* detection task, for both indoor and outdoor conditions

Table 2 Letter stimulus sizes (in font size and pixels^2) for predicted 90 % detection rates

	Vuzix M2000AR	Epson BT-200	Google glass
Indoor	102.91 pt, 3340.98 px^2	25.84 pt, 181.95 px^2	23.80 pt, 153.84 px^2
Outdoor	114.07 pt, 3859.73 px^2	25.73 pt, 180.14 px^2	27.73 pt, 213.36 px^2

Table 3 Reference conversions for font sizes, from pt to pixels^2

Pt	1	10	20	40	60	80	100	110	120	130	140
Pixels^2	1	24	109	418	1347	2276	3206	3670	4135	4600	5066

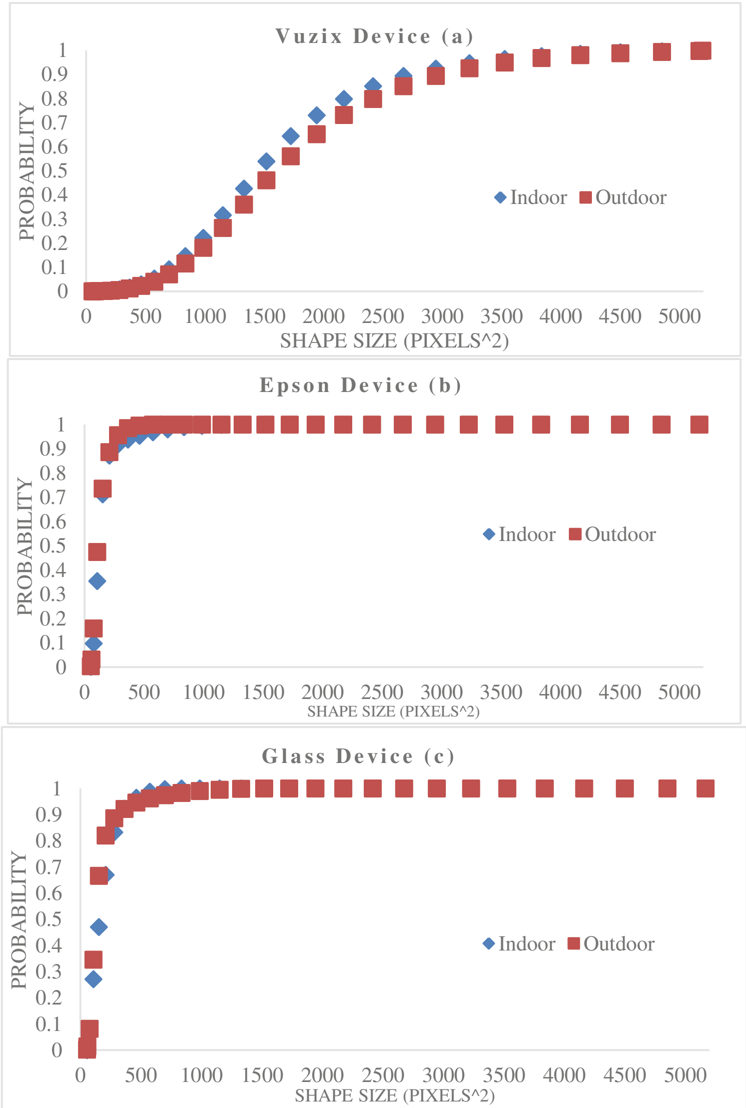


Fig. 3 The Weibull functions relating stimulus size and cumulative detection probability for the *shape* detection task, for both indoor and outdoor conditions

Table 4 Shape stimulus sizes (in pixels^2) for predicted 90 % detection rates

	Vuzix M2000AR	Epson BT-200	Google glass
Indoor	2734.09 px^2	255.17 px^2	345.42 px^2
Outdoor	3006.74 px^2	225.09 px^2	314.49 px^2

4.2 Device Comparisons

The Weibull functions relating stimulus size and cumulative detection probability for the *indoor letter* detection task, for the three devices, is plotted in Fig. 4a. Though the Epson and Glass devices show very similar functions, the Vuzix device requires considerably higher font sizes to achieve the 90 % detection threshold.

The Weibull functions relating stimulus size and cumulative detection probability for the *outdoor letter* detection task, for the three devices, is plotted in Fig. 4b. As with indoor detection rates, though the Epson and Glass devices show very similar functions, the Vuzix device requires considerably higher font sizes to achieve the 90 % detection threshold.

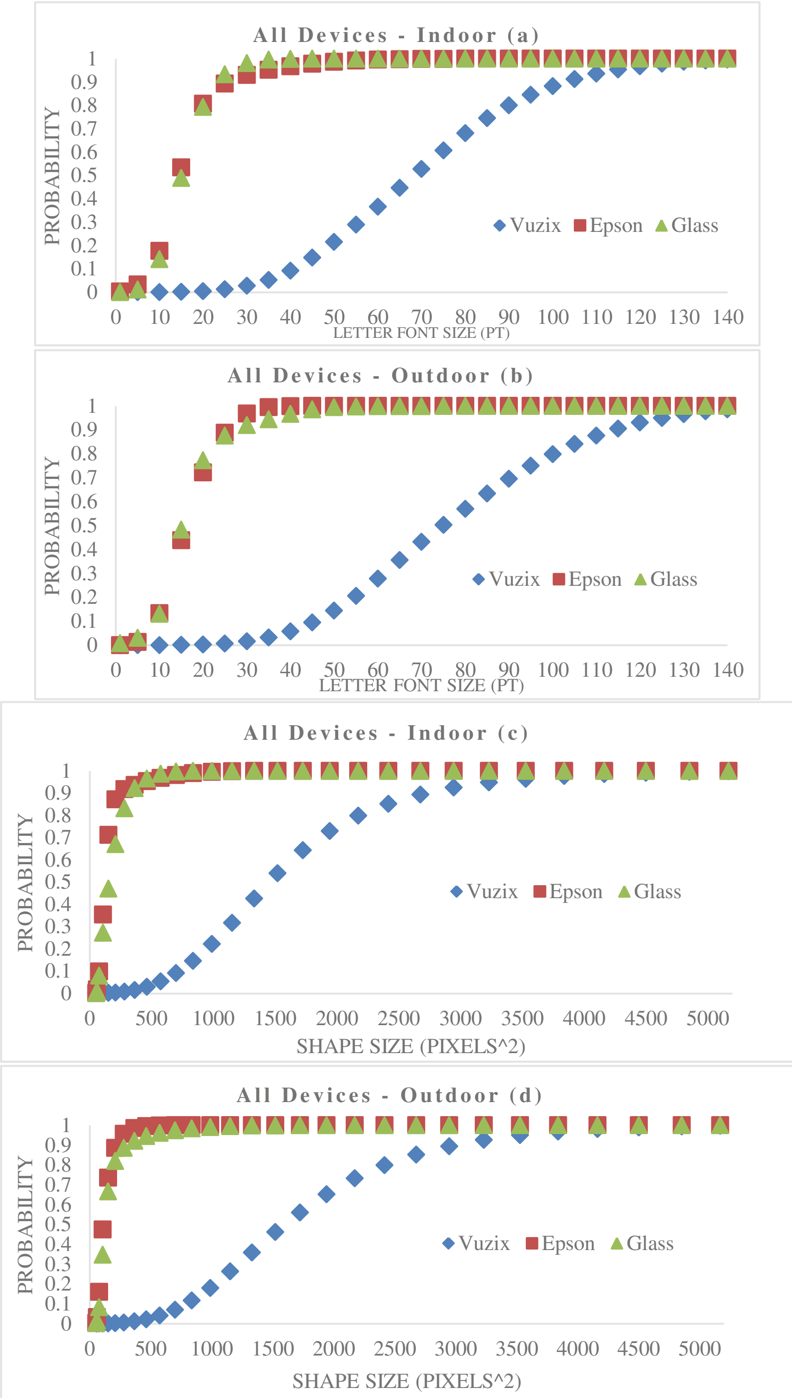
The Weibull functions relating stimulus size and cumulative detection probability for the *indoor shape* detection task, for the three devices, is plotted in Fig. 4c. As was found with the letter detection task, the Epson and Glass devices performed similarly, with relatively poor performance from the Vuzix device.

The Weibull functions relating stimulus size and cumulative detection probability for the *outdoor shape* detection task, for the three devices, is plotted in Fig. 4d. As with the indoor shape functions, the Epson and Glass devices outperformed the Vuzix device; however, the Epson did show a slight advantage to the Glass device in outdoor environments, with lower thresholds for shape detection (225 vs. 315 px^2, respectively).

Though these Wiebull functions suggest that the Vuzix device shows considerably higher psychophysical thresholds relative to the Epson and Glass devices, the Vuzix device also depicts stimuli at a substantially higher resolution. Thus, perceived size may vary due to higher pixel density of the Vuzix device; to test this possibility, the next section relates psychophysical thresholds to device resolution and information density capabilities guided by a notional use case scenario.

4.3 Development Recommendations

To inform continuing development and down-selection of head-up displays for use in tactical operations, the present section is aimed at relating device-specific psychophysical thresholds and display resolutions to real-world use cases. To do so, we relate a range of detection thresholds (i.e., 70–90 %) to the information density afforded at each threshold as a function of display device. For this analysis we collapse across Indoor and Outdoor conditions.


All Devices - Indoor (a)
PROBABILITY
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Vuzix
Epson
Glass
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
LETTER FONT SIZE (PT)
All Devices - Outdoor (b)
PROBABILITY
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Vuzix
Epson
Glass
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
LETTER FONT SIZE (PT)
All Devices - Indoor (c)
PROBABILITY
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Vuzix
Epson
Glass
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000
SHAPE SIZE (PIXELS^2)
All Devices - Outdoor (d)
PROBABILITY
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Vuzix
Epson
Glass
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000
SHAPE SIZE (PIXELS^2)

◀ **Fig. 4** The Weibull functions relating stimulus size and cumulative detection probability for the *indoor letter* detection task, for the three devices

Given stimulus sizes predicted by the Weibull functions, we provide details regarding the proportion of the display size occupied by this stimulus, and thus the maximum number of identifiable objects that could fit on each device display.

Notice that whereas the Weibull functions for the Epson and Glass devices were quite similar (Figs. 3a–d), the lower overall resolution of the Glass puts the device at a practical disadvantage to the Epson. Indeed the Epson can display approximately 1500 more letters or shapes at a 90 % detection threshold.

If we assume a high detection threshold requirement for fielded devices of 90 %, that dictates specific constraints for identifiable verbal and spatial information portrayal. Thus, even with its higher resolution display, the Vuzix device carries stricter limitations in terms of functional information portrayal. The Epson and Glass devices, in contrast, could display the FRAGO in its entirety (Table 5).

For shapes, a 90 % detection threshold imposes constraints on the depiction of spatial information, such as on maps or schematics, and these constraints vary by environment, information density, and zoom levels. For the Vuzix when using satellite imagery, if a Soldier is to visually distinguish landmarks by shape, there is

Table 5 Stimulus sizes at various detection thresholds predicted by Weibull functions

Stimulus type	Device	Detection threshold (%)	Stimulus size (pixels^2)	Proportion of horizontal space	Proportion of vertical space	Maximum number of stimuli
Letters	Vuzix	70	2550.59	0.039	0.070	361
		80	2984.31	0.043	0.076	308
		90	3625.12	0.047	0.084	**254**
		99	5021.27	0.055	0.098	183
	Epson	70	97.40	0.010	0.018	5322
		80	125.60	0.012	0.021	4127
		90	180.70	0.014	0.025	**2868**
		99	680.40	0.027	0.048	761
	Glass	70	94.94	0.015	0.027	2426
		80	116.99	0.017	0.030	1969
		90	165.64	0.020	0.036	**1390**
		99	578.3	0.038	0.067	398
Shapes	Vuzix	70	1971.41	0.035	0.062	467
		80	2317.19	0.038	0.067	397
		90	2882.06	0.042	0.075	**319**
		99	4502.03	0.052	0.093	204
	Epson	70	149.72	0.013	0.023	3462
		80	181.57	0.014	0.025	2855
		90	236.96	0.016	0.029	**2187**

(continued)

Table 5 (continued)

Stimulus type	Device	Detection threshold (%)	Stimulus size (pixels^2)	Proportion of horizontal space	Proportion of vertical space	Maximum number of stimuli
		99	697.52	0.028	0.049	743
	Glass	70	196.93	0.022	0.039	1169
		80	245.14	0.024	0.043	939
		90	336.78	0.029	0.051	**684**
		99	803.14	0.044	0.079	286

Proportion of horizontal and vertical space assumes 1:1 stimulus aspect ratio. Also provided are relations of size thresholds to the maximum number of stimuli at the device resolution. Bold values are 90 % detection thresholds for each device

a minimum required landmark size of approximately 2882 px^2. This is equivalent to, for instance, a square landmark of approximately 53 × 53 px, or rectangle of 85 × 34 px.

In representative grid-based residential areas, this defines a minimum Vuzix zoom level of approximately 1″ = 120 linear feet; at this zoom level, a map can portray approximately 2–3 residential blocks.

For the Epson device, the minimum required landmark size of approximately 237 px^2 defines a minimum zoom level of approximately 1″ = 300 linear feet; at this zoom level, a map can portray approximately 6–7 residential blocks, posing a distinct informational advantage relative to the Vuzix system. The Glass device shows similar advantage, but translates to portrayal of approximately 5–6 residential blocks.

5 Conclusions

The present study was aimed at leveraging experimental visual psychophysics procedures towards evaluating three prototype head-up display systems manufactured by Vuzix, Epson, and Google. The devices were evaluated for their ability to convey visual information (lettersk, shapes) to users embedded in varied ambient lighting conditions (indoor, outdoor).

Data derived from this effort elucidated three primary outcomes. First, the devices crossed the 90 % identification threshold at stimulus sizes that varied widely as a function of the device, with the Vuzix device requiring rather large stimulus sizes relative to the Epson and Google devices. Even after consideration of varied resolution and pixel densities of each device, the requisite stimulus sizes impose constraints on the levels of verbal or spatial detail afforded by each display system. While both the Epson and Google devices outperformed the Vuzix, the higher resolution Epson display is capable of portraying substantially more verbal or spatial information. Second, there was some suggestion that in outdoor environments the Vuzix display required considerably larger stimulus sizes to reach the same identification thresholds found indoors. The Epson and Google devices did not show highly varied detection rates in indoor versus outdoor environments.

Finally, we also demonstrated that the detection thresholds for each device carry large implications for conveying information to individuals wearing mobile head-up displays. Furthermore, each display will vary in its utility based on informational requirements of specific use cases, such as map-based route guidance, or communicating updated tactical information. In the case of maps, granularity requirements can impose constraints on requisite zoom levels for portraying comprehensible spatial information such as the unique sizes and shapes of landmarks. In the case of verbal information, requisite font sizes to increase comprehensibility will impose constraints on the quantity and quality of communicated information.

Overall, the present effort demonstrated a distinct advantage to the Epson and Google devices relative to the Vuzix device, and there was also some evidence that the Epson device is capable of conveying higher density information relative to the Google device.

References

1. Mccann, R.S., Foyle, D.C., Johnston, J.C.: Switching attention between HUDS and worlds: a chronometric analysis. In: Proceedings of the International Conference on Aviation Psychology, Columbus, Ohio (1993)
2. Gescheider, G.A.: Psychophysics: The Fundamentals. Psychology Press (1997)

Effects of Biasing Information on the Conceptual Structure of Team Communications

Michael T. Tolston, Victor Finomore, Gregory J. Funke, Vincent Mancuso, Rebecca Brown, Lauren Menke and Michael A. Riley

Abstract This study evaluated the effect of biasing information on team communication and cognition in a distributed team decision-making task. Teams received misleading or irrelevant information (control) either early or late in their information queue, and Conceptual Recurrence Analysis (CRA) was used to quantify conceptual structure in team communications. Teams in the Late condition produced a significantly greater proportion of conceptually similar utterances than teams in the Early or Control conditions. There was also a trend in the Early condition for utterances to be more conceptually similar than those in the Late condition. Additionally, the persistence of misleading information was affected by condition: teams in the Late condition were still discussing misleading information in the second half of the experiment, but teams in the Early condition were not. We take this as evidence that receiving misleading information later in the queue decreased the focus of team conversation.

Keywords Teams · Team cognition · Team performance · Conceptual recurrence analysis

---

M.T. Tolston (✉)
Oak Ridge Institute for Science and Education, Wright-Patterson AFB, Ohio, USA
e-mail: michale.tolston.ctr@us.af.mil

V. Finomore · G.J. Funke
Air Force Research Laboratory, Wright-Patterson AFB, Ohio, USA

V. Mancuso
MIT Lincoln Laboratory, Lexington, MA, USA

R. Brown · L. Menke
Ball Aerospace & Technologies Corporation, Fairborn, OH, USA

M.A. Riley
University of Cincinnati, Cincinnati, OH, USA

K.S. Hale and K.M. Stanney (eds.), *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing 488, DOI 10.1007/978-3-319-41691-5_37

1 Introduction

Teams play a central role in modern commercial, government, and military organizations [1]. As such, understanding the processes that lead to successful team outcomes is of vital importance. One process thought to shape team outcomes, team cognition, is described as emerging from mutual constraints among cognitions of team members and related team functions [2, 3]. Such essential team-level interactions can have many positive outcomes with respect to team performance, but they can also lead to negative consequences, such as the propagation of unconscious cognitive biases [4, 5]. Furthermore, team interactions may magnify individual biases, or result in new, emergent, group-level biases [4], e.g., groupthink and group polarization. While the effects of cognitive biases have been (and continue to be) extensively studied at the individual level [6], relatively little research has been conducted at the team level [4].

In one of the studies of team level bias, Mancuso et al. [6] evaluated the effects of introducing biasing information on distributed decision making. The authors found that presenting teams with misleading information early in a collaboration resulted in a tendency for those teams to focus their conversations on the misleading information, which resulted in a significant decrease in overall team performance. This behavior is consistent with the well-known phenomenon of confirmation bias. It has been pointed out, however, that team-level biases and facilitative interactions are, in some cases, fragile or difficult to predict [7, 8]. This indicates a need for continued research on the interaction between team processes, cognition, and biases.

The aim of the current research was to extend and generalize Mancuso and colleague's [9] investigation into the effects of confirmation biases in team collaborations. We attempted to do so by using a relatively new technique to evaluate the team-level structure of conceptual content in communications, Conceptual Recurrence Analysis (CRA) [10, 11], in addition to other more frequently employed measures of team performance, such as time to complete a task. CRA is a method of analyzing the semantic structure in communications, and is based on Recurrence Quantification Analysis, a non-linear time series analysis that quantifies structure in potentially complex signals [12, 13]. In essence, CRA quantifies the degree of conceptual alignment between utterances using a set of concepts as a basis. These concepts can be a set of important terms from the communications themselves, or can be extracted empirically from a corpus. This method results in a valuable property of CRA, namely a readily interpretable output in which alignment between utterances is quantified according to their mutual similarity to items in the set of concepts. This means that the output provides insight into not only the structure of team dynamics, such turn taking, leader-follower relationships, introduction of novel topics, etc. [10, 11], but also provides insight into the meaningful topics that may dominate particular subsets of exchanges. This, in turn, allows investigation of the influence of particular concepts in communications that might reveal the effects of experimental manipulations on team biases [14].

In the current study, we have expanded the participant base relative to Mancuso et al. [9] by increasing the sample size and by recruiting teams from two distal locations. We have also conducted a more controlled counterbalance with regard to the distribution of information to team members. We believed that these changes would support and expand on our earlier findings by increasing the generalizability of our results. Additionally, the previous work focused primarily on the outcome of bias, but was not able to provide insight into the underlying mechanisms by which it emerged. Our use of CRA for the evaluation of semantic similarity of communications in this study was intended to provide unique insight into this situation. In line with prior findings, we predicted that teams provided misleading information early in their information queues would exhibit communication structure significantly altered from teams in which the same misleading information was presented at a later point and from teams in a baseline condition in which the extra information was irrelevant to the task. We predicted similar outcomes with respect to the relative frequency of key terms pertinent to the partial solution of the puzzle and time to completion. Additionally, we predicted that there would be no effect of location on team performance or the semantic content of team communications.

2 Method

2.1 Participants

One hundred and thirty-six participants, organized into 34 four-person teams, took part in this study. Of these, sixty-four participants (29 men and 35 women) were recruited from the Dayton, Ohio area, and were compensated $15/h for their time. The ages of those participants ranged from 18 to 33 years (M = 23.05, SD = 3.76). An additional seventy-two participants (49 men and 23 women) from the United States Air Force Academy (USAFA), near Colorado Springs, Colorado, took part in the study for partial course credit. Ages of those participants ranged from 18 to 24 years (M = 20.57, SD = 1.34). Due to technical problems during data collection, one team from USAFA did not complete the study. All participants had normal, or corrected to normal, vision.

2.2 Experimental Design

The experiment was a 3 (bias condition) × 2 (data collection location) between-groups design. The three bias conditions were Early, Late, and Control; differences between conditions were in regard to the type and serial position of potentially biasing information presented to teams (each condition is explained in detail below). As mentioned above, data for this study were collected at two sites:

Dayton and USAFA. This provided an opportunity to evaluate the effects of different sample populations on team performance and communication.

Dependent variables assessed as part of this experiment included: accuracy of team responses, frequency counts of key terms, time to complete the task, and scores associated with CRA.

2.3 ELICIT Task Environment

The Experimental Laboratory for Investigating Collaboration Information-sharing and Trust (ELICT) [15] is a task that requires teams of individuals to work together to solve a logic puzzle. In ELICIT, teams take on the role of information analysts working to identify an imminent (fictional) terrorist attack. Team members are provided with unique information "factoids," i.e., short textual statements of varying utility for solving the logic puzzle, which they must then share with teammates over Internet Relay Chat (IRC) to reach a conclusion regarding the details of the attack. Specifically, they are asked to report *Who* will be conducting the attack, *What* will be attacked, *Where* the attack will happen, and *When* the attack will take place. Any subset of factoids that is sent to a single team member does not contain sufficient information to answer all aspects of the question correctly, though taken in total, the set of distributed factoids is always sufficient.

In the current study, each member of a team received a set of sixteen factoids (64 in total for the team). These factoids can separately be classified as *Expert, Key, Supportive,* or *Noise. Expert* factoids are unambiguous statements regarding the details of the upcoming attack; *Key* factoids help narrow down the range of applicable factoids by providing vital information related to the expert factoids; *Supportive* factoids provide supplementary information that further narrows the range of possibilities; and *Noise* factoids provide information that is irrelevant to the solution of the puzzle. Example factoids can be seen in the team communications in Fig. 1.

2.4 Procedure

Teams were assigned to conditions in a pre-established, counter-balanced order. A prior study [9] and several pilot scenarios have shown that teams in ELICIT that do not receive biasing information reliably answer the four questions accurately. In light of this, and in anticipation of a limited subject pool, conditions featuring biasing factoids were over-represented to increase power in those conditions. However, due to a clerical error, more teams in Dayton were assigned to the Early condition than the Late condition. At USAFA, seven teams were assigned to the Early condition, six to the Late condition, and four to the Control condition. In Dayton, eight teams were assigned to the Early condition, four to the Late

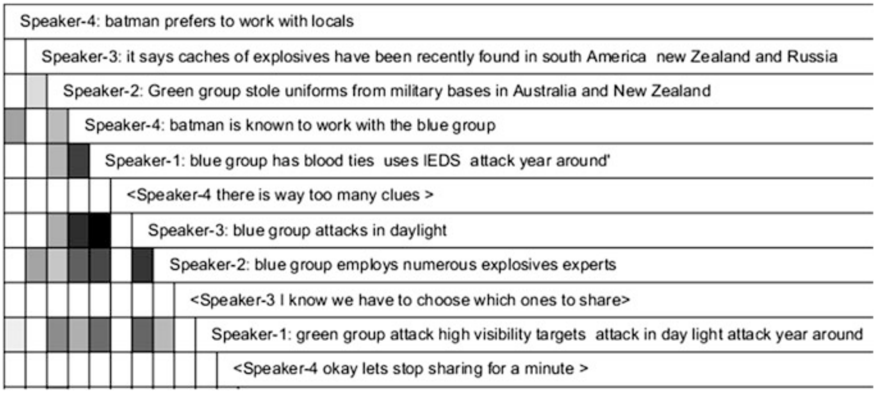


Fig. 1 Conceptual recurrence plot of an excerpt of utterances from a team in the early bias condition. Several of the shared factoids implicate the “Green” team, while others provide information about the “Blue” team. Utterances that are similar to each other have the corresponding row and column in the matrix filled in; the degree of similarity is indicated by the saturation of the shading (*darker shading* indicates greater similarity). Utterances that are enclosed with “< >” were not identified as being similar to any of the key terms extracted from the factoid data set

condition, and four to the Control condition. In all, fifteen teams were assigned to the Early condition, ten to the Late condition, and eight to the Control condition.

Upon arrival, participants completed an informed consent document and were provided instructions regarding the task and the ELICIT environment. During the task, participants sat at individual computer workstations that were visually isolated from each other. Team members received their 16 factoids serially, at a rate of approximately one every eleven seconds. Factoid distribution was initiated as soon as the task began (so all factoids were distributed to the team within three minutes of the start of the task). As a set, the factoids that each team received contained information that unambiguously pointed to the correct answer to the logic puzzle. However, the information any given team member received alone was not sufficient to solve the puzzle, and teams therefore had to collaborate to correctly respond to the four questions posed (i.e., *Who*, *What*, *Where*, *When*).

To introduce ambiguity, and potentially induce team confirmation bias, teams assigned to the Early or Late conditions received three factoids per team member (out of the 16 total per person) that provided misleading information regarding the *Who* component of the terrorist attack; teams in the Control condition received three *Noise* factoids per team member instead. Specifically, the misleading information received by participants in the Early and Late conditions implicated the “Green” group in the pending attack (an incorrect response) rather than the “Blue” group (the correct response). It is important to note that the addition of these extra factoids still resulted in a set that provided sufficient evidence to rule out all incorrect possibilities.

For teams assigned to the Early bias condition, the misleading factoids appeared for each participant as the second, third, and fourth factoids they received (the specific order of the factoids was randomized for each team). For teams in the Late condition, the misleading factoids appeared as the thirteenth, fourteenth, and fifteenth factoids they received. In both conditions, the serial order of all other factoids was randomized for each team. To ensure that there were no interacting effects between the original factoids and the misleading factoids, the factoid pairs were counterbalanced across participants and teams using a Greco Latin square.

During the experiment, participants were asked to communicate exclusively through the provided IRC interface. Participants did not receive any guidance regarding how to coordinate or strategize a solution to the puzzle. After the team decided upon an answer, each team member submitted their responses to the four questions using an electronic form and the session was ended.

2.5 Analysis Approach

CRA was conducted using the software package Discursis [11]. Discursis includes a number of features that are designed to simplify the complicated task of analyzing communications. For instance, a potential difficulty of communication analysis is the number of steps in preparing data for analysis, such as parsing and the removal of common terms, which can make replication difficult [16]. In Discursis, many of these options are streamlined via option dialogues in a user interface that can make standardization of semantic analysis between studies relatively simple.

Mathematically, CRA measures the co-occurrence of terms in a text to estimate their degree of conceptual similarity to a set of key terms, and utilizes these relations to project utterances into a lower-dimensional similarity space. The pairwise ratios of relative importance of key terms between each projected utterance (i.e., the cosine of the angle between utterances in the feature space) are then used to measure the similarity of each pair of utterances. In this way, utterances are reduced to an easily interpretable feature space and summarized in a *similarity matrix* (see Fig. 1), which serves as a basis for a variety of metrics that quantify the semantic structure of communications [10, 17].

In the current experiment, we focused on two recently established metrics to estimate the semantic structure of utterances in teams [14]. The first is the *recurrence of concepts* across utterances, measured by the density of similar utterances, abbreviated elsewhere in this document as *REC*. This measures the number of utterances that are mutually similar to at least one key concept, regardless of the degree of similarity. This is the ratio of shaded (regardless of intensity) to blank (white) squares in Fig. 1. The second is the average similarity of similar utterances, or *mean similarity,* which, given that two utterances are similar to at least one shared key concept, yields a measure of how similar they are. In other words, this is a measure of the average degree of shading of shaded (not blank) cells in the recurrence matrix. To illustrate, in Fig. 1, the utterances "blue team attacks in daylight" and "blue group has blood ties uses IEDS attack year round" have a near perfect similarity, both being projected onto the concepts "attack," "blue," and "group," and are thus darkly shaded. However, the utterances "green group attack high visibility targets attack in day light attack year round" and "batman prefers to work with locals" are only remotely similar, both being indirectly related through terms that happen to be similar to "red," and therefore have a very light degree of shading. This serves to illustrate two properties of CRA. The first is the intuitively illustrative nature of the conceptual recurrence plots, and the second is the generative capacity of the methodology, in that it often finds statistical relations between phrases that are not immediately apparent.

Prior to submission to CRA, transcripts were corrected for misspellings, and variations of key abbreviations were mapped to a common term (e.g., all instances of "NZ" were replaced with "New Zealand"). Commonly occurring words and prepositions that would otherwise saturate the similarity matrix were removed automatically in the Discursis software. We used the default list of such terms included in Discursis to identify these words, with the exception that we removed months of the year from the list, as part of the solution to the *When* question was the month "November." Data were analyzed using the "merge-word variants" option in Discursis, a stemming procedure that removes suffixes from the roots of words.

In some teams, "green" (the object of the misleading factoids) was not identified as a key concept when data were analyzed using intrinsically derived concepts. Therefore, semantic spaces were constructed using the set of all factoids [14], with the default setting of 100 concepts (an upper limit on the number of concepts extracted from the dataset); analyzed team discussions were projected onto this space. Prior to analysis, results were screened for outliers using the Grubb's outlier test. No outliers were detected in the reported measures. Tukey's HSD test was used for post hoc comparisons ($\alpha_{fw} = 0.05$). In addition to the metrics calculated from the similarity matrices output by Discursis, we examined team accuracy in answering the four ELICIT questions, counts of the key terms "blue" and "green" (both normalized by the number of utterances made by each respective team), and the total time teams took to complete the task.

3 Results

3.1 Correct Responses

Table 1 contains the proportion of correct responses for the *Who, What, Where,* and *When* categories. Teams in the Control condition answered all categories correctly, resulting in contingency table cells with entries equal to zero. Thus, two-tailed Fisher's Exact Tests were conducted on the cell contents. All teams responded correctly with respect to the *Where* category, so this category was excluded from subsequent analyses. Responses were first evaluated for an effect of data collection

Table 1 Proportion of correct responses in each category by condition

Bias Cond.	Who	What	Where	When
Early	0.73	0.93	1.00	0.80
Late	0.70	0.80	1.00	0.80
Control	1.00	1.00	1.00	1.00

Note Cond. = condition

location. There was no effect of location for any of the categories, $p = 0.688$, 0.601, 0.656, for *Who, What,* and *When*, respectively. Responses from the two locations were then aggregated, and tested for effects associated with the bias condition factor. There were no effects of bias condition on submitted responses, $p = 0.257$, 0.434, 0.563, for *Who, What,* and *When*, respectively. In sum, the manipulated variables were not found to influence the accuracy of team responses in this experiment.

3.2 Frequency Analysis

To evaluate the time dependency of focus on the concepts "green" and "blue," data from the sessions were divided in half and utterances containing the terms "blue" or "green" were counted for each half of a session. The counts were then divided by the total number of utterances to calculate the proportion of utterances containing each of the terms. These data were submitted to a 2 (location) $\times$ 3 (bias condition) $\times$ 2 (term, green vs. blue) $\times$ 2 (time, first vs. second half) mixed ANOVA, where time and term were repeated-measures factors.

The results of the analysis indicated a statistically significant main effect of term, $F(1, 27) = 60.607$, $p = <0.001$, $\eta_p^2 = 0.692$. "Blue" ($M = 0.159$, $SD = 0.045$) occurred in a higher proportion of utterances than did "green" ($M = 0.081$, $SD = 0.059$). This outcome was expected because "the Blue Group" was the correct answer to the Who question. The main effect of time was also statistically significant, $F(1, 27) = 41.855$, $p < 0.001$, $\eta_p^2 = 0.608$. Both terms were more likely to occur in the first half of the session ($M = 0.170$, $SD = 0.057$) than in the second half ($M = 0.074$, $SD = 0.062$). In addition, the term by bias condition interaction, $F(2, 27) = 5.970$, $p = 0.007$, $\eta_p^2 = 0.307$, the term by time interaction, $F(2, 27) = 10.861$, $p = 0.003$, $\eta_p^2 = 0.287$, and the three-way term by time by bias condition interaction, $F(2, 27) = 4.969$, $p = 0.015$, $\eta_p^2 = 0.269$, were all statistically significant. No other sources of variance in the analysis were statistically significant (all $p > 0.05$).

To explore the three-way interaction between term, time, and bias condition, the term by bias condition interaction was evaluated separately in the first and second halves of the session, averaging over location. For the first half, a statistically significant effect of term, $F(1, 27) = 67.184$, $p < 0.001$, $\eta_p^2 = 0.691$, and a significant term by bias condition interaction, $F(2, 27) = 10.547$, $p < 0.001$, $\eta_p^2 = 0.431$, were found; the main effect of bias condition was not statistically significant, $F(2, 30) = 2.88$, $p = 0.072$, $\eta_p^2 = 0.161$. Further evaluating the term by bias condition interaction in the first half of the session, a follow up simple-simple effects analysis for the term "green" indicated an effect of bias condition, $F(2, 30) = 10.729$, $p < 0.001$, $\eta_p^2 = 0.417$. Post hoc analyses showed that "green" occurred more frequently in Early ($M = 0.134$, $SD = 0.056$) and Late ($M = 0.168$, $SD = 0.0564$) condition communications than in those of the Control condition ($M = 0.056$, $SD = 0.036$), and that the former two conditions did not differ significantly from each other in that regard, $p = 0.005$, $p < 0.001$, and $p = 0.254$, respectively. A similar simple-simple effects analysis for "blue" indicated there were no differences between bias conditions in use of the term, all $p > 0.05$.

With regard to the second half of the session, evaluation of the proportion of communications which included "green" as a function of bias condition again showed a main effect of bias condition, $F(2, 30) = 4.117$, $p = 0.026$, $\eta_p^2 = 0.215$. Post hoc analyses indicated that "green" was more likely to be discussed in the Late condition ($M = 0.087$, $SD = 0.098$) than in the Control condition ($M < 0.001$, $SD < 0.001$), $p = 0.020$, but occurred equally often in the Early condition ($M = 0.046$, $SD = 0.050$), $p = 0.275$; however, frequency of use did not differ between the Control and Early conditions, $p = 0.243$. The simple-simple effects analysis of the term "blue" indicated no differences between bias conditions in its use, all $p > 0.05$.

To summarize, during the first half of the sessions, the biasing factoids increased the proportion of utterances teams spent discussing "green" over the Control condition. However, by the second half of the experiment, teams in the Late condition were still discussing "green" to a higher degree than in the control condition, but teams in the Early condition were not. All teams discussed "blue" to an equal extent, though both "green" and "blue" were discussed less frequently in the second half of the experiment compared to the first.

3.3 Time to Complete the Task

To evaluate the amount of time teams spent completing the task we employed a 2 (location) × 3 (bias condition) between-groups ANOVA. The results of this analysis indicated there was a statistically significant main effect of location on completion time, $F(1, 27) = 9.323$, $p = 0.005$, $\eta_p^2 = 0.257$. Teams at Dayton ($M = 32.830$ min, $SD = 14.562$ min) took more time to complete the task than teams at USAFA ($M = 21.428$ min, $SD = 8.904$ min). There was also a statistically significant effect of bias condition on completion time, $F(2, 27) = 4.218$, $p = 0.025$, $\eta_p^2 = 0.238$. The condition by location interaction was not statistically significant, $F(2, 27) = 1.670$, $p = 0.207$, $\eta_p^2 = 0.110$. In all, teams took over 40 % more time, on average, to complete the task at Dayton, possibly due to the pay incentive.

Post hoc analyses examining differences in completion time for each bias condition indicated that, after the appropriate alpha correction, none of the bias conditions significantly differed from each other. However, there was a trend for teams in the Early condition ($M = 23.018$ min, $SD = 11.367$ min) to finish faster than teams in the Late condition ($M = 33.435$ min, $SD = 17.283$ min), $p = 0.066$, with teams in the Control condition ($M = 26.244$ min, $SD = 7.240$ min) falling in between.

3.4 CRA

With respect to semantic content, bias condition significantly affected the *density of similar utterances (REC)*, $F(2, 27) = 9.840$, $p = 0.001$, $\eta_p^2 = 0.421$. There was no effect of location, $F(1, 27) = 0.319$, $p = 0.577$, $\eta_p^2 = 0.012$, nor was the interaction statistically significant, $F(2, 27) = 0.275, p = 0.762, \eta_p^2 = 0.020$. Post hoc analyses of the bias condition effect showed that the overall proportion of similar utterances was higher in the Late condition ($M = 0.220$, $SD = 0.070$) than in the Early ($M = 0.132$, $SD = 0.044$), $p = 0.001$, and Control conditions ($M = 0.129$, $SD = 0.013$), $p = 0.002$; the latter two conditions did not significantly differ from each other, $p = 0.99$.

Mean similarity was significantly affected by both bias condition, $F(2, 27) = 3.550$, $p = 0.043$, $\eta_p^2 = 0.208$, and location, $F(1, 27) = 6.571$, $p = 0.016$., $\eta_p^2 = 0.196$. The interaction was not statistically significant, $F(2, 27) = 2.102$, $p = 0.142$, $\eta_p^2 = 0.135$. Mean similarity was higher at USAFA ($M = 0.477$, $SD = 0.035$) than it was at Dayton ($M = 0.452$, $SD = 0.038$), indicating that similar utterances were more aligned at USAFA than they were at Dayton. There were no significant differences in mean similarity between bias conditions after post hoc corrections. However, mean similarity showed a trend in the opposite direction of *REC*, as scores for the Early condition were nearly significantly higher in mean similarity ($M = 0.477$, $SD = 0.03$) among similar utterances compared to the Late condition ($M = 0.444$, $SD = 0.043$), $p = 0.055$.

To summarize, teams in the Late condition produced a greater proportion of similar utterances than teams in the Early or Control conditions. It appears that the Late condition affected communication by increasing the overall recurrence of concepts throughout the discussion and by decreasing the amount of focus on a subset of concepts.

4 Discussion

In this study, we evaluated the influence on team communications of the serial position of misleading information supplied to teams working jointly on a language-based logic-puzzle. Teams that received misleading information late in their queues were more likely to discuss the concept targeted by this information for the duration of the task. Similarly, during the first half of the trial, teams in the early bias condition discussed the targeted concept more than teams who received irrelevant information, but this effect dissipated by the second half. We also found that teams differed significantly by location: teams at USAFA took less time to complete the task and focused more tightly on particular concepts than did teams at Dayton. Finally, CRA results indicated that teams receiving misleading information later in their queue tended to repeat concepts more often compared to teams in the Early or Control conditions, and trended toward being less focused in their discussion than teams receiving misleading information early in their queues.

As expected, we found that introducing misleading information altered team communications, and that this effect was evident in the semantic similarity of utterances at the team level, as well as in the time-dependent distribution of key terms. In particular, we found that adding biasing information late in the information queue increased the average density of recurrent concepts, while the trend was in the opposite direction for the average degree of alignment between similar utterances. That teams in the Late condition were still discussing "green" in the second half of the experiment suggests those teams were unable or unwilling to rule out the misleading information. This in turn increased the probability of this subset of utterances being similar over longer-range scales, and likely increased the overall density of the similarity matrix. This could also explain the decrease in the overall degree of similarity between utterances, as the larger number of active concepts could increase the probability of utterances being obliquely related, rather than being narrowly focused on a subset of explicitly shared concepts.

In addition to the main effect of bias condition on communication structure, we found that teams at USAFA exhibited a higher degree of mean similarity than teams at Dayton. We take this as evidence that teams at USAFA were more focused in their discussion of key concepts, a possibility that is corroborated by the fact that those teams took less time on average to conclude, though answer accuracy was the same between the two locations. We believe a likely explanation is that teams in Dayton were incentivized by their monetary award for participation, meaning they might have been less hurried and more open to broad discussions that extended their session. This could simultaneously explain the decrease in the average degree of alignment between all similar utterances and the increased average time to completion in Dayton. We noted that a limitation to the current study is the unbalanced distribution of teams between conditions, particularly in the Dayton sample. This might be an additional explanatory factor in the differences in time to completion between locations. However, the lack of an interaction between location and condition suggests that all conditions took longer in Dayton. Further, the condition that was over-represented was, on average, the condition in which participants finished most quickly. We believe these observations point to the conclusion that differences in location were driven mainly by incentives at the participant level.

With regard to the influence of serial position on answer accuracy, we found results somewhat inconsistent with prior research. Whereas Mancuso et al. [9] reported that introducing misleading information early in the information queue significantly altered communication patterns and caused teams to be more likely to answer incorrectly, we found evidence that introducing misleading information late in the information queue resulted in a greater effect on team communication, but produced no differences in the accuracy of team answers. Taken in sum, these results indicate that team bias may be a relatively sensitive phenomenon, and point to the importance of context as a possible key factor in the differences between the two studies. For instance, it is possible that certain combinations of factoids presented to participants by Mancuso et al. [9] were complementary, resulting in amplification of bias within an individual that might have propagated to the team. Any such interaction at the individual level might have been eliminated or reduced by the balanced distribution of factoids in the present study. Additionally, it has been pointed out that heuristic biases have in many cases been engendered by taking advantage of participants' lack of sophistication [18], and that such heuristics may prove somewhat fragile in less constrained settings that rely on participant knowledge rather than their ignorance [19]. These considerations, along with the current results, indicate that more research is needed.

Several future research directions suggest themselves from the results of the current study. For instance, it is well established that manipulation of the serial presentation of information, under sufficient time constraints, may result in systematic bias in answers [6]. It is possible that providing an additional time constraint under the given framework would reliably produce a biasing effect, though the time provided may have to be so limited as to prevent discussion of all information, thus resulting in an anchoring, rather than a confirmation, bias. Additionally, it might be that allowing individuals a greater amount of time to discuss each factoid separately might influence formation of bias. Finally, as previously mentioned, it is possible that some combinations of factoids may serve to amplify biasing effects at an individual-level, and the bias could then be transmitted to the team. Examination of each of these possibilities would further expand our understanding of team cognitive biases, suggesting that all are worth exploring.

In conclusion, the results of the current experiment suggest that the expressions of team cognitive biases are likely more nuanced than originally anticipated. Factors such as serial position of information in a queue have been demonstrated to influence team processes, such as communication, but they may not reliably influence team decision-making. Additionally, we believe our results show that CRA can provide insight that complements traditional metrics, such as frequency counts and time to task completion, and in doing so provides a deeper context for understanding team communication dynamics than can be obtained from traditional analyses alone. While the results of this study present more questions than answers, our outcomes contribute to the relatively small, but vital, body of research examining the effects of cognitive biases on team decision-making.

Acknowledgments This project was supported by grant no. 15RHCOR233 from the Air Force Office of Scientific Research (Benjamin Knott, Program Officer).

References

1. Kozlowski, S.W., Ilgen, D.R.: Enhancing the effectiveness of work groups and teams. Psychol. Sci. Public Interest **7**, 77–124 (2006)
2. Cooke, N.J., Gorman, J.C., Myers, C.W., Duran, J.L.: Interactive team cognition. Cogn. Sci. **37**, 255–285 (2013)
3. Wildman, J.L., Thayer, A.L., Pavlas, D., Salas, E., Stewart, J.E., Howse, W.R.: Team knowledge research emerging trends and critical needs. Hum. Factors J. Hum. Factors Ergon. Soc. **54**, 84–111 (2012)
4. Jones, P.E., Roelofsma, P.H.: The potential for social contextual and group biases in team decision-making: biases, conditions psychological mechanisms. Ergonomics **43**, 1129–1152 (2000)
5. Schulze, C., Newell, B.R.: More heads choose better than one: group decision making can eliminate probability matching. Psychon. Bull. Revi. 1–8 (2015)
6. Tversky, A., Kahneman, D.: Judgment under uncertainty: heuristics and biases. Science **185**, 1124–1131 (1974)
7. Theiner, G., Sutton, J.: The collaborative emergence of group cognition. Behav. Brain Sci. **37**, 277–278 (2014)
8. Kerr, N.L., MacCoun, R.J., Kramer, G.P.: Bias in judgment: comparing individuals and groups. Psychol. Rev. **103**, 687–719 (1996)
9. Mancuso, V.F., Finomore, V.S., Rahill, K.M., Blair, E.A., Funke, G.J.: Effects of cognitive biases on distributed team decision making. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **58**, 405–409 (2014)
10. Angus, D., Smith, A.E., Wiles, J.: Human communication as coupled time series: quantifying multi-participant recurrence. IEEE Trans. Audio Speech Lang. Process. **20**, 1795–1807 (2012)
11. Angus, D., Smith, A., Wiles, J.: Conceptual recurrence plots: revealing patterns in human discourse. IEEE Trans. Vis. Comput. Graphics **18**, 988–997 (2012)
12. Marwan, N., Carmen Romano, M., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex systems. Physics Reports. vol. 438, pp. 237–329 (2007)
13. Webber, C., Zbilut, J.P.: Dynamical assessment of physiological systems and states using recurrence plot strategies. J. Appl. Physiol. **76**, 965–973 (1994)
14. Tolston, M.T., Riley, M.A., Mancuso, V., Victor, F., Funke, G.J.: Analysis of Semantic Coordination in Team Communications. Unpublished
15. Ruddy, M., Nissen, M.: New software platform capabilities and experimentation campaign for ELICIT. In: Paper presented at the 13th ICCRTS, Seattle (2008)
16. Quesada, J.: Creating your own LSA spaces. In: Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W. (eds.) Handbook of Latent Semantic Analysis, pp. 71–85. Routledge, New York, NY (2007)
17. Angus, D., Watson, B., Smith, A., Gallois, C., Wiles, J.: Visualising conversation structure across time: insights into effective doctor-patient consultations. PLoS ONE **7**, e38014 (2012)
18. Hodgkinson, G.P., Langan-Fox, J., Sadler-Smith, E.: Intuition: a fundamental bridging construct in the behavioural sciences. Br. J. Psychol. **99**, 1–27 (2008)
19. Bowers, K.S., Regehr, G., Balthazard, C., Parker, K.: Intuition in the context of discovery. Cogn. Psychol. **22**, 72–110 (1990)