Advances in Intelligent Systems and Computing 592

Isabel L. Nunes *Editor*

Advances in Human Factors and Systems Interaction

Proceedings of the AHFE 2017 International Conference on Human Factors and Systems Interaction, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Advances in Intelligent Systems and Computing

Volume 592

Series editor

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

About this Series

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

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

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

Isabel L. Nunes
Editor

Advances in Human Factors and Systems Interaction

Proceedings of the AHFE 2017 International Conference on Human Factors and Systems Interaction, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Editor
Isabel L. Nunes
Dept Engg. Mecanica e Industrial
Universidade Nova de Lisboa
Caparica
Portugal

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-60365-0 ISBN 978-3-319-60366-7 (eBook)
DOI 10.1007/978-3-319-60366-7

Library of Congress Control Number: 2017943024

Printed on acid-free paper

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

Advances in Human Factors and Ergonomics 2017

AHFE 2017 Series Editors

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

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

Proceedings of the AHFE 2017 International Conference on Human Factors and Systems Interaction, July 17−21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

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

(continued)

(continued)

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

Preface

Human Factors and Systems Interaction aims to address the main issues of concern within systems interface with a particular emphasis on the system lifecycle development and implementation of interfaces and the general implications of virtual, augmented, and mixed reality with respect to human and technology interaction. Human Factors and Systems Interaction is, in the first instance, affected by the forces shaping the nature of future computing and systems development. The objective of this book is to provide equal consideration of the human along with the hardware and software in the technical and technical management processes for developing systems that will optimize total system performance and minimize total ownership costs. This book aims to explore and discuss innovative studies of technology and its application in system interfaces and welcomes research in progress, case studies, and poster demonstrations.

A total of seven sections presented in this book:

I. System Interaction in Industry
II. Human Factors in Emergency Management
III. Mixed Reality Environments and Simulation
IV. Assistive Technologies and Natural User Interaction
V. Aviation and Remotely Piloted Aircraft Systems
VI. Practical Applications
VII. Systems Usability and Device Assessment

Each section contains research paper that has been reviewed by members of the International Editorial Board. Our sincere thanks and appreciation to the board members as listed below:

Denis Coelho, Portugal
Frank Flemisch, Germany
José Manuel Fonseca, Portugal
Kazuo Hatakeyama, Brazil
Christopher Lowe, UK
Ravi Mahamuni, India
Pamela McCauley, USA
Bonnie Novak, USA
Maria Papanikou, UK
Alexandra B. Proaps, USA
William Prugh, USA
Yves Rybarczyk, Ecuador
Raquel Santos, Portugal
Michael W. Sawyer, USA
Mario Simoes-Marques, Portugal
Vesna Spasojevic Brkic, Serbia
Paolo Trucco, Italy

July 2017 Isabel L. Nunes

Contents

Human Factors in Emergency Management

System Interaction in Mixed Reality Environments and Simulation

System Interaction in Industry

Operators Working with Transmission Flexibility: Enhancing Utility Control Rooms with Dynamic Line Rating Technique

Wei Zhang(✉), Katya Le Blanc, Jake P. Gentle, and Timothy R. McJunkin

Idaho National Laboratory, Idaho Falls, USA
wei.zhang@inl.gov

Abstract. This paper reflects generally on the human factors approach of integrating the Dynamic Line Rating (DLR) technique into power utilities control room as well as other monitoring systems, investigates how to accommodate transmission operators' needs and incorporate the new technology into a control room regarding these needs.

Keywords: Control room · Dynamic line rating · Transmission operators

1 Introduction

Human-system interfaces in nuclear power plant control rooms usually include traditional analog technologies. Recently several advanced digital technologies have been proposed for incorporation into control rooms. Understanding human behaviors in control rooms and integrating this knowledge into the design of the system and user interfaces in control room has benefited both the safety and economics of plants and utilities [1]. The work in a control room is highly complex, which requires timely responses from operators in the constantly changing environment. A user-centered interface design can help operators to communicate and collaborate better with the system, lower the risk of performance failure and improve the work efficiency.

There have been many proven human factors methodologies in control room design and modernization for nuclear power plants. From the technical perspective, plant systems need to be observed and analyzed in detail in order to have a good understanding of the general operation with consideration of all possible conditions of the plant; from the organizational perspective, all the operators and engineers who will be responsible for the plant operations should be coordinated and connected. Therefore the design process is rather complex taking into account these technical and human factors under all possible scenarios [2]. These same methods for designing a nuclear power plant control room can be applied into utility control room design practices, such as a transmission control room, with considerations of its specific workplace scenarios.

Integrating New Technologies into the Control Room. Outdated systems are not the only driver for control room upgrades. Sometimes new technology development is another incentive to upgrade the control room. The industry may also seek to advance

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_1

technology targets optimizing the effectiveness and efficiency of the system, reducing operation and maintenance cost or better protecting the environment. As a system is reshaped by a new technology, the control room needs to be redesigned to match the updated system.

Integrating a new technology into current control room system seems easier than redesigning the whole control room. However, it brings more challenges related to how new technology can fit into the current system seamlessly. Typically the existing system has been operated safely for many years. The operators are familiar with the system as it is. They have learned the system well and developed a mental model of the system operation in their long-term memory. Any new technology will disrupt the current workflow and processes that are well practiced. This may ultimately unintentionally increase mental and physical demands on operators. Research on air traffic control systems has shown that numerous research projects that proposed new ideas and tools have been ultimately rejected as unusable by the controllers due to unfamiliarity [3].

Operators are required to understand complex situations and perform tasks within a limited time. Human errors are inevitable in control room and have been increasingly cited as the cause of accidents. To reduce the risk of human error and to ensure the consequences of error are minimized, the user interaction of the control room should be designed to convey clear information and provide intuitive interaction without causing operators confusion. The successful integration of a new technology into an existing control room requires careful consideration of the operator tasks, working environment, and a full understanding of how the new technology will change how work is done in the control room.

Transmission Control Room. The purpose of this paper is to explore the integration of a new technology into transmission control rooms. Power transmission utilities need operators to monitor a large amount information in control rooms and make decisions quickly during emergent situations. A good monitoring display and control system can help them better perform these tasks.

Dynamic Line Rating. The advanced technology that we consider in this work is a tool that enhances the use of existing transmission systems called Dynamic Line Rating. The increase of power use by end consumers does not only stimulate more power generation from plants, but also increases the demand on power transmission. Transmission congestion has been a critical constraint of utilizing available generation capacity, particularly during peak hours. In order to solve this issue, transmission capacity in overhead power lines needs to be improved. However, building new or improving transmission infrastructure is expensive and time consuming.

One of the key factors that limit transmission capacity is line temperature. Transferring power on transmission lines heats them, which can cause excessive line sag and other conductor deteriorations if the power flows are not carefully monitored and kept below their limits. The current carrying limits of transmission lines is generally determined by static ratings based on the worst-case scenario environmental conditions. However, this estimation is very conservative, allowing usable transmission capacity to go unused. Dynamic Line Rating (DLR) is a technology that provides timely and accurate estimation of transmission line capacity based on real-time observation of the environmental conditions of overhead transmission lines such as conductor characteristics, real time current, and weather parameters such as ambient air temperature, wind speed and wind direction [4]. The technique is valuable because it can significantly increase the capacity of existing transmission lines. DLR techniques have been explored for several decades and some of them have already been implemented such as at the electric reliability council of Texas [5].

The unique Idaho National Laboratory concurrent cooling model applies the effect of wind cooling and incorporates a weather-based system that is usually less costly than other monitoring systems such as conductor sag monitors. When the wind is blowing, it can cool the lines and provide more capacity for transmitting current, thus can increase the current limit that the line can carry. Idaho National Laboratory has developed a DLR system, which includes weather measurement equipment installed near transmission lines, computational fluid dynamics models and software to consume the weather data to model and calculate transmission line ampacity limits in real time and with forecast weather and line current information provide a predicted future limit.

In order to utilize these DLR tools without increasing operator workload, the information needs to be carefully integrated into the control room. Technologies will not be optimally used if they are not well designed and implemented in human-system interfaces. Our goal is to integrate DLR technology into current transmission control rooms using human factors proven approaches.

2 Methods

To facilitate DLR integration, human factors professionals have begun by investigating the utility control rooms and existing tools and interfaces used by operators in great details using field observations and interviews. These observations informed the design of DLR tools, which are evaluated by operators to provide feedback on the designed user interactions. Three utilities have been investigated and their operators are interviewed. We summarized the observations and generalized the design requirements and prototypes based on the observation results.

3 What Are Operators' Needs?

We have observed three utility control rooms and interviewed their operators. The goal of these investigations is to understand how the current static line rating method is being used and how the dynamic line rating is going to change the current operation process if integrated.

Among different operators in these control rooms, we targeted transmission operators because they will be the primary users of DLR information. In a transmission control room, the responsibilities of a transmission operator usually include:

1. Control the physical assets of the transmission systems in response to scheduled maintenance and activities by operating equipment.
2. Monitor power flows to ensure that system is maintained within its limits.
3. Respond to any alarms or emergent events.

One of the routine activities of transmission operators is monitoring power flows against static line ratings in real-time to make sure the lines are carrying an acceptable amount of current. Most control rooms present the current load relative to the limit (typically expressed as a percentage) on the monitor. Some of the control rooms also provide other monitoring information such as n–1 contingencies for particular lines in their transmission system. According to the interviewed operators, they do not need additional information on how line rating numbers are generated or how the limit is developed as long as they can trust that the limits are accurate and conservative. Operators also indicated that they are not likely to have time to view additional detail on line ratings, and if they need to in order to make an effective decision, it would likely overload operators.

4 How Can DLR Fit into These Control Rooms?

According to the operators' feedback and working scenarios, we learned that transmission operators have limited time to make decisions especially in emergency, so they need intuitive information to reduce uncertainty especially in a highly complex situation. If DLR is integrated, operators do not need an additional display to present the new information. Otherwise, they are likely to avoid using the DLR information because it will add additional burden. What they need is the simple and clear information that can directly scaffold their decision making. Therefore, we decided to integrate the DLR calculation in the operation system behind the scenes and replace the original static line limit information with a new current load relative to limit percentage based on DLR information. This will bring a minimum change in the information for operators without increasing any workload. Apart from that, we only add one piece of information into the operator displays, which is the time to the maximum line temperature, which will indicate to the operator how much time is left for the line to hit its maximum temperature under the current or n–1 contingency current. The timing information is only shown when the line is close to its limit. This will give operators more intuitive information for decision-making. To sum up, the following information in Table 1 will be presented to operators on the monitoring display:

Table 1. Information presented on the control room display

Information	Purpose
Line name/location	To identify lines
Percentage of current load to limit based on DLR	To present line status
Time to maximum temperature	To provide intuitive information users need

Apart from the main screen, we also provide options for operators when they want to drill down to more details. Because DLR is more complex than static ratings, operators might want to look into detailed weather information to ensure the accuracy of line status occasionally. We designed an alternative display that operators can switch to for specific weather station information over a period of line, including wind, air temperature and solar heating.

Figure 1 shows two prototype examples of the monitoring displays.

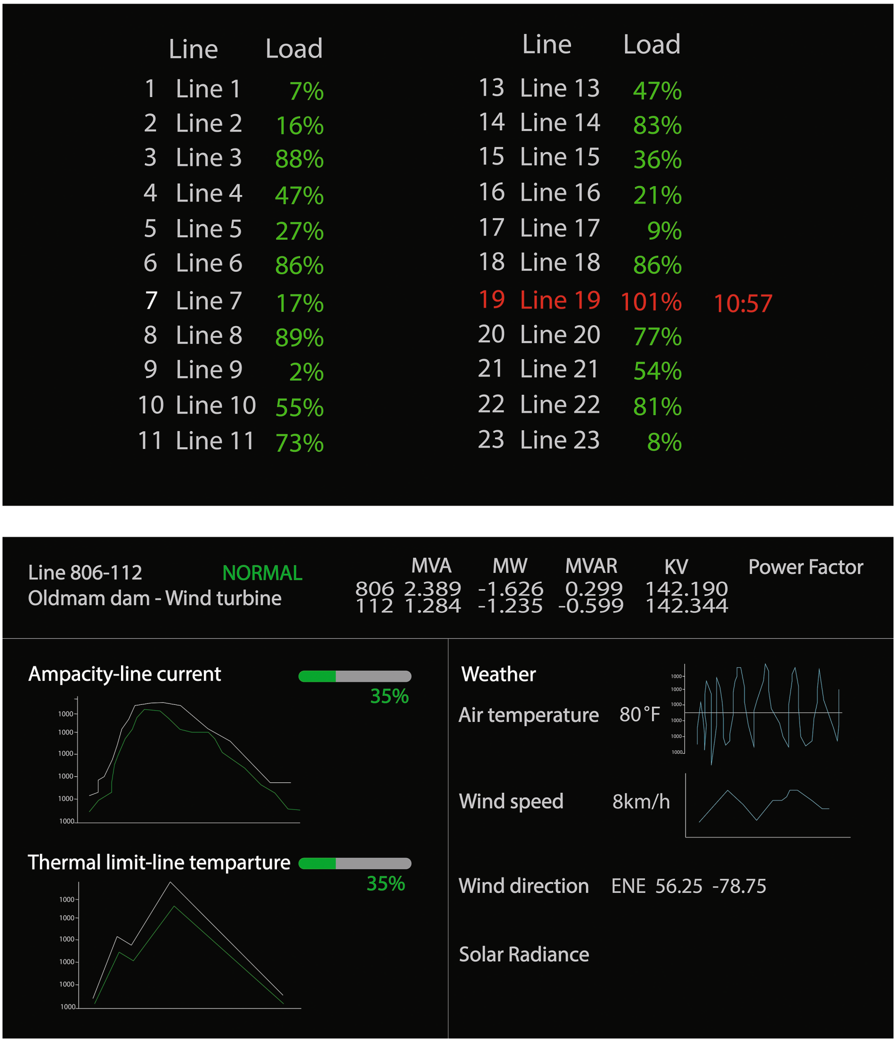


Fig. 1. Prototype examples of monitoring displays integrated with DLR information.

5 Conclusions

We have investigated the operators' activities in the control room of power utilities, including planning, scheduling and real-time operations. Based on these observations, we developed a series of tools to integrate DLR information into the current control room monitoring systems. The tools utilize DLR techniques to calculate line current behind the scenes and represent the current load relative to line limit the same as previous monitoring system shows. Additionally, we added intuitive information to show the time for a line to hit its maximum temperature limit. The system also provides detailed weather information in case operators need it. The tools will be evaluated in simulated scenarios using objective metrics in future phases of the control room integration of DLR.

As designers and implementers are responsible for introducing the new technology to users, how they frame and present the information will impact what information would be provided to operators and how this information would cue the operators' actions. Integrating a new technology is likely to increase operators' workload because it often adds new information and work processes. The additional information may also increase the uncertainty and make decision making harder. To avoid these problems, we designed a straightforward human-system interaction that is based on minimizing the changes to the information displayed and the existing work processes.

References

1. International Atomic Energy Agency: Control room systems design for nuclear power plants. Report for International Working Group on Nuclear Power Plant Control and Instrumentation, Vienna, Austria (1995)
2. Rejas, L., Larraz, J., Ortega, F.: Design and modernization of control rooms according to new I&C systems based on HFE principles. In: 2011 International Nuclear Atlantic Conference, Belo Horizonte, MG, Brazil (2011)
3. Mackay, W.E., Fayard, A.L., Frobert, L., Médini, L.: Reinventing the familiar: exploring an augmented reality design space for air traffic control. In: The SIGCHI Conference on Human Factors in Computing Systems, pp. 558–565. ACM Press/Addison-Wesley Publishing Co., January 1998
4. Gentle, J.P., Myers, K.S., Bush, J.W., Carnohan, S.A., West, M.R.: Dynamic line rating systems: research and policy evaluation. In: PES General Meeting| Conference and Exposition, pp. 1–5. IEEE, July 2014
5. Cheung, K.W., Wu, H.: Implementation of dynamic thermal ratings in the operational environment. In: FERC Technical Conference on Increasing Real-Time and Day-Ahead Market Efficiency through Improved Software, June 2014

An Approach for the Integration of Non-ergonomic Work Design as a New Type of Waste in Lean Production Systems

Uwe Dombrowski, Anne Reimer(✉), and Jonas Wullbrandt

Institute for Advanced Industrial Management,
Technische Universität Braunschweig, Langer Kamp 19,
38106 Brunswick, Germany
{u.dombrowski,anne.reimer,
j.wullbrandt}@tu-braunschweig.de

Abstract. Nowadays, Lean Production Systems are an industry standard implemented to achieve the objectives set by the management. The overall goal is to reduce costs and delivery time as well as to increase quality. By reducing waste, production processes are improved and, in turn, help to achieve these objectives. However, mostly technical or organizational changes are being made. Human Factors and Ergonomics is not seen as a potential for waste so far although non-ergonomic work causes high costs for companies. The paper outlines four different approaches for integrating non-ergonomic work in the framework of Lean Production Systems.

Keywords: Human factors · Ergonomics · Lean production systems

1 Introduction

Lean Production Systems (LPS) are commonly implemented in German companies and can be seen as an industry standard nowadays [1]. They are used to coordinate individual subsystems within the company and aim at reducing costs, increase quality and, as a result, improve the company's overall competiveness. In order to achieve these aims, processes that do not contribute to value-adding are identified and considered as waste. These non-value-adding processes often relate to technical or organizational processes such as transportation, waiting time or inventory. However, many studies examining the impact of lean production on social systems indicated that they can have a negative effect on the employees [2]. The typical lean production is often characterized by the lifting and carrying of heavy loads, static postures or repetitive activities which all have a significant impact on employees. Being exposed to those impacts result in musculoskeletal disorder and psychosocial risks [3]. For example, according to a calculation of costs arisen for the absence time of ill workers by Federal Institute for Occupational Safety and Health (BAuA) Germany, about 130 million days of incapacity to work were caused by musculoskeletal disorder in Germany 2015. As a consequence, this high number of days of incapacity to work led to a production loss of €14.1 billion [4]. Given the fact that non-ergonomic work has great potential to

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_2

contribute to the overall aim of LPS by boosting productivity, it is hardly perceived as waste. Therefore, possibilities to implement non-ergonomic work design as a new type of waste in Lean Production Systems have to be identified.

2 Work Stress and Human Factors and Ergonomics

The perspective towards the role of human work has drastically changed over the past decades [5]. In the early 20th century, employees were seen as a necessity during the production process. There was no responsibility or actions taken to keep or even improve the workers' health. In case of an accident, the worker was replaced. Over the past decades, the fundamentals of work ethics changed and the importance for workers' health and employer responsibility increased. Nowadays, the awareness for health and work-life-balance are important aims for most individuals as well as companies in order to stay as long as possible in the company.

However, production processes have a great impact on the health of the employee and result in work stresses which affect the employee. A work stress is defined in the DIN EN ISO 6385 as the total of external conditions and requirements in a work system which affect a person's physiological and /or psychological condition [6]. A work stress is not automatically a negative impact on health. It can even positively affect human health because an underchallenge of an employee can, for example, lead to a feeling of monotony and in the worst case cause illnesses [7]. Therefore, an overload as well as a underchallenge of the employee has to be avoided. Due to the diversity in personal performance prerequisites, the same work stresses lead to different strains. According to Hardenacke et al., five skills can be affected by working stress. These are endurance, cardiovascular system, perception and reaction, sensitivity and commitment as well as creativity [8]. Through the individual response of a person to a stress, various illnesses can be caused, whereby the so-called strain can be both physical and psychological. If, for example, the requirements for the employee are continuously increase by an improved production process, the stresses for the latter may become uncontrolled, which in the worst case can lead to the incapacity to work and, thus, the employee's absence times [9]. Therefore, it has to be questioned to which extent adaptions of the working system can be made without causing more stresses for the employee. If, for example, walking distances are reduced by reducing inventories or transport activities, this can lead to a reduction in load changing, thus increasing the risk of musculoskeletal disorders. In addition, endurance is less stressed by the elimination of walkways. This may be advantageous, but a minimum should be maintained to activate the cardiovascular system. The risk of cardiovascular disease, such as a heart attack, is also increasing here. Furthermore, by eliminating unnecessary processing steps and superfluous movements, compensating movements can be eliminated, which eliminates a load change and promotes musculoskeletal disorders. In addition to the physical stress, psychological stresses can arise. These are caused, for example, by work intensification or pressure to perform and can lead to a burnout in the extreme case [10].

Work-related musculoskeletal disorders are common and mainly caused by a lack of movement, lack of exercise or overload. This affects people who work on the computer as well as in the production [11]. For 2012, it is shown that musculoskeletal disorders are generally among the four greatest impacts on the populations health [12].

The manufacturing sector is characterized by high physical workloads, such as the lifting and carrying of heavy loads. As an example, the Federal Institute for Occupational Safety and Health in Germany has carried out a calculation of the economic costs related to incapacity for work for 2015 [4]. Today, the manufacturing sector is above average for with 17 days of incapacity to work per employee per year, with a large proportion of these days caused by musculoskeletal disorders. Taking into account the fact that, on average, the highest payment is paid in the manufacturing sector, this sector has the highest production loss costs of €5.13 billion per year [4].

Due to this imminent development, it is of utmost importance to focus on the health of employees. In order to achieve this objective, Human Factors and Ergonomics provides suitable methods and tools, with the focus on the prevention of occupational stresses [13]. Human Factors and Ergonomics is the scientific discipline concerned with the understanding of the interactions among humans and other elements of a system, and the profession that applies theoretical principles, data and methods to design and optimize well-being and overall performance [14].

The ergonomic workplace design is a core element of these preventive measures [15]. The design of work systems, places, products and processes are made according to criteria, which are characterized by physiological performance and psychological conditions of humans as well as their measurements [3]. Thus, new illnesses of the musculoskeletal system or an exacerbation of an existing musculoskeletal disorder can be avoided. Despite these well-known solutions, methods of Human Factors and Ergonomics are not applied adequately [3, 14]. In a review of leading experts for Human Factors and Ergonomics, four reasons were identified: limited applicability of the methods and tools, multidisciplinarity, unclear communication to the external world, and lack of awareness of the problem [14] Above all, the lack of awareness of the problem is a decisive factor in under-estimating the importance of Human Factors and Ergonomics [16]. Only the attainment of a knowledge for a problem leads people to seeking solutions. With the emerging awareness of the problem, it is to be achieved that health is maintained by preventive methods. Once this awareness has been created, this will lead to a health-conscious behavior, which will affect both the profession and private activities [3, 17]. In order to counter this lack of awareness of the problem, competences in the field of Human Factors and Ergonomics has to be developed.

3 Lean Production Systems

In modern production plants, workstations and processes are designed according to the principles of Lean Production Systems. LPS have the target to consider the aspects of technology, organization and humans equally [18]. The definition of a Lean Production System is stated in the VDI 2870 as “an enterprise-specific, methodical system of rules for the continuous orientation of all enterprise processes to the customer in order to achieve the objectives set by the enterprise management” [1, 19]. Generally, the focus is on technical and even more on the organizational process design. In order to reach higher efficiencies and, consequently, to be more competitive, elimination of wasteful activities in all company processes is the major target in LPS [14, 19, 20]. In this context, waste can be seen as all kinds of non-value-adding activities that do not contribute to the product in terms of increasing the customer value. In LPS seven different kinds of waste can be distinguished:

- Over production
- Waiting time
- Transportation
- Over processing
- Inventory
- Motion
- Defects and touch up [19].

As shown in Fig. 1, an LPS consists of different elements. On the first level, an enterprise has to define targets. In most cases, target dimensions stand for the strategic targets of an enterprise and are quality, costs and time. Since strategic targets affect the entire organizational structure they need to be referred to all enterprise processes. Enterprise processes are the second element in a LPS. Within the third element of a LPS, the strategic targets are executed. On this level, an LPS consists of different Principles that define a coherent overall framework. Each principle leads to defined methods and tools which can be used in order to achieve the targets. Methods and tools are the fourth element in a LPS [19].

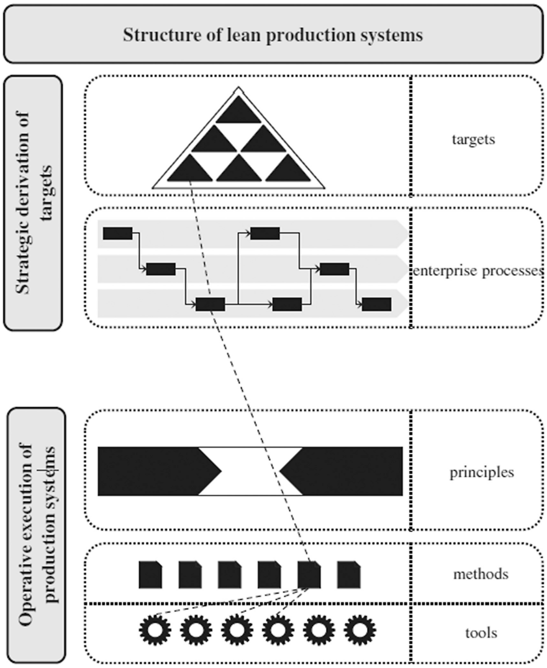


Fig. 1. Structure of lean production systems [19]

In the following the structure of a LPS is explained with an example. The top target of a manufacturing company is to improve the quality. Therefore, Sub-targets are sustainable process mastery in manufacturing and assembly-friendly product design. For this purpose, relevant manufacturing processes need to be defined, e.g., for turning, milling or grinding. As to achieve the strategic target, a suitable principle is "zero-defects-production". The principle combines methods and tools that are used to re-duce the number of defects that are passed to the next production step and to ensure a high product and process quality. Especially Six Sigma, automation, Poka Yoke and 5x Why are methods of this principle [19].

As it is mentioned above, LPS are used to comprehensively and continuously design enterprise processes. The processes are optimized to lower costs, save time or improve quality. However, methods of human factors and ergonomics are not consequently implemented in all elements of an LPS. Mostly, they are just used as methods and tools in different principles. As an example, the 5S-Method includes the cleaning of the work station. This is an important prerequisite for Human Factors and Ergonomics because it focuses the safety of the work station. However, it is not perceived as part of Human Factors and Ergonomics but to LPS. Therefore, it is assumed that there is only little awareness for Human Factors and Ergonomics in industry. In the next section, different options for the integration of Human Factors and Ergonomics in LPS are being presented.

4 Human Factors and Ergonomics in Lean Production Systems

As shown in the previous section, Human Factors and Ergonomics is a crucial factor for companies to secure employees health. It is an important mean for companies to employ healthy and motivated employees in order to stay competitive.

Lean Production Systems have the target dimensions quality, time or costs [19]. However, the way in which the product is produced is often not recognized, but has considerable implications for employees and the Society. For example, industrialization triggered such substantial changes in society that they, retrospectively, are considered technical revolutions. The change in the mode of production and thus of the production systems, has a significant influence on the quality of life and prosperity of a society [1].

Usually, the overall strategic targets of a company are to reduce costs as well as time and to improve quality. However, this is just the tip of the iceberg. Figure 2 shows different aspects which have influences on those strategic targets.

Salaries have an impact on production systems because they ultimately can reduce or, mostly, increase costs for staff. Therefore, solutions have to be found that help to compensate higher costs like automation or staff reduction. Changes like prices for resources, work-life balance, working time, education and health can have similar effects on production systems. As mentioned in Sect. 2, musculoskeletal disorders also have a significant impact on production systems [2]. However, the effects do not only work one way. Many studies over the past 20 years showed that LPS also have effects on employees. As to use the example of the LPS introduction, the zero-defects-method can not only lead to the improvement of quality but also to the work intensification for

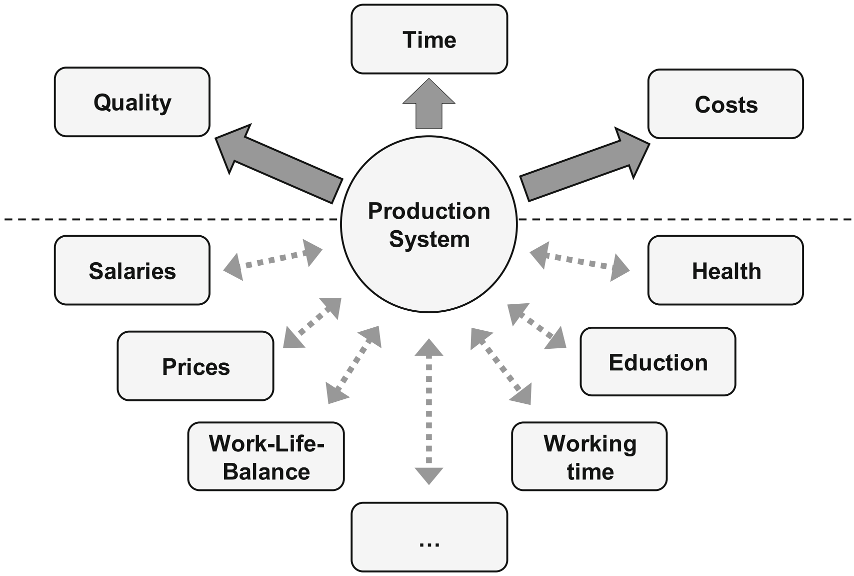


Fig. 2. Social relevance of production systems [1]

the employee. This effect can be caused by increased process controls which can e.g. lead to a reduction of hidden breaks for the employee. If the worst cause occurs, the work stress increase causes an illness for the employee and results in a long-term absence.

The example shows that LPS not only have positive effects. The studies concerning the impact of LPS on musculoskeletal disorders and psychosocial risks showed that those effects are mostly negative [2, 21]. As mentioned earlier, specific lean methods lead to an intensification of work. Therefore, compensation like additional buffers or work breaks have to be considered to avoid the negative impacts. Human Factors and Ergonomics offer several methods and tools compensate the intensification of work for the employee. For a successful integration, those methods and tools have to be integrated in an overall structure. The existing LPS framework as an industry standard can be used as a basis to implement Human Factors and Ergonomics successfully. In the next section, different possibilities are shown to implement Human Factors and Ergonomics in LPS.

5 The 9th Pillar in LPS

An analysis of existing LPS has shown that Human Factors and Ergonomics is not consequently implemented [13]. For the analysis, 20 LPS with about 800 methods were evaluated with the focus of the degree of implementation of ergonomic methods. The result was that only 23% of these methods consider Human Factors and Ergonomics. Therefore, the awareness for and implementation of Human Factors and Ergonomics has to be raised because it can create immense benefit by improving employee health as well as reducing the overall costs for enterprises. As shown in the previous section, LPS represent an industry standard in German companies and are used to give a framework for the implementation possibilities. Therefore, according to the four elements of LPS, four implementation possibilities developed. Those can be used to adapt LPS to the need of creating awareness for Human Factors and Ergonomics as well as to sensitize for musculoskeletal disorders. In the development of the solutions, particular attention was paid to the effects of the interaction of LPS in order to achieve a lasting improvement. Figure 3 shows the approaches for the sustainable integration of Human Factors and Ergonomics in LPS.

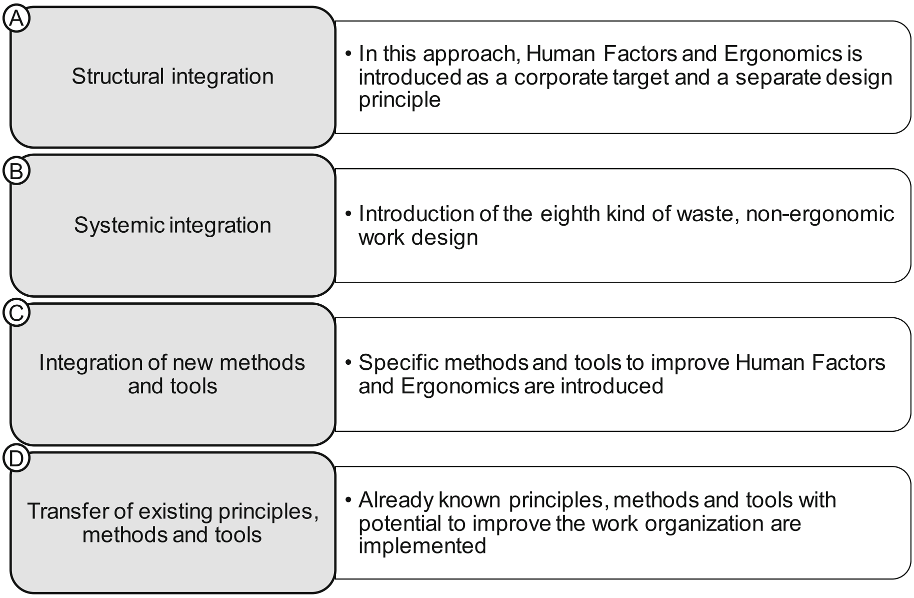


Fig. 3. Different approaches to implement human factors and ergonomics in LPS

In a structural integration, for Human Factors and Ergonomics a new corporate target and a separate design principle is integrated into the LPS. This creates the pre-requisite for bundling new methods in a design principle and ensuring a systematic application. However, in order to ensure a regular application of the design principle and the methods contained, the Human Factors and Ergonomics must also be anchored in the company's objectives. This is the only way to pass through the cascade from the target to the tool.

Furthermore, the structural integration is of particular importance because the structure of a LPS is usually visualized and seen as a symbol for the entire LPS. For this reason, the structure is usually depicted in company presentations and other marketing tools. Thus, the structural integration also leads to an improved perception of Human Factors and Ergonomics. As a consequence, all employees are obliged to life and secure methods of Human Factors and ergonomics. Therefore, is leads to a higher awareness on all levels from management to the shop floor. Especially man-agers need to not only implement these methods but also to become role models and to encourage other managers and employees to do the same. The most common forms of visualizing a LPS are the house, the circle and the product of the company [21]. Figure 4 shows the structural integration using the example of the VDI Guideline 2870.

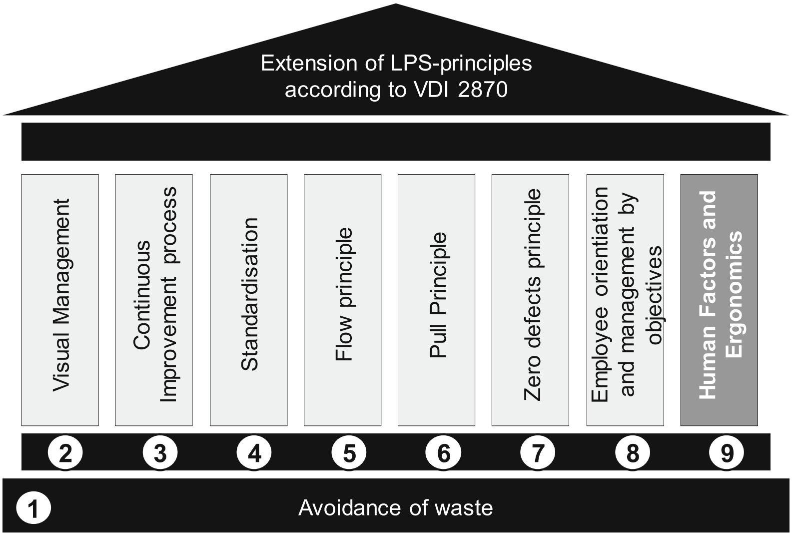


Fig. 4. Extension of LPS-Principles according to VDI Guideline 2870

The next approach proposes a systemic integration of Human Factors and Ergonomics in the LPS. In particular, the adaptation of the design principle “avoidance of waste” is an option since this is the basis for the other design principles. The focus of waste prevention is the elimination of any activities that do not contribute to the added value of the customer [4]. In the future, it will no longer be sufficient to consider the avoidance of waste only from the customer’s point of view. Rather, aspects such as the health of the employees must also be the focused. Accordingly, it is necessary to expand the seven types of waste. A possible extension is to add Human Factors and Ergonomics. Thus, the principle is kept more general in order to cover all aspects of the field.

The integration of new methods and tools into the LPS structure is necessary for the operational implementation of Human Factors and Ergonomics. This enables employees and managers to implement Human Factors and Ergonomics in their daily working routines. A study of different LPS showed that only very few methods have been anchored in the LPS for the purpose of improving non-ergonomic work [13]. In particular, methods for assessing and improving ergonomics can be integrated. Such methods are already used in companies today. However, they are not an integral part in LPS and are there not effective. In recent years, the European Assessment Worksheet (EAWS) has become a standard for industry in the assessment of ergonomics [22].

In addition to the EAWS, other methods besides ergonomic assessment like appropriate work design or occupational health management should be integrated into the LPS.

The idea of the fourth integrational approach is to transfer existing LPS design principles, methods and tools to Human Factors and Ergonomics. An important prerequisite for this approach is the enlargement of the concept of waste, which has been explained above. As a result of this extension, methods for avoiding waste can also be used in a targeted manner to shape human work (e.g. PDCA, 5× why or benchmarking). But other methods can be applied as well. By using Poka Yoke, for example, employees can be not only assisted to not produce accidental mistakes, but also to avoid lifting heavy loads. The conventional 5S method is also well transferable. By adding another S in the sense of sorting, selecting, keeping clean, making sure, standardizing and self-discipline there would be an extension to 6S. Thereby, an improvement of the work safety would be achieved.

Not only methods, but also design principles can be transferred. The design principle of standardization is particularly suitable. The transfer would not only determine the best processes in terms of quality, time and costs, but also take into account the criteria of Human Factors and Ergonomics. A zero-defects principle could also be conceived, which would combine methods and tools that would contribute to a continuous reduction of the days of incapacity to work.

6 Conclusion

Musculoskeletal disorders have a great impact on the productivity of companies. As a result, a consistent and sustainable human work design is required. Today, modern companies often design their workplaces and processes according to the principles of LPS, which in recent years have more and more developed to a standard in industry. LPS focus mainly on monetary aspects such as quality, time and costs when it comes to work design. Other aspects like Human Factors and Ergonomics are only considered little. Therefore, even though Human Factors and Ergonomics can contribute to the economic aspects by reducing days of incapacity to work, there is no awareness for this. As to raise awareness and implement Human Factors and Ergonomics in LPS, four different approaches for the sustainable integration were presented in this article. The first approach provides for the integration of Human Factors and Ergonomics into the structure of the LPS. In addition, a systemic integration of Human Factors and Ergonomics as well as the integration of new methods and tools are presented. The transfer of existing design principles, methods and tools forms the fourth part of the solution approach. By implementing Human Factors and Ergonomics into LPS, not only the awareness of all organizational members is being raised but also a systematic implementation is created. Therefore, it is integrated in an overall framework and is harmonized with the other enterprise targets.

References

1. Dombrowski, U., Mielke, T. (eds.): Ganzheitliche Produktionssysteme. Aktueller Stand und zukünftige Entwicklungen. Springer, Berlin (2015)
2. Koukoulaki, T.: The impact of lean production on musculoskeletal and psychosocial risks: an examination of sociotechnical trends over 20 years. Appl. Ergon. **45**, 198–212 (2014)
3. Schlick, C., Bruder, R., Luczak, H.: Arbeitswissenschaft. Springer, Berlin (2010)
4. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA): Volkswirtschaftliche Kosten durch Arbeitsunfähigkeit 2015. Dortmund (2017)
5. Karwowski, W.: The discipline of human factors and ergonomics. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics. Wiley, Hoboken (2012)
6. DIN EN ISO 6385: Ergonomic Principles in the Design of Work Systems. Beuth-Verlag, Berlin (2004)
7. Furnham, A.: The Psychology of Behaviour at Work. The Individual in the Organization. Psychology Press, East Sussex (2012)
8. Hardenacke, H., Peetz, W., Wichardt, G.: Arbeitswissenschaft. Carl Hanser Verlag, München (1985)
9. Dombrowski, U., Reimer, A., Ernst, S.: Occupational stresses on employees in lean production systems. In: Goossens, R.H.M. (ed.) Advances in Social & Occupational Ergonomics, Advances in Intelligent Systems and Computing, vol. 487, pp. 215–226. Springer, Cham (2016)
10. Cordes, C.L., Dougherty, T.W.: A review and an integration of research on job burnout. Acad. Manag. Rev. **18**, 621–656 (1993)
11. Breithecker, D., Teichler, N., Walter, U.: Ergonomie. In: Kaufmännische Krankenkasse, K.K.H. (ed.) Weißbuch Prävention 2007/2008. Beweglich? Muskel-Skelett-Erkrankungen – Ursachen, Risikofaktoren, und präventive Ansätze. Springer Medizin Verlag, Heidelberg (2008)
12. Vos, T., Flaxman, A.D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., et al.: Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet **380**, 2163–2196 (2012)
13. Dombrowski, U., Hellmich, E., Mielke, T.: Appropriate work design in lean production systems. In: López-Paredes, A., Pérez Ríos, J., Iglesias, C.H. (eds.) Managing Complexity. Challenges for Industrial Engineering and Operations Management, pp. 129–138. Springer, Heidelberg (2012)
14. Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W., Wilson, J.R., van der Doelen, B.: A strategy for human factors/ergonomics: developing the discipline and profession. Ergonomics **55**, 377–395 (2012)
15. Liebers, F., Caffier, G.: Berufsspezifische Arbeitsunfähigkeit durch Muskel-Skelett-Erkrankungen in Deutschland. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund (2009)
16. Wullbrandt, J., Downing, C.G.: Stress fracture: adverse effects of lean initiatives. In: ASEE's 123rd Annual Conference and Exposition. American Society for Engineering Education, New Orleans (2016)
17. Rastetter, D.: Kompetenzmodelle und die Subjektivierung von Arbeit. Verbindungslinien zweier arbeitswissenschaftlicher Ansätze. In: Schreyögg, G., Conrad, P. (eds.) Management von Kompetenz. Gabler Verlag, Wiesbaden (2006)
18. Spath, D.: Ganzheitlich produzieren. LOG_X Verlag, Stuttgart (2003)

19. VDI: Guideline 2870—Part 1: Lean Production Systems—Basic Principles, Introduction, and Review, vol. 2870-1. Beuth-Verlag, Berlin (2014)
20. Ohno, T.: The Toyota Production System—Beyond Large-Scale Production. Productivity Press, Portland (1988)
21. Landsbergis, P.A., Cahill, J., Schnall, P.: The impact of lean production and related new system of work organization on worker health. J. Occup. Health Psychol. **4**(2), 108–130 (1999)
22. MTMaktuell: Instruktoren aus fünf Ländern. European Assembly Worksheet (EAWS). MTMaktuell **1**, 24–25 (2011)

A Novel Concept for a Collaborative Dashboarding Framework

Sebastian Schöffel(✉), Greg Weibell, and Johannes Schwank

Computer Graphics and HCI, University of Kaiserslautern,
67663 Kaiserslautern, Germany
{schoeffel,weibell,schwank}@cs.uni-kl.de

Abstract. Dashboards are commonly used to visualize, analyze, interact with, or to present data in various forms. Their application domains are unlimited, e.g. they can be used as an information desk in the town hall, as a platform in meetings, as an analysis tool for business intelligence, or for providing an overview in disaster scenarios etc. Nevertheless, current dashboarding solutions mostly incorporate only rudimentary collaboration features. After an extensive state-of-the-art research on current dashboard and collaborative solutions, we present a concept for a web-based dashboarding framework with collaboration as the driving force for the design. Inspired by three highly collaborative use cases, our concept aims at innovative sharing possibilities, intuitive and powerful collaboration features, and a flexible design. We conclude with describing how our three guiding use cases can be conducted using the developed concept and give an outlook on our future work on enhancing this high-level design.

Keywords: Collaboration · Dashboards · Collaborative environment · Dashboard design guidelines · Dashboarding framework · Web-based dashboards

1 Introduction

According to the EMC Digital Universe, we will have "nearly as many digital bits as there are stars in the universe" [1]. One can expect the digital universe to grow by a factor of 300 between 2005 and 2020 [2]. Every second, millions of data points are collected and stored, whether required for further analysis or not. According to estimates from 2000, only 25% of all stored data is used for later analysis [3], a value one can expect to have decreased since then. Besides the fact that it is impossible to analyze all available information (for various reasons), an effective visualization plays a key role for gaining a more profound understanding and for successful decision making.

Dashboards can be an effective means to visualize and analyze data. A dashboard typically consists of a number of graphical widgets arranged in a grid-like layout, which may or may not allow for user interaction. Inspired by instrument panels found in vehicles, dashboards can help users to identify trends or problems, measure performances and to identify data anomalies and correlations. Thus, dashboards have become an integral part of Business Intelligence (BI) for analyzing key performance indicators (KPIs). Additionally, dashboards are used as tools in many application domains, e.g. for creating an overview of relevant data (such as machine statistics in a

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_3

manufacturing line), as information platforms on large screens in meetings, as public news/process displays, for disaster scenarios, etc. [4–6].

In most cases, dashboard solutions are developed or adapted exclusively for a specific application domain or use case (see examples above) and a single user usage. Nevertheless, there already exist some dashboard frameworks and concepts which claim to be highly flexible and customizable. After an extensive state of the art research with a focus on collaboration capabilities, we ascertained that current dashboard solutions only incorporate rudimentary collaboration features (such as sharing a URL to a dashboard or exporting snapshots) and that there is no single solution that makes use of more sophisticated collaborative approaches. Such concepts for collaborative work environments do exist, but they have not been applied to dashboarding frameworks, which has motivated the creation of this concept.

The outline of the paper at hand is as follows: First, we present different existing definitions of dashboards and a list of frameworks which inspired our concept; second, we explain certain collaborative use cases which may be solved using a collaborative dashboard approach; third, we present our design concept for a collaborative dashboard approach. This section places a special focus on solutions for supporting collaboration as well as relevant aspects for a dashboard framework in general. To show the usability of our concept, we explain how the use cases presented in the third section can be supported by the usage of a dashboard system implementing our design concept.

2 Related Work

"Dashboard" is a term widely used in various environments. Alexander et al. define dashboards as "a visual interface that provides at-a-glance views into key measures relevant to a particular objective or business process" [8]. According to Few, author of *Information Dashboard Design*, "'dashboard' is simply a new name for the Executive Information Systems (EISs) first developed in the 1980s" [7]. Dashboards are used in a wide variety of areas, such as the medical field [9], in software development [10], in the educational field [11], and in the business sector [12]. Dashboards may also be used as interactive tabletops [13], and location-based capabilities make them suitable for construction management [14].

In terms of dashboard design, Alexander et al. state: "Most dashboards are designed around a set of measures called Key Performance Indicators (KPIs). A KPI is an indicator of the level of performance of a task deemed to be essential to daily operations or processes" [8]. Visualization of these KPIs is not standardized, though design guidelines do exist [7, 15, 16].

In summary, we define a dashboard as a container typically consisting of independent graphical elements (widgets), each representing data in a specialized way (for example by using charts and graphs). More advanced dashboards update their widgets in real-time and allow for user interaction (e.g. the user can zoom into a graph). Nowadays, web-based dashboard solutions are common but differ in their requirements. Generally, web-based dashboards are meant to be displayed on a web browser — usually through JavaScript code running on the client-side — with computation, data storage, and (optionally) user management running on a dedicated server.

"Dashboarding software", though ambiguous, matches a variety of projects and products. In the paragraphs below, we list noteworthy dashboarding solutions that have considerably inspired our approach. Due to space limitations, we only mention their impact on our collaborative design:

- **Grafana** (grafana.org) is a popular open-source monitoring dashboard offering authorization support, URL sharing and allowing for sensitive data to be stripped (snapshots).
- **Dashing** (dashing.io) is a simplistic approach primarily used for presentation purposes. It offers no user management or collaborative environment but serves as a demonstration of intuitive rearranging of real-time widgets.
- **Dashbuilder** (dashbuilder.org) is a complex dashboard and reporting platform, offering user management, dashboard administration, as well as interactive widgets. It includes data export, intuitive drag and drop configuration, in-line editing of queries, and various chart library options.
- **Ganglia** (ganglia.info) uses RRDtool, a widespread graphing and visualization framework with limited user interaction. Pages are inherently static and there is no user management.
- **FusionCharts FreshForks** (fusioncharts.com/dashboards/collaboration/) is a commercial chart-based dashboard system with interactive widgets. Its collaborative features are limited to a comment section for charts and through an activity feed.

In addition, the following incomplete list of non-dashboard solutions inspired our approach. They all feature interesting collaborative possibilities, which should be more or less included in a collaborative dashboard approach.

- **Collaborative real-time editors** such as Etherpad (etherpad.org) and the office suite offered through Google Drive, enable multiple users to open and edit a document simultaneously and in real-time.
- **Wikis, e.g. MediaWiki** (mediawiki.org) allow users to work together on one document using a special markup language.
- **Trello** (trello.com) allows users to manage projects by joining teams and collaboratively working on boards and lists in real-time. Users can subscribe to changes and view a board's history through an activity feed.

The solutions above are limited with regards to either dashboarding or collaborative capabilities. The paper at hand builds on those existing concepts and introduces new approaches to allow for a more improved collaborative environment. Defining these concepts in an abstract yet comprehensive way proves to be difficult. Hence, the following chapter constructs several use cases. These, in our opinion, serve as a basis to defining the minimum requirements for a collaborative dashboard system and demonstrate its capabilities using real-life scenarios.

3 Use Cases

To convey the functionalities important to a collaborative dashboard solution, the following fictive use cases are presented. Each exemplary use case differs in setting and actors, resulting in different requirements and application domains.

3.1 Meeting Scenario

Several members of staff (SM) as well as the head of production (HP) hold a meeting in a conference room. Each SM is responsible for a certain area of a production line in building different sub-products. On a large screen, the HP's overview dashboard is displayed with different key performance indicators (such as total number of products from the last week, quality issues etc.).

Each SM can access his own personalized dashboard (individually designed according to his tasks) through his own device in front of him. Today, the CEO is also present in the meeting. He requests access to the HP's dashboard, who in turn accepts the request.

Shortly after, the HP (leading the meeting) asks the group of staff members to investigate the cause for large delays between two production steps occurring periodically. He grants them access to his dashboard for a week by sharing it with a user group they are all members of.

During the next meeting one of SM is asked to present his findings. The SM does so by interacting with the widgets on his own device, while all of his mouse movements are mirrored on the main screen, allowing participants to follow along. Soon after the meeting, member access is automatically revoked for this user group.

3.2 Factory Scenario

In a modern factory, many process steps are under computer supervision. With detailed logging functionalities, it is possible to analyze many process parameters, machine settings, input and output values, etc.

In the following setting, a screen is installed over every machine in a factory, displaying the current machine status. Additionally, key performance indicators on every production line, such as current output, delays, important values of certain machines, or elements such as a clock, remaining time on this shift, etc. are displayed on a large screen.

A member of the works committee walks through the production hall and wants to see detailed information on one production line. Because of his authority in the organization, he is allowed to access all dashboards inside the factory. Therefore, he scans the QR code visible on the dashboard on top of the production line where he is interested in and can see the dashboard on his own mobile device and analyze the data (e.g. see if the worker had a legally required breakfast break).

Another possible use case is that of maintenance: A maintenance worker is ordered to repair a machine after quality control identified this specific machine to be defective.

He scans the QR code to connect to the dashboard on his own device and is immediately given elevated *member access* by virtue of his user privileges. He can now view previous comments by other maintenance workers and has full access the dashboard and can investigate the cause of the issues.

3.3 Public Services Scenario

Nowadays, large screen displays are often used in the public (e.g. in town halls, malls, at bus stations, etc.). These could reasonably be used, for example, in disaster and crisis scenarios.

Consider a dashboard provided by the local city administration. Widgets could be used to display news from governmental institution such as law enforcement, firefighting, or the medical department, but also satellite images, weather forecasts, local twitter news feeds or other information such as important phone numbers, advices, timetables for buses, or road maps.

In an emergency, owners of public displays can display an emergency dashboard by opening a pre-shared link on all devices. Nearby users can access them (e.g. by scanning the attached QR code) and carry up-to-date information with them on their smart phone. The information displayed in this dashboard is updated by the respective organizations in real-time, allowing all users, whether connected with their device or viewing dashboards in public.

Users can interact with the dashboard on their mobile phone and access additional information (e.g. navigation options from maps), context-based content (e.g. location-related information such direction and distance of the nearest water supply) and have access to potentially life-saving information.

4 Design Concepts

This section focuses on the core design concepts of our approach. First, an overview is given, after which user management, navigation, sharing as well as collaboration features are discussed.

4.1 General

Our web-based collaborative dashboarding concept is designed for both desktop and mobile devices. It aims at an intuitive user interaction with the system as well as powerful collaborative possibilities involving multiple users in real-time.

A *widget* is a graphical element running sandboxed code, used to visualize data (e.g. through charts), to display a news feed, a form or other data, usually by connecting to one or more data sources using a *data connector*. A *dashboard* is a collection of an arbitrary amount of widgets arranged in a user-defined layout with controls to edit and share the dashboard. We refer to a person logged into the system as a *user*; if using the limited *open access* method described below, he is referred to as a *viewer*. Users can belong to an arbitrary amount of *user groups*.

Widgets can use the dashboard API for using context-based parameters (such as local time of the day or the user's location) and interaction possibilities like comments and annotations (see below). In an emergency scenario, this would allow for a dashboard to constantly display the user's location on a map, show the direction and distance to the nearest shelter, in how many minutes an emergency vehicle is to be expected to arrive, and to call for help to the current location at the press of a button.

For the sake of brevity, this paper focuses more on the concept of a collaborative dashboard as a whole and less on implementation details for widgets and data connectors. Suffice it to state that widgets are flexible in nature and can be as simple as a frame displaying data from an external website or a list of links to other dashboards, or as complex as an interactive chart allowing users to zoom into it, open a legend, highlight sections or annotate it for others to see. Widgets draw their data from an arbitrary amount of sources, either through internal means or through data connectors provided by the dashboard on a global level (e.g. for an internal database). This allows users to create flexible widgets without much background knowledge.

As outlined in Fig. 1, a dashboard is a resource that may be shared with other users or user groups. The users' interaction may be broadcast to all other users who are currently viewing this dashboard (*live mode*, described below).

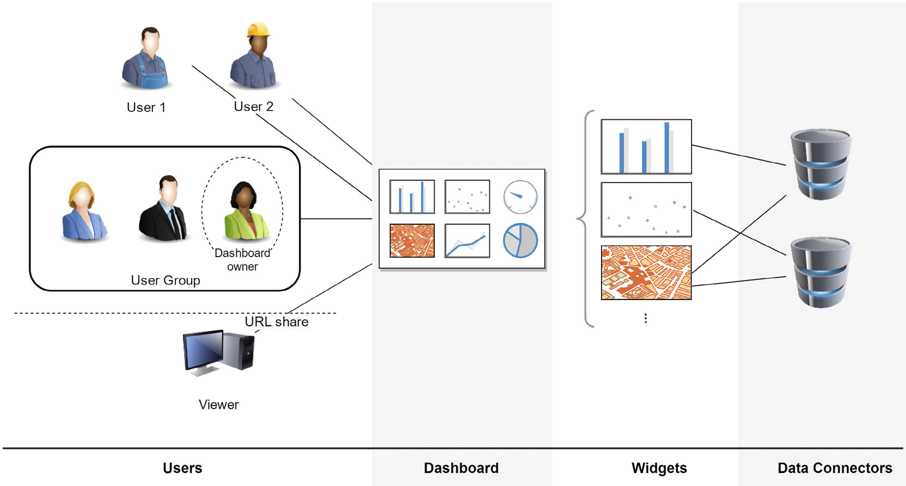


Fig. 1. Dashboard concept overview.

4.2 User Management

A dashboard user logs into the system with his *user* account credentials (email address/username and password). Users may belong to multiple *user groups*, e.g. reflecting departments in an organization or representatives meeting on a regular basis. Dashboards shared with a user group (in contrast to a single user) are accessible by each of its members. Thus, changes in company personnel only require the user group to be edited without changing with whom each individual dashboard is shared.

Users and user groups may also be organized hierarchically (and transitively if required), e.g. allowing project managers to have access to dashboards of their team leaders (but not vice-versa) and optionally of their respective team members.

Dynamic user groups are those in which members are automatically added or removed depending on specific circumstances, such as user location, time of day or week, or network-related conditions. Possible use cases include having all users in a department gain access to dashboard in the morning for planning purposes or limiting access to a specific site (location and/or network-related).

Dashboards may also be shared with others without having to log in (or not having a user account), by opening a unique URL, authorizing them to view the dashboard using the *open access* method (see below). We refer to such users with this kind of limited access *viewers*, and a common use would be to have a screen constantly display one specific dashboard (e.g. on a production line).

4.3 Navigation

After successfully logging in, users start at a main page ("Home"), i.e. the top-level page in the hierarchy. From this page, they have access to all dashboards they are members of.

A user navigates to different dashboards or other pages through hyperlinks or buttons, interacting with interface used by the respective device, i.e. keyboard and mouse input on desktops or by tapping a touchscreen on mobile devices or tabletops. Since the dashboard framework is designed as a web application, intuitive browser controls such as traversing backwards and forwards through the browser's history or setting bookmarks (for faster access) are supported as additional and intuitive navigation methods by giving each dashboard a unique URL.

Users are also able to search for dashboards by searching for their title or assigned *tags*. Tagging dashboards with descriptive keywords has the added advantage of groping them together when showing search results, provided the dashboard is not marked "private" by one of its owners. If a user is not a member of a dashboard found in the search results, he is given the option to request access (see below).

Viewing dashboards is not restricted to a single browser instance, different browser tabs or windows and even devices can display different dashboards simultaneously, thus allowing a user to both be logged in on his desktop device while walking around and accessing dashboards on his smartphone.

4.4 Dashboard Sharing

Collaboration is dependent on the user's ability to share information and the environment with others. Sharing a dashboard requires a user to have the necessary privileges (i.e. he is a dashboard owner). When sharing it with other users, they in turn become its members and can access and view it, interact with it, discuss it, and thus collaborate. Having multiple owners per dashboard ensures that it can be used in meetings using *live mode* (see below) even if one owner is not present.

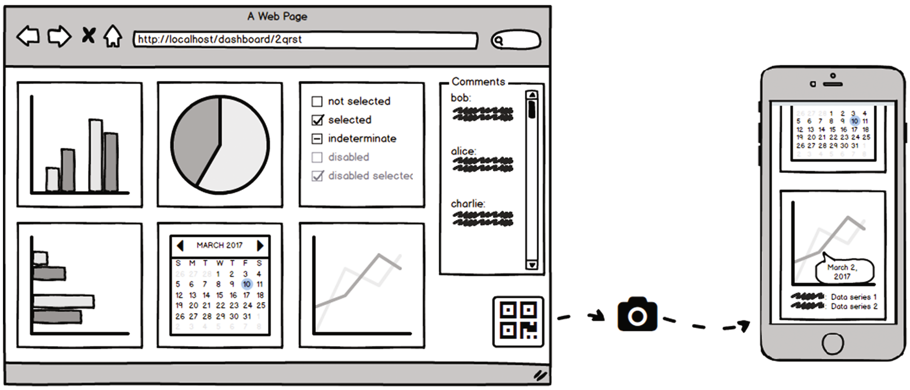


Fig. 2. Mockup of a dashboard with privileges (left) together with a smartphone (right) using the open access option through the QR code (without access to the comment section) and viewing additional information by interacting with a widget.

Owner privileges allow users to have full control over a dashboard, i.e. they can edit it and its sharing settings. Initially, the owner of a newly created dashboard is the user who created it. Through the sharing options of the dashboard, he can grant other users *owner* privileges and can add *members* — users allowed to access the dashboard and to comment on it. Owners automatically inherit member privileges.

Having *member* privileges is not the only way to access a board. We differentiate between two types of sharing methods:

- Users given **member access** can fully interact with a dashboard and its members.
- **Open access** allows viewers to access the dashboard with limited functionality.

Whereas dashboard members have full access to a dashboard and its collaborative features, viewers are restricted to a limited version. Any commenting and interactions with other users is disabled and widget functionality is limited, depending on widget configurations. This would be beneficial when a company shares a dashboard containing non-sensitive information by linking to it through their website while still allowing dashboard members to discuss the information contained therein privately.

Several options for granting other users *member* privileges exist. First, an owner can open the dashboard and select the *share option* and add one or more members to the list by either entering their username, email address or by searching for them in the directory. As an alternative, one can enter the name of a user group, thereby granting member access to all of its members at once. In addition, users may also request access to the dashboard by opening the unique URL associated with the member access request. By default, such requests result in the owners receiving a notification asking him to accept or reject the request before access is granted.

An owner may choose to constantly display a QR code with this encoded URL in one corner of the dashboard allowing nearby users to become members far easier than having the owner enter them manually. Similarly, the dashboard can enable the *open access* option and constantly display a (second) QR code with the respective URL embedded, which would allow others to access it as a viewer with limited options. When an owner sets up open access, viewers do not need to wait for permission when accessing a dashboard this way. Typical use cases for viewers are using a large screen to display a dashboard in an office, or displaying a bus schedule on a bus stop (Fig. 2).

Shares, whether through member or open access, can be customized with regards to permissions and restrictions; access can password-restricted, limited by access count, time of day, day of the week, or geographically (e.g. via GPS coordinates, Bluetooth, or Wi-Fi).

4.5 Collaboration

Our concept facilitates a collaborative environment by employing a variety of features.

Private Messaging. User can message each other privately and are notified about new messages. "Public messaging" is achieved through the commenting features described below.

Comments and Annotations. Members of a dashboard can make comments by posting to its comment section. Comments may include references, for example to other users (resulting in them being notified about it), to another dashboards (via their unique URL) or to widgets in the current dashboard.

Widgets supporting the comment API allow users to refer to specific data points in a comment. For example, a user can click or tap on such a widget displaying a city map and have the comment field populated with a reference to the precise location. When the comment is submitted and another user clicks on the reference, the associated widget and the data point referenced (in this case the location on the map) are highlighted.

In a similar fashion, certain widgets can be annotated by submitting a comment about a data point. If, for example, a member wishes to explain an anomaly in a line chart and submits a comment, this comment can be displayed as an annotation in the chart.

For privacy reasons, a dashboard's comment feed and annotations are only visible to its members and, consequently, not through open access.

Notifications and Watchdog. Users are notified about certain events and can open their notification feed containing a recent history at any time by clicking or tapping on the corresponding icon. Users can choose to be sent emails whenever they receive a new notifications and rate limit these emails as needed.

Members of a dashboard can subscribe to a dashboard and/or its comment feed to be notified whenever any changes occur, e.g. if a dashboard owner edits or adds a widget, or a when new comment is posted.

Live Mode. One of the main features of our concept is real-time interaction with other users. When a dashboard owner enables *live mode* and other members become participants by *joining* this mode, any interaction made by one user is seen by all other participants on their own screen (Table 1).

In this mode, not only are widget interactions mirrored, such as zooming into a graph, but also any mouse cursor (pointer) movements, by displaying a cursor icon for each participant. Three potential problems with this feature are addressed as follows: First, touch devices use the most recent point of contact; second, since devices have different screen sizes, only pointer movements while within a widget are mirrored; third, in order to avoid confusing disorder, (1) each participant is assigned a different color, (2) the pointer fades out over time and becomes fully opaque once it moves again. This allows the current speaker in a meeting to direct other participants' attention to a specific widget and its contents.

Table 1. Differences contrasting access and live mode options.

	Live mode	Private mode
Member access	All participant interactions are shared with all other participants	Interactions are neither shared nor visible by user
Open access	Live mode not available	Limited access without comment section

5 Use Case Analysis

In the following section, we briefly describe the components of our concept which have been used for performing the use cases of section three.

In the **meeting scenario**, we have different users: the CEO, the head of production (HP) and different staff members (SM, all of which are members of the same user group "SMG"). The HP opens his dashboard on the large screen device. The CEO requests access to the HP's dashboard by scanning the QR code displayed on screen, resulting in the HP receiving a notification to accept the request. The HP gives the staff members *member access* by sharing the dashboard with the pre-defined user group "SMG" and setting it to expire in one week.

During this time, the SM can access the dashboard and collaborate using the comment section. During the next meeting, the SM presenting their findings asks the HP to enable *live mode*, allowing everyone participating to see his mouse movements and interactions mirrored on the main screen and (if desired) on their own screen. After the meeting, membership of the user group "SMG" expires.

With regard to the **factory scenario**, we have a number of "open access" viewers, represented by multiple screens inside the factory, each showing the corresponding dashboard. All dashboards display a QR code, enabling either the member of the working council or the maintenance worker to connect to each dashboard and analyze the data needed on their own device. Furthermore, the maintenance worker has access permissions to modify the visible widgets on the screen to view additional details (e.g. access to temperature values in real-time).

Concerning the **public services scenario**, our concept can also be fully applied. The widgets in this dashboard behave similarly to iframes, in that they simply show excerpts of web pages from governmental institutions, possibly through RSS feeds. There is no need for them to maintain an extra service, resulting in no extra effort and a high acceptance rate. In case of an emergency, the administrators of the public displays need only open the web browser with the URL they received from the local administration with everything being ready to go.

Users simply scans the QR code displayed with the public dashboard to gain *open access* and to receive updated information in real-time.

6 Conclusion and Future Work

In this paper, we presented a high-level concept for a dashboarding framework utilizing collaborative functionalities. Designed as a web-based application, a wide range of possible application domains are served. In our concept, we defined key terms for dashboard frameworks as well as concepts with regard to user management, navigation, sharing possibilities, as well as powerful collaboration functionalities.

For the purpose of understanding how potential facilities might benefit from using collaborative dashboards, we described possible use case and discussed these scenarios, particularly in terms of transferability, after the presentation of our concept.

The next steps are to extend our concept to a more low-level, fine-granularly description with more details, especially concerning user roles or privileges (including sharing modalities), widget functionalities, and inter-widget communication possibilities. We also plan to extend the only slightly introduced context-based dashboard concept in a future step to enhance e.g. proximity functionalities or to support more extensive collaborative possibilities.

Although different use cases are already discussed, we plan to extend them and search for more application scenarios for evaluating our concept, making it widely applicable. Finally, our future work includes the concept's implementation as a prototype using current web technology and design standards.

Acknowledgments. This research was partially funded the German Federal Ministry for Economic Affairs and Technology in the context of "Smart Data - Innovations in Data", Grant No. 01MD15004E.

References

1. Executive Summary on Data Growth, Business Opportunities, and the IT Imperatives. https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm
2. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf
3. Müller, W., Schumann, H.: Visualisierung: Grundlagen und Allgemeine Methoden. Springer, Berlin (2000)
4. Olech, P.S., Cernea, D., Meyer, H., Schoeffel, S., Ebert, A.: Digital interactive public pinboards for disaster and crisis management-concept and prototype design. In: Proceedings of the International Conference on Information and Knowledge Engineering (IKE, WorldComp) (2012)

5. Karavakis, E., Andreeva, J., Khan, A., Maier, G., Gaidioz, B.: CMS dashboard task monitoring: a user-centric monitoring view. J. Phys. Conf. Ser. **219**(7), 072038 (2010). IOP Publishing
6. Cheng, C.K., Ip, D.K., Cowling, B.J., Ho, L.M., Leung, G.M., Lau, E.H.: Digital dashboard design using multiple data streams for disease surveillance with influenza surveillance as an example. J. Med. Internet Res. **13**(4), e85 (2011)
7. Few, S.: Information Dashboard Design: The Effective Visual Communication of Data. Analytics Press, Oakland (2006)
8. Alexander, M., Walkenbach, J.: Excel Dashboards and Reports. Wiley, Hoboken (2013)
9. Nagy, P.G., Konewko, R., Warnock, M., Bernstein, W., Seagull, J., Xiao, Y., et al.: Novel, web-based, information-exploration approach for improving operating room logistics and system processes. Surg. Innov. **15**(1), 7–16 (2008)
10. Biehl, J.T., Czerwinski, M., Smith, G., Robertson, G.G.: FASTDash: a visual dashboard for fostering awareness in software teams. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1313–1322. ACM, New York (2007)
11. Kim, J., Jo, I.H., Park, Y.: Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. Asia Pac. Educ. Rev. **1**, 13–24 (2016)
12. Haupt, R., Scholtz, B., Calitz, A.: Using business intelligence to support strategic sustainability information management. In: Proceedings of SAICSIT 2015. ACM, New York (2015)
13. Martinez Maldonado R., Kay J., Yacef K., Schwendimann B.: An interactive teachers dashboard for monitoring groups in a multi-tabletop learning environment. In: Intelligent Tutoring Systems, ITS 2012. LNCS, vol. 7315. Springer, Heidelberg (2012)
14. Guerriero, A., Zignale, D., Halin, G.: A zoomable location-based dashboard for construction management. CDVE 2012. LNCS, vol. 7467. Springer, Heidelberg (2012)
15. Juice Analytics: A Guide to Creating Dashboards People Love to Use: Translating Delicious Data into a Beautiful Design, Version 2.0. (2015). https://www.juiceanalytics.com/white-papers-guides-and-more/. Accessed in 2017
16. Maeda, J.: The Laws of Simplicity. MIT Press, Cambridge (2006)

Group Characteristics and Task Accuracy in Distributed Remote User Controlled Manufacturing as Collaborative Environment

Vesna Spasojevic Brkic[1(✉)], Goran Putnik[2], Zorica A. Veljkovic[1], Vaibhav Shah[2], and Ahmed Essdai[1]

[1] Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, 11120 Belgrade 35, Serbia
{vspasojevic, zveljkovic}@mas.bg.ac.rs, essdai64@hotmail.com

[2] Department of Production and Systems Engineering, University of Minho, Campus de Azurem, 4800-058 Guimaraes, Portugal
{putnikgdm, vaibhav.shah}@dps.uminho.pt

Abstract. Collaborative environments today are very successful when used in multiplayer games and for meetings organizing, which gives an idea to explore its potentials in remotely controlled decentralized manufacturing. This survey is based on the experiment that involved 34 small collaborative groups including 68 students in Serbia, that have used the interface for remote collaborative control of manufacturing systems to control of CNC machine located in Portugal. Our previous surveys have shown that group work consumes less time than individual work in most working options. This paper examines influence of homogeneity of the groups. Results of statistical examination show that percentages of errors are significantly or highly significantly higher in homogeneous compared to a non-homogeneous groups. Therefore, it could be recommended to form heterogeneous groups when remotely controlling decentralized manufacturing processes.

Keywords: “Wall” and “Window” interface · Group characteristics · Desktop and video beam presentation mode

1 Introduction

In order to remain competitive in a turbulent and very competitive climate product manufacturers today are forced to search for solutions in terms of new business models and strategies, management principles, organizational models, processes and technological capabilities to satisfy their customers. Smart, connected products offer many expanding opportunities for new functionality, enhanced reliability, much higher utilization and capabilities that change traditional boundaries [1]. Collaborative environments today are very successful when used in for multiplayer games and for meetings organizing, which gives an idea to explore its potential in remotely controlled decentralized manufacturing.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_4

This survey is based on the experiment that involved 34 small collaborative groups including 68 students at the University of Belgrade, Serbia, that have used the interface for remote collaborative control of manufacturing systems to control of CNC machine located at Universidade Minho, in Portugal. Our previous surveys [2, 3] have shown that group work consumes less time than individual work in most working options. The experiments on collaborative environment are done using two types of "client" user interface and two types of display, when working in small collaborative group consisted of two users. This research investigates samples of small collaborative groups regarding their homogeneity that is based on gender, age, average grade trough studies, computer literacy and desirability to use new software's, learning style and capability of knowledge transfer. With this aim, groups are divided in heterogeneous and homogeneous with main aim to examine task accuracy and draw conclusions from them.

1.1 Previous Research

The manufacturing environment has changed from traditional single site manufacturing to decentralized multi-site manufacturing networks in the last decades. There are also developed possibilities for remote control. The concept of collaboration is also emerging [4, 5]. The collaboration is derived from the Latin word "*collaborare*" that means "to work together," differs significantly from cooperation and describes a process of shared creation, where a group of entities enhance the capabilities of each other and that implies sharing of risks, resources, responsibilities, losses and rewards [7].

Accordingly, responsiveness to frequent market changes through share of resources in integrated manufacturing environments becomes efficient mode of operation that forces collaboration among different sites together with new business models, business strategies, governance principles, processes and technological capabilities [6, 7]. Small and medium size enterprises (SME's), characterized by limited skills and scarce resources, when using collaboration are able to overcome those limitations [7].

Collaborative manufacturing networks could be seen as mechanisms to facilitate agility and resilience and a way to mitigate the effects of unexpected changes and disruptions at market [8]. In that aim it is very important to combine dynamically and find the best fitting of set of competencies and resources [8]. Collaborative manufacturing networks and manufacturing grids are also emphasized by Liu and Shi [9], especially in sense of coordinated resource sharing and problem solving in dynamic, multi-institutional organizations. Large number of previous surveys pointed out to the importance of immersion and presence [10–12], while surveys in [10, 13] pay special attention to the fidelity of representation. But, soft dimensions which are critical to implementation of ubiquitous manufacturing till today are rarely surveyed, together with relation to culture and other constructs, both input and output [2, 14–16].

In that aim Spasojevic-Brkic et al. [2] have surveyed group and individual working options in distributed remote user controlled manufacturing system and proved that group collaborative work consumes less time than individual work in most working options. Authors in [2] have tested two versions of interface to remotely control the manufacturing cell and recommend "Wall" interface as collaborative environment for further usage and development, since it shows far better results than the "Window" interface.

According to previous research results it is obvious that there is a need to survey more deeply collaboration characteristics, such as group characteristics and task accuracy in distributed remote user controlled manufacturing as collaborative environment.

2 Method

The methodology includes examination of distributed remote user controlled manufacturing system that herein includes developed "client" user interface for the distributed manufacturing system, manufacturing cell, that belongs to the Ubiquitous Manufacturing System Demonstrator and is settled in Universidade Minho (as shown in Fig. 1), in Portugal and its control by 34 small collaborative groups including 68 students at the Faculty of Mechanical Engineering, University of Belgrade, Serbia. The experiments on collaborative environment are done using two types of "client" user interface, "Wall" and "Window" shown both in desktop and video beam presentation mode, when working in small collaborative group consisted of two users, as shown in Figs. 2 and 3. "Wall" interface shows the live video feed from the remote cell on wall panel and the human client" operator is watching the remote cell through on wall while controlling the remote CNC machine, while "Window" interface uses a window panel on the interface and the human "client" operator is watching the remote cell out into a window while controlling the remote CNC machine.

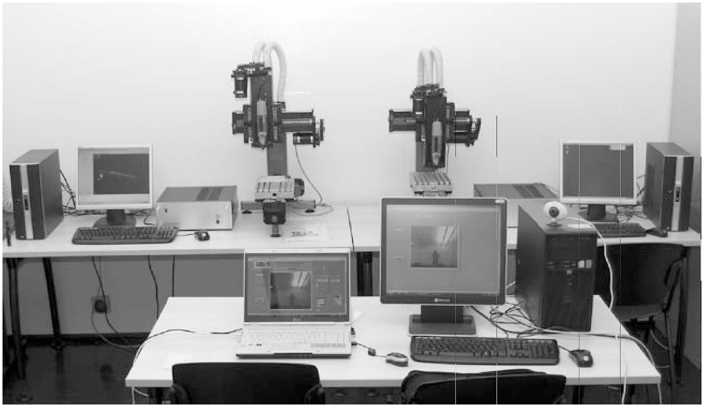


Fig. 1. Manufacturing cell at Universidade Minho, in Portugal

Each group had task to remotely control the manufacturing cell in different environment settings, namely using both "Wall" and "Window" interfaces in desktop and video beam presentation mode. The users' task consisted of the following subtasks: to connect to the remote cell, start a CNC machine, upload a CNC program to conduct

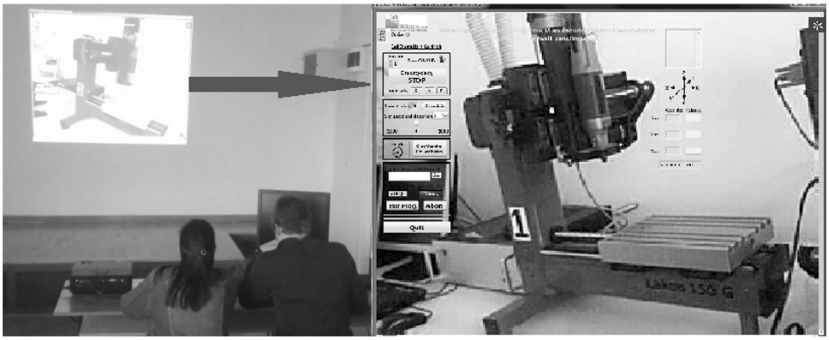


Fig. 2. Group performing task on “Wall” interface [2]

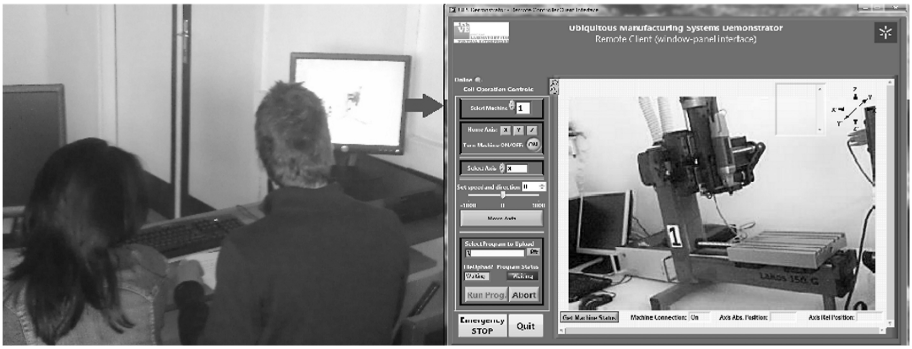


Fig. 3. Group performing task on “Window” interface [2]

operations on the machine, remotely use the emergency stop button, move axes, assess the status of the machine and positions of all axes.

3 Results

3.1 Applied Methodology

Recorded results are obtained by using following methods - descriptive statistics, correlation analysis, comparisons between groups and test for proportion. for basic characteristic defining homogeneity [17].

Descriptive statistics was used to obtain additional data for results in order to determine statistical procedures.

Tests within the groups were conducted for variables influencing homogeneity using correlations for linear regression or Spearman, depending of variable type.

For examination relations between the groups Student t-tests for parametric examinations, and U* Man Withney test for nonparametric examination were used.

In cases of binomial distribution of data z-test approximation was used.

3.2 Descriptive Statistics

Descriptive statistics is shown at Table 1. for both homogenous and heterogonous groups. It includes number of measurements (entities in groups), mean, median, standard deviation, coefficient of variation and Kolmogorov test for normality with it's *p*-values, in order to determine if the variable is parametric or nonparametric.

Table 1. Descriptive statistics on students' groups

Variable	*N*	*Av*	*Me*	*Sd*	*cv*(%)	*Kol.*	var.type
AGE HO	14	22.607	21.75	2.949	13.046	$p > 0.2$	par
GPA HO	14	8.214	8.00	0.378	4.601	$p < 0.01$	non par
CLT HO	14	3.929	4.00	0.730	18.583	$p > 0.2$	par
CKT HO	14	3.357	3.25	0.497	14.812	$p > 0.2$	par
LNS HO	14	4.071	4.00	0.646	15.872	$p > 0.2$	par
AGE HE	20	23.400	23.50	1.896	8.102	$p > 0.2$	par
GPA HE	20	8.300	8.00	0.715	8.609	$p > 0.2$	par
CLT HE	20	4.025	4.00	0.550	13.653	$p > 0.2$	par
CKT HE	20	3.750	3.75	0.500	13.333	$p > 0.2$	par
LNS HE	20	3.975	4.00	0.444	11.158	$p > 0.2$	par

Although there is certain low indication that the statistical significance of the number of homogeneous groups is greater than the number of heterogeneous group (*p*-value = 0.0867), with respect to the size of the sample it can be concluded that in this experiment the number of homogeneous and heterogeneous group was the same. Number of male participants in experiments was significantly higher than the number of female participants ($p < 0.001$) due to the fact that Faculty of Mechanical Engineering has higher percent of male students.

3.3 Homogeneity Variables Examination

Participants in collaborative groups had 23.16 years in average and grade point average 8.27 on the scale from 6 to 10. Average age in homogenous groups is 22.607 years, while in heterogeneous it counts 23.4 years. Students' computer literacy was 3.99, capability of knowledge transfer was 3.59 and desirability to learn new software was graded with 4 in average, all on the Likert scale from 1 to 5. Students' computer literacy in heterogeneous groups was 3.93, capability of knowledge transfer was 3.36 and desirability to learn new software was graded with 4.07 in average. Students' computer literacy in homogeneous groups was around 4, capability of knowledge transfer was 3.75 and desirability to learn new software was graded with 3.98 in average. Grade point average is slightly higher in heterogeneous groups (8.3 vs. 8.214).

Examination within the groups shows that for homogenous groups there are the highly significant correlation grade point average and the rest of variables defining homogeneity (Table 2). Reason for that can be the small group and application of nonparametric examination. Also correlation relation exists between capability of knowledge transfer and desirability for learning new software ($p < 0.05$).

There's no significant difference between male and female gender, while similar learning stile is lower in case of the same preferences comparing to different learning stale, for p-value at $p < 0.001$.

Table 2. Correlations between variables for homogenous groups

	Sp.corr.	p-value
GPA HO vs. *AGE HO*	0.0557	$p < 0.001$
GPA HO vs. *CKT HO*	–0.0114	$p < 0.001$
GPA HO vs. *CKT HO*	0.09358	$p < 0.001$
GPA HO vs. *LNS HO*	0.034	$p < 0.001$
	$r^2(\%)$	p-value
CKT HO vs. *LNS HO*	26.316	$p < 0.05$

In case of heterogonous groups where all variables are parametric (Table 1), correlations exists between desirability for learning new software's and computer literacy and, as well as capability of knowledge transfer for $p < 0.05$ (Table 3).

Table 3. Correlations between variables for heterogonous groups

	$r^2(\%)$	p-value
CLT HE vs. *LNS HE*	42.32	$p < 0.05$
CKT HE vs. *LNS HE*	38.81	$p < 0.05$

For heterogeneous number groups with same gender there is significantly higher than number of groups with mixed gender ($p < 0.001$). Also, groups with same learning style are significantly smaller than groups with different learning style, for $p < 0.01$.

Comparing homogenous and heterogonous groups indicates that capability of knowledge transfer is smaller for first groups then for the second at significance level $p < 0.05$. Furthermore, occurrence of the individual learning style is far greater in homogeneous groups than in heterogeneous, for significance level of $p < 0.001$ (Table 4).

Table 4. Comparison of significant characteristics between homogenous and heterogeneous groups

Homogenous group		Heterogonous group	Significance
CKT HO	$<$	*CKT HE*	$p < 0.05$
LST HO1	$>>>$	*LST HE1*	$p < 0.001$

3.4 Task Accuracy

The accuracy of executed task is measured by number of errors made during tasks' execution, namely probability to make an error. Comparison of errors depending on the homogeneity of the group is shown in Table 5.

Table 5. Comparison of errors depending on the homogeneity of the group

Homogenous group		Heterogonous group	Significance
NOE HO WA D	=	*NOE HE WA D*	n.s.
NOE HO WA V	>	*NOE HE WA V*	$p < 0.05$
NOE HO WI D	=	*NOE HE WI D*	n.s.
NOE HO WI V	>	*NOE HE WI V*	$p < 0.05$
NOE HO WA	>	*NOE HE WA*	$p < 0.05$
NOE HO WI	=	*NOE HE WI*	n.s., $p = 0.054$
NOE HO D	=	*NOE HE D*	n.s.
NOE HO V	>>>	*NOE HE V*	$p < 0.001$
NOE HO	>>	*NOE HE*	$p < 0.01$

On number of errors made during the execution task tests of proportions is applied in different environment settings. Accuracy of task execution regarding the interface, comparing the homogeneity of the groups is shown at Fig. 4, while overall adequacy of task execution for homogenous and heterogeneous groups is shown at Fig. 5.

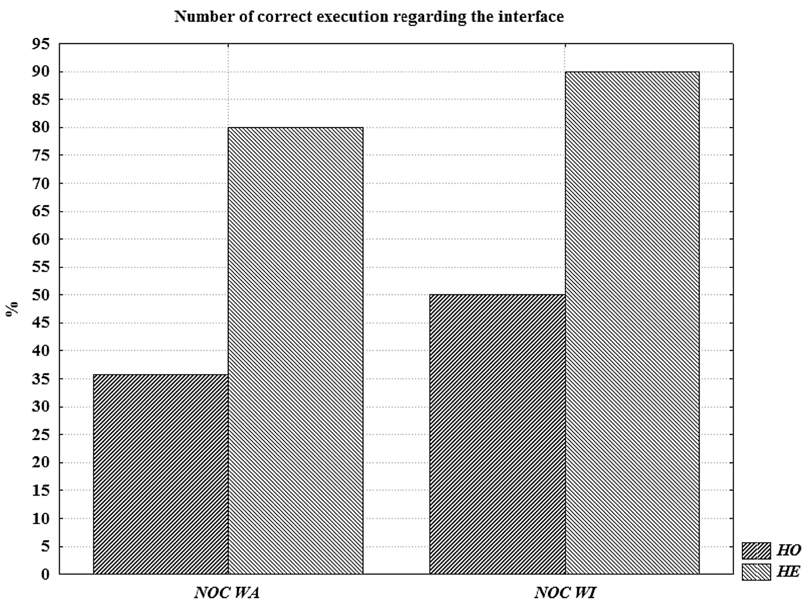


Fig. 4. Number of accurate task executions for homogeneous and heterogeneous groups regarding the type of interface

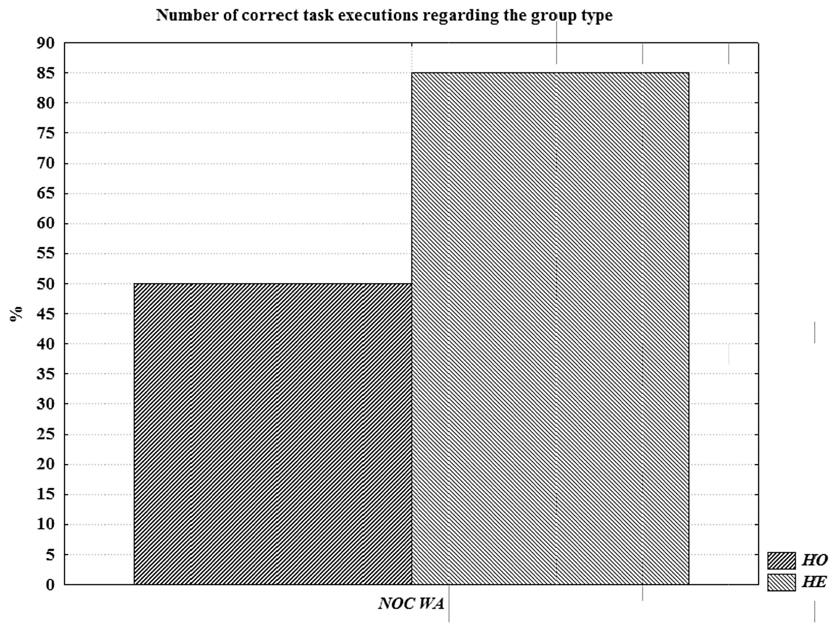


Fig. 5. Overall number of accurate task executions for homogeneous and heterogeneous groups

Results from Table 5 and Figs. 4 and 5 show the following:

1. The probability to make an error is significantly or highly significantly higher in homogeneous compared to non-homogeneous groups.
2. The percentage of errors when using “Window” interface does not show statistically significant difference, but is on the border of statistical significance of 0.05, so the results should be taken with a fence.
3. The number of errors in homogeneous group is greater than the number of errors in heterogeneous groups when working with video beams under the Wall interface.
4. The number of errors in homogeneous group is greater than the number of errors that is made in heterogeneous groups when working with video beam under the Window interface.
5. The number of errors that homogenous groups have made is much greater than the number of errors that heterogeneous groups have made when the experiment was carried out using video beam.
6. The number of errors that homogenous groups have made is much greater than the number of errors that heterogeneous groups have made.

4 Conclusion

Collaborative environments today are very successful tool for product manufacturers to remain competitive in a turbulent and very competitive climate. This survey uses an idea to explore its potential in remotely controlled decentralized manufacturing.

This survey is based on the experiment that involved 34 small collaborative groups, both heterogeneous and homogeneous by structure. Each group had task to remotely control the manufacturing cell using both “Wall” and “Window” interfaces and desktop and video beam interfaces. On number of errors made during the execution task tests of proportions is applied given environmental settings.

Statistically significant results show that percentages of errors are significantly or highly significantly higher in homogeneous compared to a non-homogeneous groups. The percentage of errors when using “Window” interface does not show statistically significant difference, but is on the border of statistical significance of 0.05.

Engineering conclusions show that: 1. The number of errors for homogeneous group is greater than the number of errors in heterogeneous groups when working with video beam under “Wall” interface, 2. The number of errors that homogenous groups have made is greater than the number of errors that heterogeneous groups have made when working under “Wall” interface in general, 3. The number of errors that homogenous groups have made is greater than the number of errors that heterogeneous groups have made when using video beam interface and 4. The number of errors that homogenous groups have made is greater than the number of errors that heterogeneous groups have made, in general.

It could be recommended to form heterogeneous groups when remotely controlling decentralized manufacturing processes.

Since our previous research [3, 10] give advantage to “Wall” interface in sense of collaborative effort, involvement, participants’ activity, satisfaction, awareness, representational fidelity and co-presence measures it could be added that according to results of this survey video beam mode of presentation could be recommended, too.

Future research ideas are to test similar platforms with more than two users and to study aspects such as group leader choice and role, possible conflicts etc.

Acknowledgements. This work is supported by grants for bilateral cooperation project “Design and Evaluation of User Interfaces for Remote Collaborative Control of Manufacturing Systems” between Serbia and Portugal, financed by Serbian Ministry of Education, Science and Technological Development (MESTD) and Portuguese Foundation for Science and Technology - Fundação para a Ciência e a Tecnologia (FCT), respectively and grant SPRINCE - SAF€RA by MESTD.

Appendix: Nomenclature

Nomenclature	
AGE	Average age per group
GPA	Grade point average per group with scale 6 to 10
CLT	Average computer literacy per group
CKT	Average capability of knowledge transfer per group
LNS	Average desirability for learning new software per group
CKT	Average capability for knowledge transfer per group
LST	Learning style - individual or in group
GEN	Gender - male of female
NOE	Number or percentage of errors in task execution
NOC	Number or percentage of correct task executions
HO	Homogenous group
HO1	Homogenous group for individual learning style
HE	Heterogeneous group
HE1	Heterogeneous group for individual learning style
WA	"Wall" interface
WI	"Window" interface
D	Desktop screen
V	Video beam presentation
N	Number of participants in the group
Av	Mean value of the variable in the group
Me	Median value of the variable in the group
Sd	Standard deviation of the variable in the group
cv (%)	Coefficient of variation for the group
Kol	Kolmogorov normality test
p	*p* value(level) of the test
n.s.	Statistically non-significant result
$r^2(\%)$	Coefficient of determination
Sp. corr.	Spearman correlation

References

1. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming competition. Harvard Bus. Rev. **92**(11), 64–88 (2014)
2. Brkic, V.S., Putnik, G., Veljković, Z., Shah, V., Essdai, A., Castro, H.: Interfaces for distributed remote user controlled manufacturing: working individually or in collaborative group? Proc. Manuf. **3**, 747–753 (2015)
3. Brkic Spasojevic, V., Putnik, G., Veljković, Z. A., Shah, V., Essdai, A.: Interfaces for distributed remote user controlled manufacturing as collaborative environment. In: Nunes, I.L. (ed.) AFHE 2016. LNCS, vol. 497, pp. 335–347. Springer, Heidelberg (2017)

4. Ding, K., Jiang, P., Leng, J., Cao, W.: Modeling and analyzing of an enterprise relationship network in the context of social manufacturing. Proc. Inst. Mech. Eng. B J. Eng. Manuf. **230**(4), 752–769 (2016)
5. Wang, W.M., Lünnemann, P., Neumeyer, S., Hayka, H., Stark, R.: Product development in collaborative networks – an expert view on current challenges and future trends. In: Afsarmanesh, H., Camarinha-Matos, L.M., Lucas Soares, A. (eds.) PRO-VE 2016. LNCS, vol. 480, pp. 302–312. Springer, Heidelberg (2016)
6. Filip, F.G., Zamfirescu, C.B., Ciurea, C.: Collaborative activities and methods. Comput. Support. Collaborative Decis. Making. **4**, 71–120 (2016)
7. Camarinha-Matos, L.M., Afsarmanesh, H., Galeano, N., Molina, A.: Collaborative networked organizations—concepts and practice in manufacturing enterprises. Comput. Ind. Eng. **57**(1), 46–60 (2009)
8. Camarinha-Matos, L.M.: Collaborative networks in industry and the role of PRO-VE. Int. J. Prod. Manag. Eng. **2**(2), 53–56 (2014)
9. Liu, Q., Shi, Y.J.: Grid manufacturing: a new solution for cross-enterprise collaboration. Int. J. Adv. Manuf. Technol. **36**(1–2), 205–212 (2008)
10. Spasojević-Brkić, V.K., Putnik, G., Veljković, Z.A., Shah, V., Castro, H.: Representational fidelity in distributed and remote lab environment. FME Trans. **42**(3), 243–248 (2014)
11. Mikropoulos, T.A.: Presence: a unique characteristic in educational virtual environments. Virtual Real. **10**(3–4), 197–206 (2006)
12. Mikropoulos, T.A., Strouboulis, V.: Factors that influence presence in educational virtual environments. Cyber Psychol. Behav. **7**(5), 582–591 (2014)
13. Dalgarno, B., Lee, M.J.: What are the learning affordances of 3-D virtual environments? Br. J. Educ. Technol. **41**(1), 10–32 (2010)
14. Dubey, R., Gunasekran, A., Chakrabarty, A.: Ubiquitous manufacturing: overview, framework and further research. Int. J. Comput. Integr. Manuf. **30**, 1–14 (2015)
15. Spasojevic-Brkic, V.K., Putnik, G.: User evaluation of interfaces for remote control manufacturing systems. Serb. J. Manag. **8**(2), 201–212 (2013)
16. Mejía-Gutiérrez, R., Osorio-Gómez, G., Ríos-Zapata, D., Zuluaga-Holguín, D.: Ubiquitous conceptual design of a ubiquitous application: A textile SME case study for real time manufacturing monitoring. Comput. Aided Des. **59**, 214–228 (2015)
17. Montgomery, D.C., Runger, G.C.: Probability and Statistics for Engineers, 3rd edn. Wiley, New York (2003)

The Use of Context Aware Pervasive Systems in the Area of Human Factors and Ergonomics

Beza Mamo(✉)

Department of Computer Science, Kotebe Metropolitan University,
Addis Ababa, Ethiopia
bezamamo@yahoo.com

Abstract. The major attempt of pervasive computing is to reduce the required user effort in using applications in their surrounding by identifying and recognizing resources, and reacting involuntarily based on the user concern. Pervasive systems are also focusing on improving efficiency, productivity and job satisfaction, with the goal of minimizing errors which are the core issue in human factors too. Spontaneous interaction of heterogeneous resources to achieve context awareness requires the identification of entities together with its description in a given environment is also the concern of Human Factors and Ergonomics as it deals with human – system interface technology in broad-spectrum. Fundamentally, context acquisition is one of the challenging and essential elements in pervasive computing; this concept is also elaborated from the aspect cognitive ergonomics standpoint. This work proposes algorithms used for context preference identification and discusses how context aggregation takes place by a range of context provider.

Keywords: Context awareness · Human factors · Ergonomics · Context provider · Low level context

1 Introduction

Today most computing components are context sensitive, inconspicuous, almost unseen and interconnected through wireless networks. This was the vision of Mark Weiser in early 1990s as he forecasted computers are applied everywhere and anywhere in the 21st century [1]. Pervasive-computing environment are crowded, heterogeneous and always changing. Additionally pervasive-computing applications must discover and use resources based on the current context. The discovery and use of resources are one of the prevailing concepts in the discipline of Ergonomics.

Ergonomics in general, deals with technical and work circumstance of a particular entity. It put human needs and capabilities at the focus of designing technological systems. Software ergonomics in particular, is about software design of systems, that encompass user needs, interface design, user support and usability test. A systematic approach is essential in the design of software process by considering Human Factors and, thus, usability of the software can be addressed adequately in succession.

One of the important design element in pervasive computing is how applications obtain information they require in order to implement adaptive behavior [2]. Predicting

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_5

events is one of the challenging phenomenons in pervasive environment. There are a number of research works that deal with event prediction based on context histories and current action of a user, devices and other resources in the user environment.

As learnt from the work of [3–5], context information can be obtained from a number of sources including hardware sensors, network information, device status, user profiles and other entities in the user environment. Consequently, context information can be divided in to physical context (i.e., location, sound, movement, and touch) and logical context (user goal, tasks, business process and emotional states). Correspondingly our previous generic architecture [6] elaborate context information as dynamic context (same as physical context) and static context (subsets of logical context).

2 Related Works

2.1 Context-Awareness

Guanling Chen and Dauid Kitz [7] define context as: "Circumstance in which an application runs, and may include: physical state, computational state, and user state".

The following lists are identified as feature of context-aware applications [8, 9]:

- presentation of information and services to the user;
- automatic execution of services to the user; and
- tagging of context information to support later retrieval

Two recent trends highly facilitate the vision of ubiquitous or pervasive computing: wireless hotspots and the explosion of mobile devices. Network technology facilitates, the spontaneous connection of a user personal device to immediate services seamlessly [10, 11]. But the difficulty lies on knowing the available services in the user environment.

The dynamic environment of pervasive computing proactively provides the participating entities with a rich set of capabilities and services all the time, everywhere, and in a transparent, integrated and extensible way. A context is any situation of an entity (i.e. a person, place, location, etc.).

Obtaining the exact context of an entity is not a simple task and requires in-depth consideration of stored profile of user data in relation to user interaction with other entities/resources as this study indicates.

2.2 Smart-Office

The setup of modern day workplaces relies on the use of information technology. Optimal and effective utilization of physical infrastructure and IT resource is the core concern in smart office. Devices such as, display, a mouse, a keyboard, or other peripheral appliances are considered as infrastructure in smart office environment that contains embedded processor with wireless capability. Integration and interaction of such resources in a smart office environment should be transparent to its user.

Moreover, [12] identifies four basic challenges and put forward resolution techniques to establish smart spaces in a workplace environment; (i) power dependencies,

(ii) network dependencies, (iii) peripheral dependencies and (iv) application dependencies.

To the best of our knowledge most of smart office appliance strive to address the above stated challenges reasonably and tries to accommodate smartness from space/allocation point of view.

2.3 Human Factors

Software development is primarily a human and collaborative endeavor, as software is typically developed by individuals or groups working together in an organization. The Study of Human Factors is essential for every software manager to understand how development team interacts with each other.

Numerous companies came to recognize success of a product depends upon a solid Human Factors design and Human Computer Interface (HCI) design, which "is a sub-discipline concerned with the specification, design, evaluation/testing and implementation of interactive computing systems for human use" [13]. As a result, HCI evolve into a discipline, which has its own defined and managed processes.

Ergonomics is a branch of Human computer interaction that can be defined as the application of scientific information concerning humans to the design of objects, systems and environment for human use. It deals with the technological and work situations of a particular individual. It puts human needs and capabilities at the focus of designing technological systems. Office ergonomics, in particular, includes workspace design, environmental factors and teamwork [14].

3 Challenges

Addressing human factors in system development requires the investigation of human ability, limitation and associated uniqueness in designing task, job or system. In this regard, pervasive computing creates a better opportunity with respect to developing context sensitive application/services, which react based on the changing behavior of its user. On the other hand, such context sensitive application requires in-depth understanding of user, task, system and resources preferences. Analyzing preferences of a given entity is still challenging in context aware pervasive system, and hence demand end-to-end understanding of entities, components, methods, behaviors and abstraction for a given circumstance.

Identification of preferences in relation to: user, furniture, and changing environment are difficult. The three basic preferences (user preferences, furniture preference, and total environment preference) provide set of low level context information and demand aggregation to support context awareness with the goal of human factors and ergonomics at time of preprocessing.

4 The Proposed Approach

This section contains part of the extension of our work [6], with some change at the lower level components to accommodate smart office setup and detailed elaboration of introduced components for the overall context determination. The proposed architectural diagram (Fig. 1), illustrates how pervasive system consume context preferences systematically.

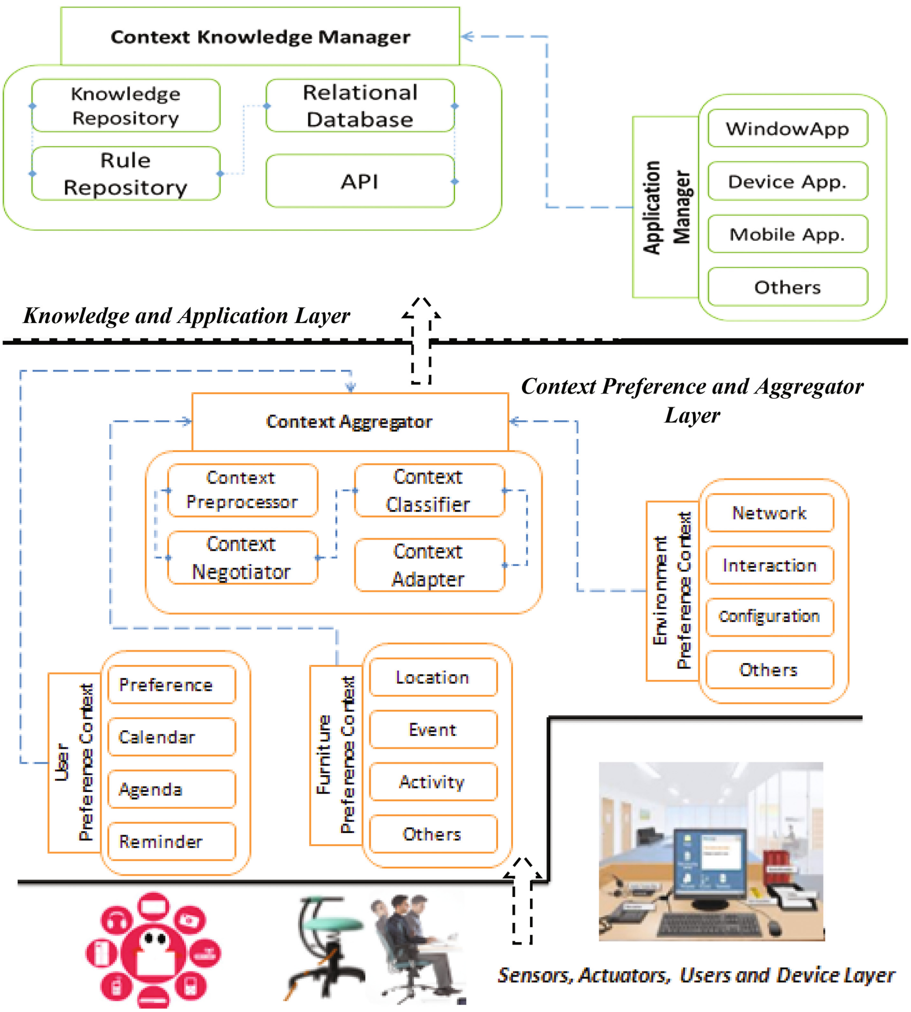


Fig. 1. Proposed context aware system architecture tuned for human factors and ergonomics application.

4.1 Component Description

This section elaborates the components identified in Fig. 1, by giving emphasis to lower level components.

User Preference Context. This is static context information to be filled by the smart office personnel as a preference to all intelligent appliances in the office environment. For instance, a person could have a preference of seat while composing mail with respect to type of device (such as PC, Laptop, PDA, Smart phone and other personal apparatus). Thus, the user is supposed to define his/her preference while using the suitable pervasive application, preferably via hand held device. The envisioned pervasive application encourages office personals to fill their preference as soon as using the intelligent office furniture and appliances for the first time and utilize it for subsequent usage. Moreover, user data available in the hand healed devices are considered as a major input for the overall context predication and regarded as source of human factors in using context sensitive applications/devices.

User preference is not limited to the usage of explicit intelligent furniture for some specific purpose, but also a preference of doing things using a number of technological solutions, in general.

Furniture Preference Context. Intelligent office furniture has always a defined preference up on the different context in the smart office environment. The furniture's are assumed to have a location preference in relation to other entities in smart office specifically at time of usage. For instance the orientation of white board in a given room should be visible to all chairs (positions) in the area, as long as chairs are occupied by participants (asserted by pressure sensor available at chairs).

Thus, activities in smart office environment play a significant factor for the overall setup of intelligent office furniture. Employees in smart office might involve in a number of activity while using a limited intelligent resource (such as chair). For instance, a person while seating in a chair, he/she might involve in different activities such as reading, telephone conversation, writing memo or discussion with client or colleagues; for each of the different activities the setup of the intelligent chair is different so as to facilitate better level of task accomplishment. Office furniture for two office personnel should react according to their preference and the total environment context.

Environment Preference Context. All entities in a given smart office environment would vote for their preference as per the configuration, interaction, network and other contextual data (such as date and time). For instance, considering the intelligent white-board at time of demonstration, it is governed by the absence or presence of light in the room. Consequently, the light sensors determine the situation of the intelligent whiteboard as a global preference.

In general, resources in smart office environment would have two type of preference; local preference and global preference. The local preference is the one set as a default preference for specific intelligent furniture. On the hand, the global preference is the one considered as environmental preference. The global preference prevails over the local preference, in case of preference variance.

Context Aggregator. Detail of context aggregator and other high level components are already described in our previous publication [6] and accessible directly online at www.scincedirect.com.

4.2 Proposed Algorithms and Techniques

This section reveals algorithms consumed by the different context preferences (such as user, furniture and environment). Consequently, the algorithm detailed in Fig. 2, is permitted for all context prediction at lower level. The context achievement is analyzed by the context lookup data given in Table 1.

Table 1. Context lookup table

Context variables			Context value and action		
CV1	CV2	CV3	Context response value	Number of variables return 1	Context support
Curent Contexte (CV1, CV2, CV3)					
0	0	0	0	None (0)	No
0	0	1	1	1	Limited
0	1	0	2	1	
1	0	0	4	1	
0	1	1	3	2	Intermediate
1	0	1	5	2	
1	0	1	6	2	
1	1	1	7	All (3)	High

Remarks: **For User Context:** CV1 = Agenda, CV2 = Reminder, & CV3 = TodoList; **For Furniture Context:** CV1 = Event, CV2 = Location, & CV3 = Activity; **For Environment Context:** CV1 = Network, CV2 = Integration, CV3 = Configuration

The algorithm in Fig. 2, will be consumed by the algorithm detailed in Fig. 3. The preprocess_context method at line 5 of Fig. 3 relies on the aggregate context response recommended by Table 1.

As learned from Fig. 4, and the context lookup table there are four set of context support condition governed mainly based upon the context variables responses. The current context of user, furniture or environment return the Boolean value one to assure the situation outlook availability for a given context variable is true, otherwise it will be zero, to indicate the lack of context knowledge.

Consequently, no context support is a circumstance where all of the context variable lacks knowledge about the user, environment, and other contextual situation and hence requires adaption of the smart environment to the user handheld device, intelligent furniture and other nearby devices in the environment. On the contrary, high context support is facilitated, if all the context variables have knowledge about the user, the environment and connected appliances entirely, in such cases the device will react

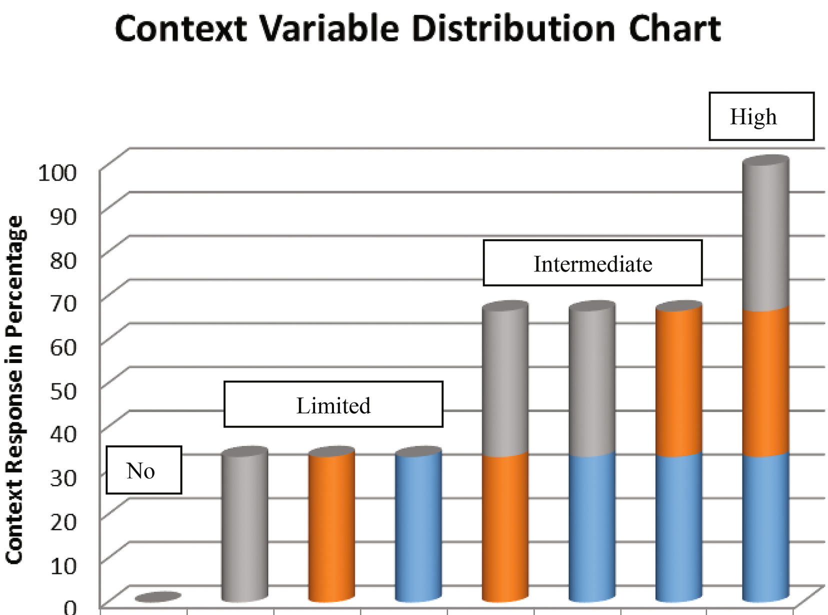
```
Input:
   SP (Static Profile, accessible from the database)
   ci (Set of context for user, furniture, and environment)
      d (Current date)
      st (Activity start time)
      et (Activity end time)                         //will be append as the context changes
Output: CCi (Current Context)                        //aggregate information of
Processes:
      d ← current_date                               // initialization of system date information
      ci ←SetContext(C)                              // initialization context
      ip ← i_profile                                 // initialization of user, chair or environment profile
      st ← Start_Time                                // initialization of current System time
For every Context do
      If Join ((iP(d,st), ci) >=1)                   // if any intersection
               PrioritizeContext(ci)                 //refer figure 3, below
      Else
               SP(Default)                           //No Context binding
  End if
      et ← End_Time                                  //Activity accomplishment or Context change time
      return CCi                          /*where i=u, for user context, i=f, for furniture context,
                                          and i=e, for environment context */

End do
```

Fig. 2. Algorithm used for context preference determination

```
Input:
      CCu (set of user Context)                      //return of fig 2
      CCf (set of Furniture context)                 //return of fig 2
      CCe (Set of Environment Context)               //return of fig 42
Output:
      AggregatContext
Process:
      C= Preprocess_Context (CLoockup(CCu), CLoockup(CCf), CLoockup(CCe))
  Begin :
      Context_Negotator(
      If Context Chage(C) = True
               AggregatContext = Create(Context_Classifier(C))
      else
         Search_for_Context(C, ContextDB)
            AggregatContext = Consume (Context_Classifier(C))
         End If
   End
```

Fig. 3. Algorithm used for context aggregation

involuntarily based on the predefined context knowledge. The remaining, situation lack the assertion of one or two variable set and facilitates limited or intermediate context support respectively, such situation requires context refinement to adjust the variables with the value equals to zero.



CV3	0	33.3	0	0	33.3	33.3	0	33.3
CV2	0	0	33.3	0	33.3	0	33.3	33.3
CV1	0	0	0	33.3	0	33.3	33.3	33.3

Fig. 4. Context variable distribution chart

As per the exploration in this investigation, the context response of a given context preference is highly influenced by the context variables and computationally determined using a simple mathematical approach. Thus, $CR = 2^{CV}$, Where CR - is the context response and CV - is the context variable. The variable values are binary (i.e., 0 or 1) in nature as clearly indicated in Table 1.

4.3 Implementation Scenario

- User 1 has two hours meeting agenda schedule before a week at 4:00 AM in the meeting room.
- User 1 had a discussion with his/her colleague before the meeting section starts, the position of the chair was adjusted as per his/her furniture default preference, to be loaded/adjusted to the smart chair from the user hand held device application.

- The context of the chair will be adjusted as the meeting starts, by considering the status (on/off) and position (vertical/horizontal and orientation) of the meeting room projector, the configuration and related preference of the projector will be broadcast to all intelligent chair and chair occupants.
- The context aware server side application (referring our previous study [6]) will consume the published context preferences (user, furniture, and environment), the moment the projector is on.
- Once the diverse context variable preferences aggregate based on the algorithms and the context lookup table responses, the context of the intelligent furniture will be adjusted/corrected based on temporary preferences.
  – If all context variable preference has knowledge about the circumstance, then context variation is very minimal and hence the intelligent chair will behave based on the predefined preference.
  – If context variable preference lacks knowledge about the cir circumstance, then context variation is very significant and hence the intelligent chair requires training/preference setup by its user.
- Finally, the high level preference will be directly push to the intelligent furniture actuator to take action based on User 1 preferences.
- Eventually, at the end of the meeting user 1 will have a telephone conversation; the intelligent furniture will adjust its position based on the default preference of chair (i.e., local preference).
- The moment User 1 out f the meeting room, the context preference will be stored in the server application, for successive usage for all chairs available in the meeting room by associating with User 1 hand held device.

5 Conclusion and Future Work

This article brings concepts that have common characteristics; critical and requires a separate detail analysis of context preference, primarily for context aware pervasive system. Three basic preferences (user, furniture, and environment) were considered for the actualization of low level context information and stipulate aggregation to support context awareness with the goal of human factors and ergonomics at time of preprocessing. Knowing the preference of entities in context aware environment can be attained using the suggested techniques, and algorithms have enormous input for human factors in software system design and ergonomics. The preferences aggregated will be pushed to the knowledge and application layer. The application layer, consequently push the refined context information to the desirable appliance to bring context awareness and by considering human factors in the changing user environment.

This work gives emphasis only to low level context prediction, algorithms, and related techniques. Further, investigation with regard to the high level context information and actual deployment of the proposed techniques are considered as part of the prospect work.

References

1. Weiser, M.: The Computer for the 21st Century. Scientific American, September 1991
2. Garlan, D., Siewirek, D.P., Smailogic, A., Steenkiste, P.: Project aura: towards distraction-free pervasive computing. IEEE Pervasive Comput. **1**(2), 22–31 (2002)
3. Mobile Sensing Towards Context Awareness, September 2017. https://web.fe.up.pt/~dee10014/documents/individual_topics.pdf. Rua Dr. Roberto Frias
4. Baldauf, M., Dustdar, S., Rosenberg, F.: A survey on context-aware systems. Int. J. Ad Hoc Ubiquitous Comput. **2**(4), 263277 (2007)
5. Hofer, T., Pichler, M., Leonhartsberger, G., Altmann, J., Retschitzegger, W.: Context-awareness on mobile devices the hydrogen approaches. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences, pp. 292–302 (2002)
6. Mamo, B., Ejigu, D.: A Generic layered architecture for context aware applications. In: The 9th International Conference on Future Networks and Communications (FNC 2014)/The 11th International Conference on Mobile Systems and Pervasive Computing, MobiSPC 2014, ScienceDirect ELSERIER, vol. 34, pp. 619–624, August 2014
7. Guanling, C., David, K.: Context-sensitive resource discovery. In: Proceeding of the First International Conference on Pervasive Computing and communications, pp. 243–252, March 2003
8. Day, A.: Understanding and using context. Pers. Ubiquitous Comput. **5**(1), 4–7 (2001)
9. Dey, A., Abowd, G.: Towards a better understanding of context-awareness. In: First International Symposium on Handheld and Ubiquitous Computing, HCC 1999, June 1999
10. Bottaro, A., Gerodolle, A., Lalanda, P.: Service composition in the home network. In: Advanced Information Networking and Applications, pp. 596–603, May 2007
11. Guinard, D., Steng, S., Gellersen, H.: Related Gateways: A User Interface for Spontaneous Mobile Interaction with Pervasive Services. CHI, San Jose (2007)
12. Akyol, B.A., Fredette, M., Jackson, A.W., Krishnan, R., Mankins, D., Partridge, C., Shectman, N., Troxel, G.D.: Smart office spaces. In: Proceedings of the Embedded Systems Workshop, Cambridge, MA, USA, 29–31 March 1999
13. Seffah, A.: Human Computer Interaction—A Brief Introduction: Past, Present and Future
14. Ergonomics. http://www.ergonomics.org.uk. Accessed Mar 2017

Learning the "Language" of Road Users - How Shall a Self-driving Car Convey Its Intention to Cooperate to Other Human Drivers?

Nina Kauffmann[1(✉)], Frederik Naujoks[2], Franz Winkler[1], and Wilfried Kunde[3]

[1] BMW Group, Knorrstraße 147, 80788 Munich, Germany
{Nina.Kauffmann,Franz.Winkler}@bmw.de
[2] Wuerzburg Institute for Traffic Sciences (WIVW), Robert-Bosch-Straße 4, 97209 Veitshöchheim, Germany
naujoks@wivw.de
[3] Department of Psychology (III), Röntgenring 11, 97070 Würzburg, Germany
kunde@psychologie.uni-wuerzburg.de

Abstract. Communication between road users is ruled by road traffic regulations, but there are also implicit laws of communication. Especially lane changes in dense traffic scenarios require not only communicating one's intention but also cooperating with other drivers. Self-driving vehicles will need to communicate with conventional vehicles on the road during the transition period to full automation. But how does a driver show his willingness to cooperate? A driving simulator study with N = 28 drivers in a dense traffic scenario on the highway was conducted. It was assumed that different lag vehicle reaction behavior on turn signals of the ego driver would influence the ego driver in his subjective evaluation of the situation. Three main effects, deceleration, the amount of velocity reduction and reaction time concerning perceived cooperation were found. The results of the study can be used to design cooperative driving strategies between self-driving and manually driven vehicles.

Keywords: Cooperative driving · Communication · Driving behavior · Interaction between Self-driving and conventional vehicles

1 Introduction

During the past decade, autonomous cars attracted the focus of researchers' attention and interest. It is known that self-driving cars offer many advantages such as enhancing road safety, comfort and reducing emissions [1]. Studies regarding the public opinion towards autonomous driving indicate that autonomous cars express a positive image and the expectations towards their abilities are high [2]. Nevertheless, regarding traffic safety, there are still concerns raised by customers [2].

Considering the period of the transition between assisted and fully automated driving, one essential question remains unsolved: How should self-driving vehicles interact with drivers of conventional vehicles on the road to be accepted? Due to the actual state of development of advanced driver assistance systems, it is impossible to

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_6

avoid a mixture of vehicles with different levels of automation sharing the same road, as stated by Sivak and Schoettle [2]. Hence, gaining acceptance from end users that are still driving conventional vehicles will be crucial to the deployment of self-driving vehicles. Experiencing a positive image of self-driving cars as an initial impression could lead to positive emotions and subsequently qualifying for a long-term acceptance of autonomous vehicles [3, 4].

However, available studies on which behaviors of self-driving vehicles are perceived to be acceptable are limited. Numerous studies try to replicate human behavior by analyzing objective drive parameters, but fundamental research concerning the appearance of the self-driving car is only reported by [5–7]. Färber [8] emphasizes the importance that the intention of the car must be understood to gain acceptance by other road users. Consequently, it is necessary to observe the interaction of drivers with self-driving cars to build a broader empirical basis for the derivation of recommendations relating to actions for self-driving cars.

Looking at the acceptability of different behaviours during interaction sequences between conventional vehicles may serve as a starting point for the development of acceptable driving strategies of self-driving cars. Rules of the road not only consist of formal laws of traffic regulations but also of informal ways understood clearly between drivers [9]. Merten pointed out different strategies how to cope with other drivers in traffic situations [9]. According to the author, drivers tend to perform anticipatory indications of their intended driving action they are planning to execute. For example, when drivers aim to change lanes, a slight adjustment in the lateral position can be noticed [10]. Moreover, there exist certain gestures and mimics illustrated by Risser [11] which help the driver to communicate his intentions. Furthermore, interactions between road users comprise an emotional component leading to an interpretation of actions in a specific way, causing emotional reactions such as anger or frustration [12]. Road users have a certain imagination of which behaviour is perceived positively or negatively in traffic [11]. An example for an aggressive behavior is driving too slow or tailgating other cars initiating negative feelings [13]. Maag [14] characterizes traffic behavior as being either cooperative or aggressive, describing a behavior in the sense of giving a competitive edge at the expense of others. Benmimoun [15] defines cooperative behavior as a general term for fair and professional behavior. The advantages of showing the willingness to cooperate can be seen in a positive perception of the driver but also in the prevention of reactance [13] meaning the other driver has all options to decide what to do. On a macroscopic level, according to Zimmermann et al. [16], cooperative driving strategies can optimize security, comfort and traffic flow.

In the present study, the perception of the braking behavior of the lag driver during a lane change is investigated. The aim of this study is to obtain recommendations how a self-driving car should behave to be accepted by surrounding traffic when signalling the willingness to cooperate. Additionally, the lane change behavior of the merging car is analyzed. The present study consists of two parts: the pre-study and main study. In the pre-study, participants were interrogated verbally concerning their lane change behavior from the perspective of the merging car. It was analyzed which parameter has the greatest influence on the decision to change the lanes. In the second part of the experiment, the main intention was to find out which parameters human drivers apply to judge whether a lane change is possible and how the lag driver is perceived. In the following section, an introduction to cooperative driving is given. Subsequently, the two parts of the study are described and the results are discussed. Finally, a summary and an outlook, as well as implications for further studies, are presented.

2 Method

2.1 Independent Variables

The following sections will introduce the independent variables used in the study in detail, as well as the reasoning why these variables were chosen.

When a driver reacts to a stimulus applying a braking behavior (e.g., when reaction to a cutting-in vehicle), the reaction can be decomposed into a sequence of parameters. Since not all of them are equally effortless to be implemented, it was intended to find out which parameters are used by observers to judge whether another vehicle will cooperate or not during lane changes. In a preliminary study, N = 35 participants (16 females) between the ages of 20 to 68 years (M = 40, SD = 15.9) were asked, about relevant parameters in their decision to change lanes. The questions asked were:

(a) Who is most relevant to your judgment whether to change lanes on a two lane road?
(b) What indicators did you use to decide, if the lag driver on the target lane is willing to cooperate with you?

Most of the participants referred to the lag driver (n = 25) as the most important indicator concerning their judgment. Additionally, the lag and lead driver were mentioned (n = 10). These findings are supported by [17, 18]. The results of the second question are shown in Fig. 1. They show that the indicators for cooperation are mainly be perceived by means of parameters resulting from driving behavior of the other vehicle, especially the amount of speed reduction, deceleration and the reaction time of the other vehicle. Consequently, these factors were used as independent variables in the simulator experiment. The exact factor levels used in the study are explained in the following paragraphs.

Reaction Time. We define the reaction time as a combination of mental processing time and the beginning of the movement time. The reaction time is influenced by different factors such as expectancy, age, gender, the type of the wheel and urgency [5, 19, 20]. When an event is expected, reaction times vary from 0.6s to 1.4s [19] and 0.8s to 2.4s when events are unexpected [21]. In field studies, reaction times to turn signals in dense traffic situations vary between 1 and 2 s as can be found in [5, 22]. Based on the available literature, reaction times of 0, 1 and 2 s are chosen, whereby a reaction time of 0 s is considered as the anchor point of an anticipating driver.

Deceleration. Concerning braking behavior in dense traffic situations Ehmanns (2000) conducted a field study and found decelerations of between -0.5 to -1.5 m/s^2 [23]. Ahmed [5] reported similar findings for the same setting between -0.6 m/s^2 and -1.6 m/s^2 depending on the speed difference, THW (time head way) and traffic densities. Kesting [24] suggested deceleration values up to -4 m/s^2 for his simulation

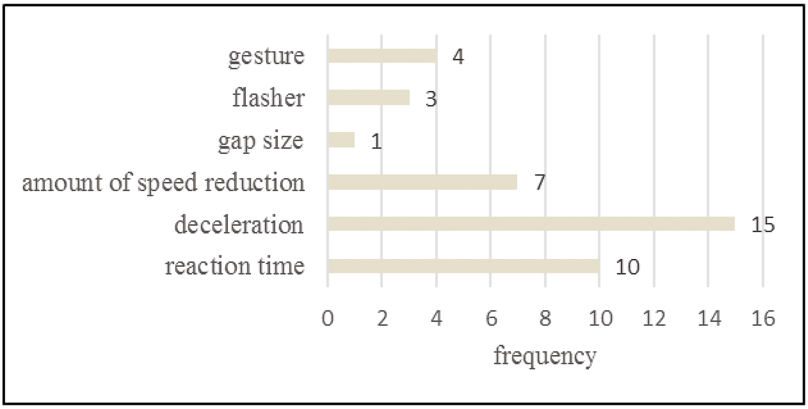


Fig. 1. Indicators for perceived cooperation.

model MOBIL. In the current study values of -0.5 m/s^2, -1.0 m/s^2, -1.5 m/s^2 were used.

Amount of Speed Reduction. We define the amount of speed reduction as the difference between the velocity before the lane change announcement was observed and right after the deceleration process finished. Values can scarcely be found in the literature due to vague definitions [19, 25]. Barreit et al. [26] examined the period from the initial contact of the foot on the brake pedal until the brake pedal was fully pressed. Values between 0.5 and 1 s could be found, where driver never pressed the brake pedal completely down. Based on a pre-study which was conducted in the same simulator as in the main study we decided to use values of 7 m/s, 4 m/s, 2 m/s. Based on the original speed these values result in a reduction of speed of 80% (7 m/s of 8.33 m/s), 48% (4 m/s of 8.33 m/s), and 24% (2 m/s of 8.33 m/s).

2.2 Apparatus

The experiment was conducted in a fixed-based driving simulator at the Wuerzburg Institute for Traffic Science. The scenery is projected on five canvases (front, left, right, left behind, right behind) to simulate a driver's 300° field of vision. The simulator has an automatic gearbox, and the driver sits in a driver's seat with a steering wheel with force feedback, acceleration pedal, and brake pedal (Fig. 2).

2.3 Experimental Design

The test track was a three-lane highway with congested traffic conditions. Convoys of vehicles were driving in the middle lane with a THW (time head way) of 2 s between them. The headway between the vehicles was chosen as 2 s because literature defines this measure as critical gaps [17], but they are still in the normal driving zone of the lag driver on the target lane [27, 28]. The experiment was divided in 27 scenarios,

Fig. 2. The fixed based driving simulator of the WIVW.

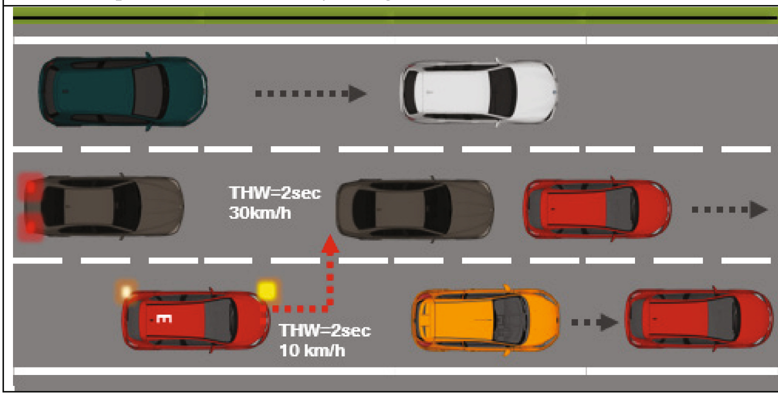
Scenario schematic
Description of the test situation
After a sound signal, the participant (E) had to decide whether a lane change is possible. Afterward, a subjective evaluation of the situation concerning perceived cooperation and criticality was given.


Fig. 3. Subjective, dependent measures employed in a simulator study.

consisting of convoys of cars, which reacted with varied braking behavior when a turn signal appeared (see Fig. 3). The traffic situations were presented in the randomized order to control carry-over effects.

2.4 Test Design

In the experiment, each convoy of cars represented a braking behavior of the lag vehicle as a reaction to the attempt of the ego driver to change lanes. A full-factorial $3 \times 3 \times 3$ within-design was selected with the factors described above: Reaction time,

Independent Variable	Levels
Deceleration	0.5 m/s², 1m/s², 1.5 m/s²
Amount of speed reduction	2 m/s, 4 m/s, 6 m/s
Reaction time	0s., 1s., 2s.

Fig. 4. Overview of independent variables.

deceleration and the amount of speed reduction of the simulated car. Figure 4 summarizes the independent variables and its levels.

2.5 Dependent Variables and Inferential Statistics

In the experiment, participants were asked to decide after each situation whether they would conduct a lane change in the next moment. For this purpose a handout with the scale was located on the seat next to the driver. Subjective ratings of perceived cooperation and criticality [29] were collected after each situation (see Fig. 5). Perceived cooperation was rated using a 15-point scale that also uses verbal labels as anchoring points based on the category-subdivision scale by Heller [30]. The rating was given in a two-step category–scaling procedure: after an initial categorization of

Parameter Description

Scale for perceived cooperation assessment of driving

How cooperative was the lag driver on the target lane?

Not at all	very little cooperative			a little cooperative			intermediate cooperative			cooperative			very cooperative		
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Scale for criticality assessment of driving

Fig. 5. Subjective, dependent measures employed in the simulator study.

perceived cooperation (e.g., "very cooperative"), a numerical rating was given. The criticality scale is a uni-dimensional scale that uses verbal labels as anchoring points (e.g., "imperceptible", "uncontrollable"), ranging from 0 to 10.

2.6 Procedure

The experimental session took approx. 1.5 h and was structured in four parts:

1. Instruction (duration up to approx. 10 min): Instruction and demonstration of the vehicle. The vehicle was equipped with an ACC (adaptive cruise control), which guaranteed that a certain THW of 2 s to the lead car was fulfilled, to guarantee the conditions for every participant.
2. Practice drive (duration approx. 10 min): Short drive in order to familiarize the drivers with the driving simulation, usage of the ACC as well as exercising the lane change task and get used to the questions asked after every situation.
3. Test drive (duration approx. 60 min): The driver was asked to follow the lead car until an acoustic sign was given to decide if a lane change is possible. Every two minutes a convoy of vehicles appeared in the middle lane with modified parameters. After each decision whether to change lanes, the participants were asked to give a subjective rating on their level of perceived criticality and cooperation.
4. Closing interview and debriefing (duration approx. 5 min).

2.7 Instruction

Participants were asked to trail a leading car on the right lane. During the drive, faster convoys of cars were passing in the middle lane (see Fig. 3) and the participants had to execute a lane change if possible, using the turn signal.

Participants were instructed verbally to complete the test course promptly without violating the traffic code (maintaining safe headway, complying with the driving in the right lane regulation, no passing on the right lane, etc.). The participants were asked to perform the task to change lanes only when they evaluated the traffic situation as to be safe enough.

2.8 Sample

A total of N = 28 participants (12 females) between the age of 21 to 69 years (M = 31.5, SD = 11.5) took part in this study. At the time of the test, the participants had between 5 to 44 years of driving experience (M = 14.2, SD = 11.2). In the preceding year, the participants had driven an average of 9,764 km (SD = 9,422 km). The participants were selected from an existing WIVW test driver pool and received extensive simulator training before the start of the study. The standardized training [32] aimed at making participants familiar with the handling of the simulated vehicle (e.g., accelerating, negotiating curves, executing turning maneuvers, braking) and reducing simulator sickness. Three subjects had to be removed due to not following or misunderstanding of the instruction. The participants received an expense allowance for taking part in the test.

3 Results

To assess the influence of the experimental factors on the dependent measures, one statistical procedure was used:

Analysis of Variance (ANOVA) with Repeated Measures: To evaluate the effects of the experimental factors on the subjective (perceived cooperation and perceived criticality) measures, a full-factorial ANOVA for repeated measures was applied. An alpha level of 5% was used.

For the analysis a 3 (deceleration) by 3 (amount of speed reduction) by 3 (reaction time) ANOVA with repeated measures was conducted with the perceived cooperation as the dependent variable. No significant interaction effects between the independent variables were found. A significant main effect of deceleration on perceived cooperation was found ($F(2, 48) = 83.0$; $p < .001$ with a partial $\eta^2 = .78$). The post hoc analysis with Bonferroni correction showed significant effects between every factor level, indicating that higher deceleration values of the car lead participants to choose higher values in perceived cooperation. Secondly, a significant main effect of amount of speed reduction was found ($F(2, 48) = 30.9$; $p < .001$ with a partial $\eta^2 = .56$). The post hoc analysis with Bonferroni correction indicated significant differences only between the lowest (24%) and the highest level (80%). Third, a significant main effect of reaction time of the lag vehicle was found ($F(2, 48) = 63.7$; $p < .001$ with a partial $\eta^2 = .73$) with significant post-hoc differences between every level, indicating that lower reaction times lead to higher values in perceived cooperation.

The same analysis was done with the perceived criticality as the dependent variable. Again, significant main effects of deceleration ($F(2, 48) = 50.3$; $p < .001$ with a partial $\eta^2 = .68$) amount of speed reduction ($F(2, 48) = 29.6$; $p < .001$ with a partial $\eta^2 = .55$) and reaction time of the lead vehicle ($F(2, 48) = 33.0$; $p < .001$ with a partial $\eta^2 = .58$) were found. No significant interaction effects were found. Bonferroni

Table 1. Means and standard deviations of perceived cooperation.

Parameter description		M	SE	Verbal category label
Deceleration (m/s^2)	$-0.5\ m/s^2$	6.6	.51	A little cooperative
	$-1\ m/s^2$	9.2	.35	Intermediate cooperative
	$-1.5\ m/s^2$	10.7	.28	Cooperative
Amount of speed reduction (%)	24%	6.8	.59	A little cooperative
	48%	9.7	.32	Cooperative
	80%	10.1	.35	Cooperative
Reaction time (s)	0 s	10.5	.33	Cooperative
	1 s	9.1	.34	Intermediate cooperative
	2 s	6.9	.48	A little cooperative

Table 2. Means and standard deviations of perceived criticality.

Parameter description		M	SE	Verbal category label
Deceleration (m/s^2)	−0.5 m/s^2	4.9	.33	Unpleasant
	−1 m/s^2	2.9	.28	Harmless
	−1.5 m/s^2	2.6	.23	Harmless
Amount of speed reduction (%)	24%	5.4	.52	Unpleasant
	48%	2.7	.20	Harmless
	80%	2.3	.21	Harmless
Reaction time (s)	0 s	2.5	.28	Harmless
	1 s	3.3	.23	Harmless
	2 s	4.7	.34	Unpleasant

corrected post hoc tests revealed a significant unidirectional influence for all levels of the independent variables. Higher deceleration and amount of speed reduction values led to the choice of lower perceived criticality, whereas lower values of reaction time led to lower criticality values. The means and standard deviations of both analyses can be found in Tables 1 and 2.

4 Discussion and Future Work

The findings of the pre-study showed that deceleration, the amount of speed reduction and reaction time are important indicators concerning the judgment, how cooperative the driver of the lane-changing vehicle perceives the lag driver on the target lane. The results are reinforced by qualitative findings [13, 23] that braking behavior is an important indicator when evaluating cooperative behavior. The main finding of the study refers to the intensity of the signal. The stronger the stimuli, in this case, the deceleration, the amount of speed reduction and the faster the reaction time, the more cooperative it is observed. We studied here changes of driving behaviour of the lag vehicle that were clearly perceptible. It might be an interesting question for future studies to determine the threshold of perceptibility according to signal detection theory [31]. Besides that recommendation can be derived which driving parameters are relevant for autonomous driving and how these parameters should be integrated into autonomous driving algorithms. In a future study, based on the subjective and objective parameters, an algorithm can be developed, which reacts in the same range of speed as described in the results of the study: the combination of the factors deceleration (1,5 m/s^2), amount of speed reduction (4 m/s) and reaction time (0 s), which were perceived as being most cooperative. Additionally, the results can be used when addressing questions concerning comfort aspects of the person seated in an autonomous vehicle thus implying which intensities of braking parameters are necessary to be perceived from the surrounding traffic.

The development of algorithms for autonomous vehicles based on cooperatively perceived human lane changing behavior is the scope of future work as well as testing the algorithm under natural conditions in a field study.

References

1. EC: WHITE PAPER: roadmap to a single European transport area. In: European Commission, p. 31 (2011)
2. Sivak, M., Schoettle, B.: Road Safety with Self-Driving Vehicles: General Limitations and Road Sharing with Conventional Vehicles. University of Michigan, Ann Arbor, Michigan (2015)
3. Nees, M.A.: Acceptance of self-driving cars: an examination of idealized versus realistic portrayals with a self- driving car acceptance scale. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **60**(1), 1449–1453 (2016)
4. Hoff, K.A., Bashir, M.: Trust in automation. Hum. Factors **57**(3), 407–434 (2015)
5. Ahmed, K.I.: Modeling drivers' acceleration and lane changing behavior. Transportation (Amst) Ph. D., 189 (1999)
6. Gipps, P.: A model for the structure of lane-changing decisions. Transp. Res. Part B **20**(5), 403–414 (1986)
7. Choudhury, C.F., Ben-Akiva, M.E., Toledo, T., Lee G., Rao, A.: Modeling cooperative lane-changing and forced merging behavior. In: 86th Annual Meeting of the Transportation Research Board, Washington, DC. (2007)
8. Färber B.: Kommunikationsprobleme zwischen autonomen Fahrzeugen und menschlichen Fahrern. In: Autonomes Fahren, pp. 127–146. Springer, Berlin (2015)
9. Merten, K: Kommunikationsprozesse im Straßenverkehr (1977)
10. Maurer, M., Gerdes, C.J., Lenz, B., Winner, H.: Autonomes Fahren (2015)
11. Risser, R.: Kommunikation und Kultur des Straßenverkehrs (1987)
12. Watzlawick, P., et al.: Menschliche Kommunikation: Formen, Störungen, Paradoxien. Huber, New Jersey (2007)
13. Ellinghaus, D.: Rücksichtslosigkeit und Partnerschaft. IFAPLAN, Köln (1986)
14. Maag, C.: Fahrer und Verkehrsklima (2004)
15. Benmimoun, A., Neunzig, D., Maag, C.: Effizienzsteigerung durch professionelles/partnerschaftliches Verhalten im Straßenverkehr. FAT-Schriftenreihe no.1811, Frankfurt/Main (2004)
16. Zimmermann, M., Fahrmeier, L., Bengler, K.J.: A Roland for an Oliver? Subjective perception of cooperation during conditionally automated driving. Int Conf. Collab. Technol. Syst. CTS **2015**, 57–63 (2015)
17. Kandada, S., Bham, G.H.: Driver mandatory lane change behavior: use of governing gap in critical gap estimation. In: Transportation Research Board, 92nd Annual Meeting (2013). Paper No. 13-5151
18. Henning, M.J.: Preparation for lane change manoeuvres: behavioural indicators and underlying cognitive processes, p. 163 (2010)
19. Green, M.: How long does it take to stop? Methodological analysis of driver perception-brake times. Transp. Hum. Factors **2**(3), 195–216 (2000)
20. Johansson, G., Rumar, K.: Drivers' brake reaction times. Hum. Factors **13**(1), 23–27 (1971)
21. Broen, N.L., Chiang, D.P.: Braking response times for 100 drivers in the avoidance of an unexpected obstacle as measured in a driving simulator. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **40**(18), 900–904 (1996)
22. Ellinghaus, D., Welbers, M.: Vorschrift und Verhalten. Eine empirische Untersuchung über den Umgang mit Verkehrsregeln, 3. Auflage. IFAPLAN, Köln (1978)
23. Ehmanns, D.: Simulation Model of Human Lane Change Behaviour. In: Der Fahrer im 21. Jahrhundert?: Tagung Berlin, 3. und 4. Mai 2001, pp. 203–216 (2001)

24. Kesting, A., Treiber, M., Helbing, D.: General lane-changing model MOBIL for car-following models. Transp. Res. Rec. J. Transp. Res. Board **1999**(1), 86–94 (2007)
25. Gerlough, J., Daniel, L., Huber, M.: Traffic Flow Theory: A Monograph, pp. 17–31, Transportation Research Board, Washington D.C. (1975). Special re.
26. Barreit, G.V.: Feasibility of studying driver reaction to sudden pedestrian emergencies in an automobile simulator. Hum. Factors **10**, 19–26 (1968)
27. Evans, L.: Traffic safety. Science Serving Society, Bloomfield Hills (2004)
28. Ohta, H.: Individual differences in driving distance headway. In: Gale, S., Brown, A.G., Haslegrave, I.D., Moorhead, C.M., Taylor, I. (eds.) Vision in Vehicles, pp. 91–100. Elsevier Science, London (1993)
29. Neukum, A., Naujoks, F., Kappes, S.: Kontrollierbarkeit unerwarteter Eingriffe eines Bremsassistenzsystems aus Perspektive des Folgeverkehrs
30. Heller, O.: Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). Psychol. Beiträge **27**, 478–493 (1985)
31. Green, D.M., Swets, J.A.: Signal Detection Theory and Psychophysics. Peninsula Pub, Baileys Harbor (1988)
32. Hoffmann, S., Buld, S.: Darstellung und Evaluation eines Trainings zum Fahren in der Fahrsimulation. In: VDI Wissensforum (ed.) Integrierte Sicherheit und Fahrerassistenzsysteme, pp. 113–132 (2006)

Augmented System Verification Using Automated Testing

Amir Schur[1(✉)] and Mike Williams[2(✉)]

[1] Odyssey Systems Consulting Group, 19 Crosby Drive, Suite 220, Bedford, MA 01730, USA
amirschur@odysseysystemsconsult.com
[2] Raytheon Integrated Defense Systems, 50 Apple Hill Drive, Tewksbury, MA 01876, USA
mike.williams@raytheon.com

Abstract. A verification process is an evaluation of whether or not a product satisfies its requirements, which typically includes a visual verification activity. This is sometimes referred to as acceptance testing. Minimal or no human interaction is the construct of fully automated testing, which is often not sufficient for verification process/user acceptance testing. An approach for writing automated testing is described here, which augments the human verification process while still leveraging the benefits of test automation.

Keywords: Human-systems integration · Systems engineering · Agile testing · Automated testing

1 Introduction

Throughout the software development life cycle, developers conduct levels of testing which progress to include larger portions of the system from unit testing, component testing, integration testing, component integration testing, system integration testing, system testing and acceptance testing [1]. The goal of this paper is to discuss a process that utilizes automated system level testing in a manner that supports witnessed acceptance testing.

System testing consists of positive and negative path testing, corner cases, fault insertion, and other tests performed using some or all of the target system. Within the full set of System Tests is a subset of tests that best show that the product meets the system level requirements. This subset of tests will be performed for customer witnesses during acceptance testing to provide evidence for acceptance. Due to the importance of these tests, they will be included in the Regression Test Suite (RTS), which are tests executed throughout the development cycle whenever the system changes, or upon release of a new software build. There is significant motivation to automate the tests in the RTS to realize the following benefits: reduction of time to execute and analyze the RTS, reduction/elimination of human involvement in execution, reduction/elimination of operator errors during execution, and to make possible off-hours scheduling of RTS to increase test asset utilization. However, automation of tests is initially more expensive than just straight execution of tests, as an automation

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_7

infrastructure needs to be procured and installed, and the automated tests need to be generated. The Return-On-Investment (ROI) of automation comes primarily from the fact that every iteration of the RTS comes at a reduced cost. An additional ROI can be achieved if the formal acceptance testing can use the automated tests as is, without additional test or procedure development.

Acceptance testing is an evaluation of whether or not a product satisfies its requirements. In order to ensure the system performs correctly, the tester/system engineer must verify each product requirement using verification methods of inspection, analysis, demonstration, or test [2]. A core feature of traditional acceptance testing is that the testing is performed for witnesses who, based on their observations, will accept the functionality being demonstrated. Fully automated testing executes with minimal or no human interaction. This is suitable for software unit testing, regression testing, and often system testing. In our agile software development project, this automated testing capability is not sufficient for acceptance testing, as human visual verification is still required. The method of verification, as it satisfies the system requirements, needs to be explicitly written in the test procedures to include a visual method of verification.

With some modifications of our automated test development methodology, we are able to utilize automated testing to augment the human verification process. We would like to describe how we accomplish this. Human verification during acceptance testing is still the primary focus; we employ automated testing as a tool to assist the human verification process.

2 System Verification Process

Well-written requirements are the foundations of a project. A clear understanding and communication of the requirements of the system is critical to development and test. This is only possible through a close collaboration between developer and customer. Classic waterfall development processes accomplish this through development of documents, which go through multiple levels of review. Agile development methods replace the process of extensive documentation and review cycles with rapid prototyping and development iterations. A technical partnering review of the strengths of and desired modifications to a prototype is generally more effective than attempting to agree on documentation describing a final product, which ultimately all parties may be interpreting differently [3].

This concept extends to acceptance testing as well. A clear understanding of the verification approach to satisfy each requirement is necessary. Whenever possible, customer and developer should collaborate on verification steps over mock-ups, prototypes, or the actual system itself, if available. A technical working group that works on the system or prototype and walks through the operator interactions, expected system responses, and analysis of test output data, leads to clear agreement on the procedure that will be executed during the formal acceptance activities. This partnered procedure can then go into automated test development with the expectation that there will be little rework necessary.

Classical methods of verification in systems engineering includes Analysis, Demonstration, Test, and Inspection. Industry definitions of these methods vary, but commonly Inspection and Analysis do not require the actual operation of the system, rather they involve physical measurements, specification sheets, or other documentation or data. Demonstration and Test methods both require the system to be operated in some manner and displayed to a witness or to generate data that is then analyzed. As such, the process we describe in this paper best applies to Test and Demonstration Verification Methods (VM).

Within Demonstration or Test requirements, the focus of our process varies. For Test Requirements, the tester executes the test procedure, collects test data, and analyzes that data for evidence for acceptance. While the test procedure is important and must be correct and repeatable, the focus is not on how the operator interacts with the product, or what is displayed to the witness during execution. Rather, provided system inputs, do the product algorithms and processes make the desired decisions and provide the expected answers. For Demonstration Requirements, the focus is precisely on the interaction of the operator with the system, and the display of data to that operator and witness. The difference in focus influences test automation methods, which we will review in the next section.

Regardless of the Verification Method, human verification is the essence of the activities. A person must witness the demonstration and verify test results to ensure requirements are fully satisfied. The goal is an automated Regression Test Suite that is run for each new release leveraging the benefits of automation, but which can also be conducted as-is for a human witness to view the system interactions, responses, and data that allow them to make an acceptance decision.

We will describe an example of automated testing capabilities in the next section. Keep in mind that full automation will not necessarily be sufficient in requirement verification methods. A machine cannot perform independent analysis, nor can it perform test verification. In some circumstances, software automation could perform the activities, but a human must still be in the loop to perform the final validation.

3 Automated Testing to Support Acceptance

Test Automation is the process of using automation tools and processes to develop test cases, which are initiated, executed, and analyzed without human intervention. One framework available for developing automated tests is Cucumber. It is a behavior-driven development tool intended to automate various tests, including acceptance tests. The Cucumber framework includes the Gherkin programming language, a business readable language that supports a simple test-oriented structure and acts as the “orchestration” layer of the automated test. There are other language options than Gherkin, and it is used here merely for convenience. Each line of Gherkin is bound to a lower level language that actually executes the test actions. If these test actions involve interaction with a system Graphical User Interface (GUI), a product such as EggPlant™ can drive the GUI. This allows the Gherkin to remain at the test procedure level, although it is actually executable software.

The key for use of automated tests in Acceptance Testing is writing Gherkin in a manner that clearly supports verification. The only requirements that Gherkin places on the developer are that the lines start with a key word and are then correctly bound to executing code. It is up to the automated test developer to follow a convention such that the Gherkin code provides a clear procedure that the witness can follow, and which provides them with all of the information necessary to verify the requirement.

An example of Gherkin written in a manner that supports verification of demonstration requirements and one that does not, follows. We will use simple requirements for a fictional Air Traffic Control System, which we will call ATCS. The function of ATCS is to display to the operator aircraft information as it becomes available. Since these are demonstration requirements, their acceptance would be via the system displays:

Reqt 1: The ATCS shall visually provide the aircraft tail-number on the main display for all aircraft.
Reqt 2: The ATCS shall provide a separate display with the following details for each aircraft: tail-number, transmission frequency, altitude, heading, speed, and final destination.
Reqt 3: The ATCS shall provide a separate display with landing allocations for each aircraft with runway, approach altitude, and approach route.

Below is a short example of Gherkin code written in a manner that supports acceptance testing. The primary Gherkin keywords in use, and their meaning, are:

Feature: Defines the start of a group of automated tests, with description
Scenario: Defines the start of a single automated test, part of a larger feature
Given: Preconditions - these provide the witness with context
And: Connector that continues the previous keyword
When: Actions - any test input or action relevant to the verification
Then: Results - the action that the witness must verify to pass the requirement
$ - denotes a variable

```
Feature: Air Traffic Control
  As an Air Traffic Controller, I want to maintain situa-
tion awareness and provide landing information to ap-
proaching aircraft.

Scenario: Single Aircraft enters system airspace
  Given GUI automation software is active
  And $operator is logged onto workstation $ws1
  And Air Traffic Controller Application is started
  And Air Traffic Simulator is started on Simulation
Server
  And Scenario $Single_Approaching_Plane is running

  When 1 Aircraft enters airspace
  Then Operator Main Display shows Aircraft icon with
     tailnumber
  Then Entity Information Menu displays Aircraft tail
     number, transmission frequency, altitude, heading,
     speed, route
  Then Recommendations Menu displays allocated runway,
  approach altitude, and approach route
```

This Gherkin is executable code, and each line kicks off lower level software to perform the actions listed: logging onto the selected workstations as the selected user, starting the Air Traffic application, etc. When the preconditions are complete, the trigger for activity that the witness must view is that an air track has entered the airspace, signified by the keyword "when". The automated test waits on this line for this condition, and any activities occurring in the scenario prior to this happening are not relevant to verification of this requirement, so are not represented in the procedure. When the air track enters the airspace, verification of multiple items occur in separate lines. Each line contains information about what display will provide the information, and what items need to be present. For these lines to pass automated verification, the items in the display are compared to expectations, and assessed as pass or fail. A pass/fail for the entire test is the roll up of the pass/fail of each separate line. For these items to pass human verification, the witness needs to visually confirm them on the screen.

An exaggerated and simplified example of Gherkin code that fully supports the automated verification, but would not provide a witness the required information for Acceptance Testing follows. The underlying code for this example could be the exact same underlying code as the previous example, so the test case would be just as valid and robust for automated execution:

```
When 1 Air Track enters airspace
Then Reqt1 is verified
Then Reqt2 is verified
Then Reqt3 is verified
```

The actual executing lower layers of software would be exactly the same for the two examples. However, the second example is not a sufficient acceptance testing procedure because it does not contain the key elements to support verification, namely those items that the witness must observe to confirm verification.

As mentioned previously, requirements with a Verification Method of Test may be handled differently than Demonstrations. While Demonstrations must interact with the product GUI to be valid, Tests requirements often do not, and avoiding the GUI may make the test more robust. GUI automation tools use image comparison to locate the necessary buttons and GUI features. The developer “snaps” an image of a desired feature (button, textbox, etc.) which the automation tool stores, and when interaction with that feature is desired, the GUI automation scans the screen until it is found or the scan times out. The Image comparison search has failure modes that include that the feature is either: partially or totally obscured by something else on screen, slightly changed in appearance (color, etc.), or not on screen due to timing reasons. There is motivation to eliminate these failure modes by having the automation software bypass the GUI and interact with the software layer beneath the GUI. There are well-documented methods for implementing this, which we will not elaborate on, but the shift away from GUI interaction causes a change in the development of the Gherkin software. The focus of the Gherkin code continues to be to provide the witness with all information necessary to perform requirement verification. This is demonstrated in the following simplified example that includes a Test Requirement and the Gherkin that would support it. This is a Test requirement because human visual observation of the scenario will not have sufficient precision to confirm compliance with the timing aspect of the requirement. Time-stamped data logging and post-run analysis is necessary.

> Reqt 4: The ATCS shall determine runway allocation for each new aircraft within 1 s of an aircraft entering ATCS airspace

```
Scenario: Single Aircraft enters system airspace
  Given GUI automation software is active
  And $operator is logged onto workstation $ws1
  And Air Traffic Controller Application is started
  And  Air  Traffic  Simulator  is  started  on  Simulation
Server
  And Scenario $Single_Approaching_Plane is running

  When 1 Aircraft enters airspace
  Then ATCS records airspace entry timestamp
  Then  ATCS  completes  runway  allocation  and  records
  timestamp
  When scenario completes
  Then  $runwayAnalysisTool  performs  analysis  to  verify
  runway allocation in less than 1 second
```

This simplistic example provides the witness with the key high-level events that will inform the pass/fail assessment when the test is complete. The witness may very well wish to see the lower levels of software to determine that the timestamp mechanism will provide valid results, but when convinced of that will accept the results of scenarios that include thousands of planes.

There is an additional element necessary to allow use of automated tests for acceptance of requirements. There are constraints associated with acceptance testing using automated tests that do not exist for regression testing. For regression testing, the desire is to execute the test suite as rapidly as possible to minimize the scheduling needs on the testing assets. For witnessed acceptance testing, the steps must occur at a rate such that the witness can confirm each verification item. Our project found a simple solution for this issue by running the automated tests at full-speed for regression testing, and in line-by-line debug mode for acceptance testing. When written in the correct manner, as demonstrated above, this worked very well. Each line of Gherkin can be stepped through, and for each line, the automation brings forward the necessary displays, performs its image verification, and then pauses waiting for the next action. At that point, the witness can view all verification evidence, and when complete, advance to the next step. For time sensitive tests, such as recording the timestamps between steps, the test execution can utilize breakpoints that allow the execution of several steps at once to preserve the validity of the test.

4 Conclusions

We have shown a method of development of automated tests that supports acceptance testing. This method may differ depending on the particular requirement and Verification Method of that requirement, but the key to this method is in the generation of high-level automated code that is acceptance testing focused and provides the witness with a clear roadmap for acceptance. It does this by including only those steps that are test relevant, by utilizing the Gherkin key words to identify those steps that are providing verification evidence, and by including scenario elements and actions that allow the witness to make an informed assessment. This process allows the program to realize the benefits of automated testing throughout the many iterations of regression testing, with the knowledge that the exact tests that are being regression tested will be utilized for acceptance testing with no additional procedure development.

References

1. Padmini, K.J., Perera, I., Bandara, H.D.: Applying agile practices to avoid chaos in user acceptance testing: a case study. In: 2016 Moratuwa Engineering Research Conference (MERCon), pp. 96–101. IEEE, April 2016
2. Alexander, M.G.: Requirement Assurance: A Verification Process, Langley Research Center, Hampton, Virginia (2011). https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110014948.pdf
3. Paetsch, F., Eberlein, A., Maurer, F.: Requirements engineering and agile software development. In: WETICE, vol. 3, p. 308 (2003)

Task Design in Human-Robot-Interaction Scenarios – Challenges from a Human Factors Perspective

Patricia H. Rosen(✉) and Sascha Wischniewski

Human Factors, Ergonomics, Federal Institute for Occupational Safety and Health (BAuA), Friedrich-Henkel-Weg 1-25, 44149 Dortmund, Germany
{rosen.patricia,wischniewski.sascha}@baua.bund.de

Abstract. The production environment is facing a constant change. New technologies emerge and the latest development in the field of direct human-robot-interaction will create new forms of hybrid working systems. The question arises on how to design tasks in these new working systems. The aim of this paper is to outline current conditions of German production work, focusing on task design characteristics. Based on that analysis conclusions are drawn on how to design hybrid work systems using light-weight robots. The paper presents task specific results of two national employee surveys and of a scoping review on job control aspects, as specific task design features. The analyses reveal that task variability, timing and method control have a strong impact on employees' wellbeing. Based on the findings guidelines for task design in HRI scenarios in manufacturing cells with light-weight robots are presented and discussed.

Keywords: Task design · Human-robot-interaction · Job control · Production work · Health

1 Introduction

The manufacturing sector and therefore production work is of great importance for the German economy. It has been continuously representing more than 20% of the German gross domestic product (GDP) during the last two decades. Due to the differences in labor costs in Germany (38 € per hour in 2015 for the manufacturing sector) compared to other countries in the world, productivity is essential for the competitiveness of work [1, 2]. In the production environment, increasing the degree of automation (DOA) in work systems is a possible answer to cope with this challenge. Light-weight robots (LWR) start closing the gap between automation costs and production volume.

However, increasing the DOA can result in stressful work conditions for example characterized by a higher workload or a loss of situation awareness [3]. Automation might also have an effect on task characteristics like job control and resulting monotony or time pressure, which again affect employees' health. The following work will address this issue.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_8

Latest technological developments take place especially in the field of sensors and actors. For this reason new generations of light-weight robots including necessary security technologies arise, allowing new forms of interaction and a close cooperation or collaboration of humans and technology. A recent study examined companies' motives to implement LWR technologies: Main reasons were to increase operational efficiency followed by innovation and improving ergonomics [4]. This brings up the question of how to design production tasks for these new forms of hybrid work systems using the flexibility of robots and at the same time reaching a human-centered and competitive DOA.

2 Production Work in Germany 2016

About one fourth of the working population in Germany can be assigned to the manufacturing sector. This proportion decreased from 2000 to a stable level of 24.5% since 2010 (see Fig. 1).

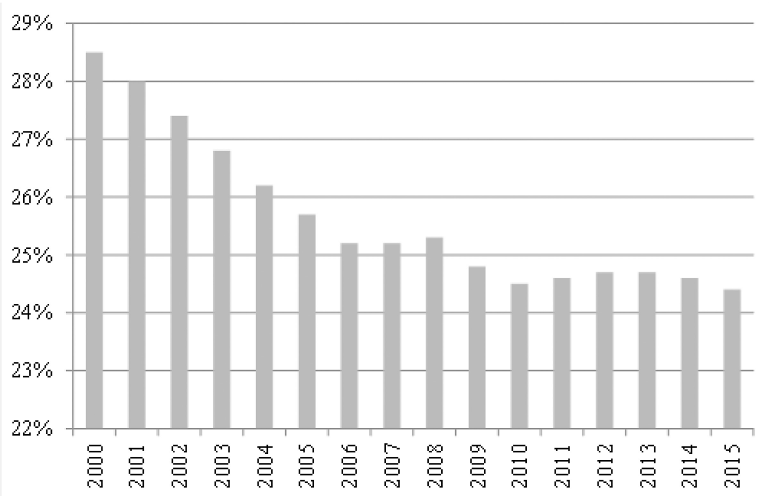


Fig. 1. People employed in the manufacturing sector in Germany [5].

The following information on production work in Germany were taken from the representative employment survey "Working time reporting for Germany" recently carried out on behalf of the German Federal Institute for Occupational Safety and Health (BAuA). For the reported weighted results, data from individuals was analyzed who stated to work fulltime in production systems in industry (weighted and therefore calculated n = 1211). The dataset included subjects aged between 18 and 67 with a mean age of 43.14 (SD = 11.31). The gender distribution shows that the sample contains 85.2% male participants and 14.8 female participants.

When analyzing production work the task characteristics machine-determined work rate (mdwr) as well as job rotation are two major aspects to consider: Within the given dataset 78% state to rotate between workstations. 45.7% indicate to work in machine-determined work rates. 37% state to rotate between workstations and to be machine-determined in their work. This implies that 82% of the subjects who report working in machine-determined work rates also report to rotate between workstations. 8% of the sample is exposed to mdwr only and not to job rotation (see Fig. 2).

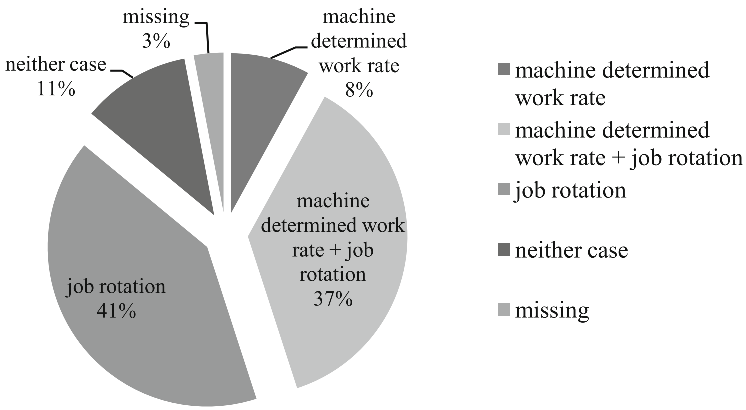


Fig. 2. Reported task characteristics by interviewed employees.

Taking a closer look at the reported average machine-determined work rates shows an accumulation of cycle times below 10 min in two thirds of the work systems. Within this group half of the subjects report cycle times of less than 2 min (see Fig. 3).

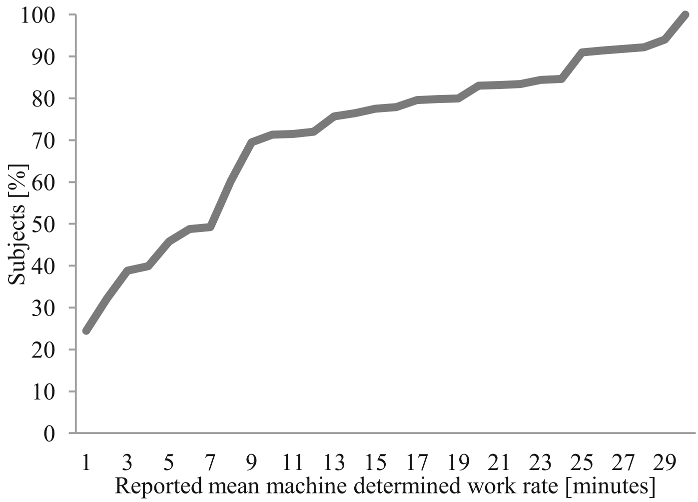


Fig. 3. Cumulated percentage of the reported average machine-determined work rates.

Further, self-reported health-related outcomes were analyzed within the sample of employees stating to work in industry production systems. Figure 4 presents the frequencies of complaints that employees state to suffer from. Data is presented for the industry sample and for the overall German working population. In both samples, participants suffer from general tiredness and exhaustions, neck pain, shoulder pain and back pain the most. In the production work sample 56% participants state to suffer from general tiredness and 47% of the participants suffer from physical exhaustion. 39% report sleeping problems and 23% state to suffer from emotional exhaustion.

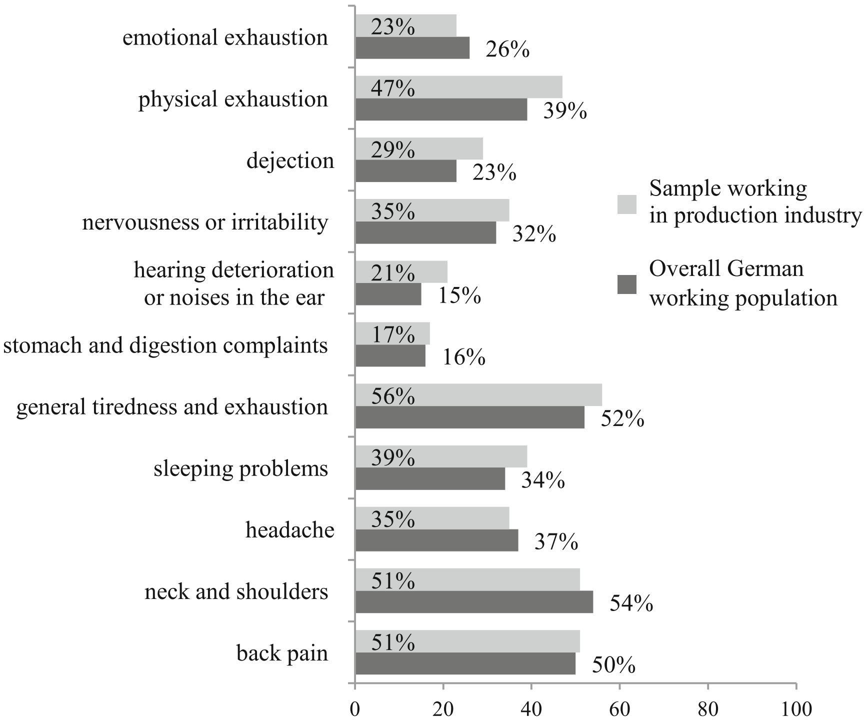


Fig. 4. Frequencies of self-reported health variables for production industry and overall working population.

3 Task Characteristics and Health, Well-Being and Performance Outcomes in Literature

The described task characteristics like machine-determined work rates or job rotation, that dominate the nature of production work, comprise different features of the theoretical construct of job control [6, 7]. The concept of job control is well known in work psychology and incorporates various aspects. For example the aspect of timing or method control, the possibility to rotate between different workstations or the amount of latitude an employee is given.

The concept of job control in the area of production was addressed in an extensive literature review. A systematic literature search was conducted in interdisciplinary databases applying a complex search algorithm. Only studies with samples stemming from the manufacturing sector were included. After a two-stepped selection process, 106 studies in total were included in the review. In this paper, the focus is put on the results of 21 studies dealing with similar structured tasks. The tasks in scope do not vary in their complexity and mainly address timing- and method control, routinization, job rotation, monotony or the concrete cycle time. Studies dealing with aspects of decision authority like skill discretion or decision latitude are reported elsewhere [8].

The review explores the relationship of these specific job control features and employees' health, wellbeing and performance. 13 studies address the association between the aforementioned job control aspects and different health outcomes. Table 1 gives an overview of the distribution of outcomes analyzed with job control in the included studies. Especially health outcomes and aspects of wellbeing or job satisfaction have been studied frequently. Only a very little number of studies address the association between aspects of job control and performance (see Table 1).

Table 1. Distribution of outcomes associated with job control.

Outcome variable	Number of analyzed studies	
	No. of studies finding a relation	No. of studies not finding a relation
Physical health	6	1
Mental health	6	1
Wellbeing, satisfaction	10	0
Performance	1	–

The analysis shows that three longitudinal and five cross-sectional studies observe weak to medium strength effects between low levels of timing- respectively method control and health aspects. Relations to physical health outcomes like musculoskeletal complaints are found here [9] and here [10]. Both studies use odd ratios (OR) and correlations coefficients (r) as well as their significance (p) to show that higher control in work pace reduces neck- and back pain (OR = .44; 95% CI 0.26 – 0.76) respectively less control increases tension in the upper extremities ($r = 0.30$, < 0.01). While an association between general musculoskeletal complaints and timing control was observed, an association with the more specific diagnosis carpal tunnel syndrome was only found within a cross-sectional design (OR = 1.59; 95% CI 1.04 – 2.43) [11] but not in longitudinal data [12]. Physical complaints can also be observed among employees being exposed to short work cycles. Working shorter cycles than ten seconds (OR = 1.90; 95% CI 1.04 – 3.48) respectively fifteen seconds ($r = 0.69$, $p < 0.01$) is associated with musculoskeletal complaints [10, 11].

Beside physical outcomes several studies address outcomes related to mental health. Three cross-sectional studies observe effects of timing- and method control on depressive symptoms, occupational burnout or self-rated mental health. The three studies show, that the lower the method- and timing control an employee exerts, the higher the self-rated level of psychological stress in form of depressive symptoms ($r = -0.26$, $p < 0.001$) [13], ($r = -0.35$, $p < 0.01$) [14] and the lower the self-reported mental health ($r = -0.16$, $p < 0.001$) [15]. The aforementioned studies analyzed depressive symptoms and did not focus on diagnosed mental disorders. In contrast Inoue and colleagues emphasized long-term sick leave due depressive disorders. Their longitudinal study shows a medium strength relationship between diagnosed depression and job control (OR = 1.7; 95% CI 1.3 – 2.2). In the large sample of 15.256 employees the authors observe, that people with higher timing- and method suffer less from the mental mood disorder depression [16].

Task variability and job rotation also show significant associations with physical and mental health aspects. Three publications indicate, that a lack in task variability increases musculoskeletal complaints (OR = 1.7; 95% CI 1.3 – 2.2) [17], (OR = 1.36; 95% CI 1.35 – 1.63) [18], (OR = 2.3; 95% 1.9 – 2.8) [19]. Concerning mental health aspects, Hsieh and Chao show in their results, that employees who rotate between tasks suffer less from emotional exhaustion (r = −0.11, p < 0.05). Though no relation was found with the other burnout dimensions cynicism or professional efficacy [20]. Melamed et al. analyze subjective monotony as well as work cycles shorter than one minute and their relation to psychological distress. For both variables, a significant association can be observed. Indicating that higher monotony ($r = 0.21$, $p < 0.05$) and short work cycles ($r = 0.14$, $p < 0.05$) are related to greater distress [21].

Apart from general health variables, the review addresses aspects of wellbeing, motivation and job satisfaction and their relation to job control. Two studies report a significant effect for reduced work-related anxiety when being exposed to higher levels of timing- and method control ($r = -0.33$, $p < 0.001$; $r = -0.25$, $p < 0.001$) [13, 14]. The same applies to the association of timing- respectively method control and self-reported short term strain ($r = -0.34$, $p < 0.001$; $r = -0.12$, $p < 0.05$) [22, 23], as well as for repetitive tasks and short term strain (OR = 1.3; 95% CI 0.6 – 2.2) [24]. In total six studies analyze the association between timing control, method control, task variability and measurements of work satisfaction. The analyses show, that all results point in the same direction. Greater levels of timing and method control as well as higher task variability show positive associations with job satisfaction [14, 15, 21, 25], motivational aspects like generating new ideas [26], fewer numbers of fluctuation [27] and higher self-reported work ability [28]. More detailed results are presented elsewhere [8]. Compared to health and wellbeing, performance measures and their relation to aspects of job control have not been studied in such detail. Nevertheless Jorgensen et al., show, that job rotation is employed by manufacturing companies in order to reduce workplace accidents ($r = 0.42$, $p < 0.05$) [27].

The literature review on aspects of job control and their relation to various outcomes in the field of manufacturing shows, that job control plays an important role for employees' health and wellbeing. Observed effects sizes vary between small and medium, nevertheless indicating a stable association between the variables in question. Although the presented results seem to show in one direction, a very concrete description of the analyzed working system is missing. Very little is known about the settings the research was conducted in. Apart from subjective research methods, no attempt is made to describe the individual task characteristics or the technical environment in more detail. Although this can be useful information, merely the branch is mentioned briefly. Further the concrete impact of each job control facet remains unclear, as different measures are used in the analyzed studies and seldom a differentiated picture of the concept is drawn.

The data show the relevance of job control for the manufacturing sector, yet the question arises: How can we use this knowledge in order to design tasks and future manufacturing jobs with semi-automatized tasks and human-(light-weight-) robot-collaborations?

4 Chances and Challenges for Task Design in Hybrid Human-Robot Work Systems

4.1 Facets of Job Control and Their Importance

As can be seen in chapter 3, job control plays an important role in terms of task design in manufacturing. The association with employees' health and wellbeing is undeniable. However, the concept itself is very broad and it is challenging to draw concrete conclusions in terms of work design. Most literature on job control addresses the concept as a whole, whereas practitioners are more interested in explicit knowledge on how to design specific workplaces. Generic design recommendations for example state, that individual job control should be increased. This advice might refer to various job control aspects like timing control, method control, production responsibility, problem-solving demand, monitoring demand [29] or decision authority. Yet it remains unclear, whether all facets are equally important or whether one is more important than the other. Especially new work-systems, like hybrid human-robot teams need precise recommendations on how to design humane jobs.

In order to specify the most important aspects of job control for manufacturing work places, a sub-sample from a representative employment survey of the German working population from the year 2012 [30] was analyzed. The sub-sample comprises manufacturing jobs like assembly tasks or operating machines. In total 313 people were asked to rate, to what extend (frequently, sometimes, rarely, never) they are exposed to different working conditions that all relate to the concept of job control. If people stated that they were exposed to one condition frequently, they were then asked whether this condition was stressful for them. The results show, that aspects referring to timing and method control had the highest impact on people:

- 46% reported that they had to achieve a target piece number, performance or time, i.e. they only had little timing control, of these more than half (54.2%) rated, that this condition causes stress.
- 47.9% suffered from time pressure often, again a lack in timing control, and here two-thirds (66%) felt stressed due to that condition.

The subjects were also questioned, whether they were exposed to constantly repeating work cycles and whether their tasks were compulsory in every detail, both aspects referring to method control.

- 74.1% were exposed to repeating work cycles often, but only 18.5% felt that this condition caused stress.
- 51.6% reported to have tasks, which are compulsory in every detail, often. Here almost a third (31.7%) stated to feel stressed by that working condition.

4.2 Implications for the Design of HRI

The results of the review as well as the analysis of the German datasets show, that job control plays an important role for employees' health and well-being in production tasks. Especially the aspects of machine-determined work rate and timing control seem to have a significant impact. Therefore these aspects should be considered when designing human-robot-interaction (HRI) scenarios.

When designing work systems like human-robot-teams, it is necessary to define the tasks of each team member. In their study Bauer et al. outline the different types of main applications and their frequencies LWR are used for. They identify three groups of tasks. The first group, of most frequently used applications, includes tasks like „gripping" or „joining" of parts. Tasks like "pick & place", "machine loading" and "packing & palletizing" comprise the second group. The last group, with the least used application, includes tasks like "screw tightening", "gluing" or "welding" [4]. Further, the authors define different types of interaction between humans and LWR based on their shared workspace (see Fig. 4 and 5).

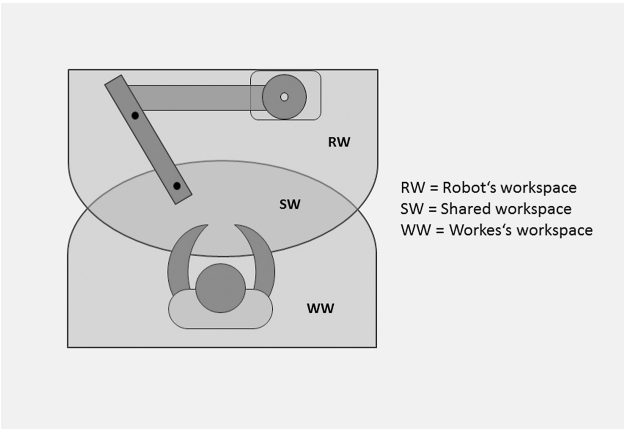


Fig. 5. Workers' and robots' workspaces according to [4].

Bauer et al. defined five levels of cooperation with a LWR depending on the amount of overlap in the robots' and workers' work zones and the type of task [4]:

- Level 1 (cell): No genuine cooperation, robot is located in a cage
- Level 2 (coexistence): No shared workspace, human and cage-free robot work alongside
- Level 3 (synchronized): Human and robot share a workspace but only one is present in the workspace at a given time
- Level 4 (cooperation): Human and robot have to perform tasks at the same time in the shared workspace but do not work on the same product at the same time
- Level 5 (collaboration): Human and robot work on the same product at the same time

In terms of job respectively task design different recommendations arise, depending on the level of cooperation.

In case of cell or coexisting levels of cooperation, workplace design addresses a more superior level. Planners have to allocate the tasks in question either to the human worker or the robot. Here the question arises, whether the human worker is left with residual tasks like stocking the robot. Only after designing the work system on this meso-level, aspects of timing- and method control can be directed in a second step. Here the requirement applies that machine-determined work rates should be avoided. A dynamic task allocation, specified by the worker depending on the daily production demands, might be considered [31] to substantially increase job control for these kinds of work systems.

For level three to five task design recommendations directly apply to the two facets timing and method control. Both aspects should be given attention. In terms of timing control the following recommendations should be considered:

Avoiding machine-determined work rates by creating possibilities

- to interrupt the working process, e.g. by starting the LWR self-determined;
- to initiate the process on one's own pace, e.g. adjusting the robots work pace;
- to add buffers in the working system.

Apart from the aspects of timing control, planners should address the following task design recommendations regarding method control:

- a greater variance in terms of task procedures, i.e. enabling more than one way to conduct an assembly task for example;
- giving the worker the possibility to change task sequences and procedures, meaning that a form of task rotation takes place without actually moving to a different working station.

The aspect of greater variance in task procedures can be contrary to the strategy of standardization in order to achieve the most competitive process. On the one hand, giving standards for each step related to a task this is a common industrial engineering approach. On the other hand, the flexibility light-weight robots bring with them, allows alternative task designs. Like Bauer et al. point out, the lightest form of cooperation "coexistence" is a good way to start with human-robot-interaction. More possibilities in terms of task design arise from the others levels of cooperation. Considering the above described task design aspects already in the planning stage, enables a human-centered human-robot-interaction.

5 Conclusion

As technologies develop further, new forms of working systems will emerge. Especially the field of direct human-robot-interaction generates challenges in terms of human-centered task design. Light-weight robots in particular can fill the gap between automation costs and production volume and will increase the degree of automation.

Physical complaints like musculoskeletal pain and mental complaints like fatigue or exhaustion are common health problems in the production industry. When looking at productions tasks more closely, it is evident that job control is the core concept. The results presented in this paper show, that in production tasks especially the job control aspects timing and method control play an important role for employees' health and wellbeing. Moreover, the results of the presented literature review clearly indicate the importance of task variability, job rotation, method- and timing control not only for several health outcomes but also for employees' motivation and job satisfaction.

New work systems including LWR are more flexible and therefore allow greater variance when designing collaboration tasks. Different levels of cooperation between human workers and robots enable a distinct task design. The presented task design recommendations make a first attempt to give concrete advice for practitioners and planners of modern human-robot-interactions based on a valid empirical fundament.

This article presents frequencies of specific features of production jobs like machine-determined work rates, job rotation and their combination. In a next step, the impact of these task characteristics on health related outcomes in the described sample will be analyzed. Further research is also necessary in order clarify the specific impact of the various job control facets. In addition, the job control aspects of decision latitude and skill discretion and their role in manufacturing have not been focused in depth. Both aspects will be addressed in future research.

In this manner, new emerging work systems like modern human-robot-interactions can be designed using well-known human factor guidelines and perspectives.

References

1. Destatis: EU-Vergleich der Arbeitskosten 2015: Deutschland auf Rang acht. Federal Statistical Office of Germany, Wiesbaden (2016)
2. Destatis: Anteil der Industrie am BIP seit 20 Jahren nahezu konstant, Federal Statistical Office of Germany, Wiesbaden (2015)
3. Onnasch, L., et al.: Human performance consequences of stages and levels of automation: an integrated meta-analysis. Hum. Factors **56**(3), 476–488 (2014)
4. Bauer, W., et al.: Lightweight robots in manual assembly—best to start simply! Frauenhofer-Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart (2016)
5. Destatis: Labour market—employment (national concept) by industries. Federal Statistical Office of Germany (2016)

6. Hacker, W.: Allgemeine Arbeits- und Ingenieurpsychologie : psychische Struktur und Regulation von Arbeitstätigkeiten, vol. 472. VDR, Berlin (1973)
7. Karasek, R., Theorell, T.: Healthy Work: Stress, Productivity, and the Reconstruction of Working Life. Basic Books, New York (1990)
8. Rosen, P.H.: Psychische Gesundheit in der Arbeitswelt - Handlungs- und Entscheidungsspielraum, Aufgabenvariabilität. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund (2016)
9. Bergstrom, G., et al.: Risk factors for new episodes of sick leave due to neck or back pain in a working population. A prospective study with an 18-month and a three-year follow-up. Occup. Environ. Med. **64**(4), 279–287 (2007)
10. MacDonald, L.A., et al.: Covariation between workplace physical and psychosocial stressors: evidence and implications for occupational health research and prevention. Ergonomics **44**(7), 696–718 (2001)
11. Leclerc, A., et al.: Carpal tunnel syndrome and work organisation in repetitive work: a cross sectional study in France. Occup. Environ. Med. **55**(3), 180–187 (1998)
12. Leclerc, A., et al.: Upper-limb disorders in repetitive work. Scand. J. Work Environ. Health **27**(4), 268–278 (2001)
13. Wall, T.D., et al.: The demands-control model of job strain: a more specific test. J. Occup. Org. Psychol. **69**(2), 153–166 (1996)
14. Mullarkey, S., et al.: The impact of technology characteristics and job control on worker mental health. J. Org. Behav. **18**(5), 471–489 (1997)
15. Dvash, A., Mannheim, B.: Technological coupling, job characteristics and operators' well-being as moderated by desirability of control. Behav. Inf. Technol. **20**(3), 225–236 (2010)
16. Inoue, A., et al.: Organizational justice, psychological distress, and work engagement in Japanese workers. Int. Arch. Occup. Environ. Health **83**(1), 29–38 (2010)
17. Alipour, A., et al.: Four-year incidence of sick leave because of neck and shoulder pain and its association with work and lifestyle. Spine (Phila Pa 1976) **34**(4), 413–418 (2009)
18. Ghaffari, M., et al.: Low back pain among Iranian industrial workers. Occup. Med. **56**(7), 455–460 (2006)
19. Alipour, A., et al.: Occupational neck and shoulder pain among automobile manufacturing workers in Iran. Am. J. Ind. Med. **51**(5), 372–379 (2008)
20. Hsieh, A.-T., Chao, H.-Y.: A reassessment of the relationship between job specialization, job rotation and job burnout: example of Taiwan's high-technology industry. Int. J. Human Resour. Manag. **15**(6), 1108–1123 (2004)
21. Melamed, S., et al.: Objective and subjective work monotony: effects on job satisfaction, psychological distress, and absenteeism in blue-collar workers—correction. J. Appl. Psychol. **80**(4), 538 (1995)
22. Shimazu, A., Shimazu, M., Odara, T.: Divergent effects of active coping on psychological distress in the context of the job demands-control-support model: the roles of job control and social support. Int. J. Behav. Med. **12**(3), 192–198 (2005)
23. Parker, S.K., Sprigg, C.A.: Minimizing strain and maximizing learning: the role of job demands, job control, and proactive personality. J. Appl. Psychol. **84**(6), 925–939 (1999)
24. Bonde, J.P., et al.: Understanding work related musculoskeletal pain: does repetitive work cause stress symptoms? Occup. Environ. Med. **62**(1), 41–48 (2005)
25. Dwyer, D.J., Ganster, D.C.: The effects of job demands and control on employee attendance and satisfaction. J. Org. Behav. **12**(7), 595–608 (1991)
26. Holman, D., et al.: Job design and the employee innovation process: the mediating role of learning strategies. J. Bus. Psychol. **27**(2), 177–191 (2012)

27. Jorgensen, M., et al.: Characteristics of job rotation in the Midwest US manufacturing sector. Ergonomics **48**(15), 1721–1733 (2005)
28. Frieling, E., Buch, M., Weichel, J.: Ältere Beschäftigte in gewerblich-industriellen Tätigkeiten – ausgewählte Ergebnisse und Handlungsfelder am Beispiel der Montage. Wirtschaftspsychologie **3**, 120–128 (2008)
29. Jackson, P.R., et al.: New measures of job control, cognitive demand, and production responsibility. J. Appl. Psychol. **78**(5), 753–762 (1993)
30. Rohrbach-Schmidt, D., Hall, A.: BIBB/BAuA Employment Survey 2012. F.I.f.V.E.a. Training, Editor, Bonn (2012)
31. Ulam, P., et al.: Integrated mission specification and task allocation for robot teams-design and implementation. In: Proceedings 2007 IEEE International Conference on Robotics and Automation. IEEE (2007)

Human Factors in Emergency Management

Interlinking Standardized OpenStreetMap Data and Citizen Science Data in the OpenData Cloud

Werner Leyh[1(✉)] and Homero Fonseca Filho[2]

[1] Department of Computer Science, University of São Paulo (USP), São Paulo, Brazil
WernerLeyh@yahoo.com
[2] Environmental Management, School of Arts, Sciences and Humanities, University of São Paulo (USP), São Paulo, SP, Brazil
homeroff@usp.br

Abstract. The aim of this work is to explore the opportunities offered by semantic standardization to interlink primary "spatial data" (GI) from "OpenStreetMap" (OSM) with repositories of the "Linked Open Data Cloud" (LOD). Research in natural sciences can generate vast amounts of spatial data, where Wikidata could be considered as the central hub between more detailed natural science hubs on the spatial semantic web. Wikidata is a world readable and writable community-driven knowledge base. It offers the opportunity to collaboratively construct an open access knowledge graph that spans biology, medicine, and all other domains of knowledge. In this study we discuss the opportunities and challenges provided by exploring Wikidata as a central integration facility by interlink it with OSM, a popular, community drives collection of free geographic data. This is empowered by the reuse of terms and properties from commonly understood controlled vocabularies that represent their respective well-identified knowledge domains.

Keywords: Metadata · Dublin Core Standard (DwC) · Semantic web · Semantic interoperability · Namespaces · Wikidata · OpenStreetMap (OSM) · Spatial Data Infrastructure (SDI) · Citizen Science (CS) · Ontology · Controlled vocabulary

1 Introduction

A General Overview. In this paper, we use the semantic web technology, notably RDF, SPARQL and Linked Open Data, in the context of scientific data sharing. The goal is to facilitate the discovery and exploitation of datasets by volunteered "citizen scientists" and related communities. This mainly involves relying on the use of Wikipedia for specifying the semantics of datasets, and the use of the identifiers employed by Wikidata (the Wikipedia knowledge base) for annotating these datasets and thus facilitating their discovery [1].

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_9

1.1 Related Work

Data Sharing and Reuse: Interoperability and Standardization. The Semantic Web offers the prospect of facilitating data sharing and reuse. New challenges have emerged from the huge increase in data production capacity by the wide range of experimental devices [1]. Owing to the large amount of data, sharing has to rely on computers. This may prove to be complicated because of interoperability concerns between different protocols and software [1].

According to Rafes and Germain [1] the wide range of data formats reflects the history of scientific data: "The obvious solution is standardization" [1].

The World Wide Web Consortium (W3C), recommends Linked Data for publishing Web metadata. Linked Data is based on the following concepts: (I) Conceptual things that are unambiguously denoted by universal resource identifiers (URI). (II) URIs that are identifiable on the Web and denote objects that can be referred to and looked up ("de referenced") by people and user agents. (III) The fact that when a URI is looked up, its "semantics" includes links to other related things. The semantics is represented in a standard format, namely Resource Description Framework (RDF) [1].

The use of RDF makes it possible to employ the SPARQL query language [1].

OpenStreetMap, API's and LinkedGeoData. The crowd sourced project (OpenStreetMap) is playing a key role in cyber-geography due to its openness, ease of use and capacity for integration in third-party applications [2]. OSM is a geographical database that is maintained by Web users and contains a huge amount of data that can also be displayed on a map. Its database is updated every 15 min and as of today it contains over 5,000,000,000 GPS points and also more than 2,500,000 users who contribute to the project [3].

All OSM data is either available via weekly data dumps or is query able through a public API's [3]. It is easy to see that having this kind of data available in the Linked Open Data cloud must enhance the amount and quality of the information available within the so- called Web of Data [3]. According to Anelli et al. [4] this is the rationale behind the Linked Geo Data project [2]. It aims at triplifying OpenStreetMap dumps every six months by mapping OSM TAGs and OSM KEY properties with reference to a publicly available ontology. This is a very useful resource because it makes available classes that map KEYs and TAGs used in OpenStreetMap nodes [3].

In the last few years, we have witnessed the spread of knowledge intensive applications that rely on the flourishing of the datasets available in the Linked Data (LD) Cloud. The richness of semantic data that they expose, paves the way for a new generation of services and tools that can exploit the ontological knowledge they encode. As well as this, they can make it easier to mash up data that comes from different sources [3].

The availability of a common query interface i.e., SPARQL, and the crowd-driven standardization of ontological vocabularies, allows an intelligent application to grab data from diverse datasets and join them together. One can imagine different scenarios where this kind of feature can be an important asset when providing a high-quality service such as context-aware recommendation systems, on-line shopping, etc. or public services needed in situations of disaster management where the quality and timeliness of the data is of crucial importance [3].

Wikidata, a Fully Open Public Knowledge Graph. Wikidata is the community-created knowledge base of Wikipedia, and the central data management platform for Wikipedia and most of its sister projects. Since its public launch in late 2012, the site has gathered data on more than 15 million entities, including over 34 million statements, and over 80 million labels and descriptions in more than 350 languages. One reason for this strong community participation is the tight integration with Wikipedia: as of today, almost every Wikipedia page in every language incorporates content from Wikidata [5].

According to Good et al. [6], the Creative Commons license of Wikidata, its Semantic Web compatible implementation and the active community provides a unique opportunity to assemble and disseminate knowledge and is currently the only major Semantic Web resource that supports open, collaborative editing. With regard to distributing knowledge, its direct integration with the Wikipedias can allow its content (which is vetted by the community) to be shared with literally millions of consumers in hundreds of languages. Since it has a truly open access status and its implementation is standard compliant, it could become the central component of the long promised Semantic Web in the life sciences [6].

Like Wikipedia, Wikidata is arranged in pages, and this is also how the data is structured. Every subject on which Wikidata has structured data is called an entity, and every entity has a page. So far, the system has distinguished two types of entities: items and properties. In the familiar terms of semantic technologies, items represent individuals and classes, and Wikidata properties resemble RDF properties [5].

The title of any Wikidata page is “Q42” rather than “Douglas Adams” since Wikidata is a multi-lingual site. For this reason, items are not identified by a label in a specific language, but by an opaque item identifier, which is assigned automatically when creating the item and which cannot be changed later on [5]. Every Wikidata entity is identified by a unique URI (such as http://www.Wikidata.org/entity/Q42 for item Q42, Douglas Adams). Item identifiers always start with “Q” followed by a number. By resolving this URI, the tools are able to obtain item data in the requested format (through content negotiation). This complies with Linked Data standards for data publication, and makes Wikidata a part of the Semantic Web while supporting the integration of other semantic Web data sources with Wikidata [7].

Connecting Wikidata to the Linked Data Web. A key characteristic of linked open data is the interconnection of datasets. Wikidata, too, makes many connections to external datasets from many domains, ranging from international authority files, such as ISSN or VIAF, to highly specialized databases [5]. In spite of this, the same authors stated in the same work [5] that they only found 17 widely used Wikidata properties in which direct links were established with other RDF datasets. They argue that their main goal is to generate RDF exports that faithfully represent the original data by appropriately using the language of RDF and linked data and that they restrict direct links to target datasets for which Wikidata has a property [5].

In addition to its function as a structured data store for Wikimedia projects, Wikidata is being used to integrate and disseminate knowledge [6]. Wikidata provides links to concept unique identifiers in a growing number of controlled vocabularies and ontologies, thus making integration easier to achieve with and between the existing knowledge bases [6].

The Five Stars of Linked Data Vocabulary Use. Janowicz et al. take [8] (the original proposal from Berners-Lee's [9]) to the next stage by introducing a 5-star rating for Linked Data vocabulary use.

In 2010, Berners-Lee introduced a 5-star rating for Linked (Open) Data [9]. The first star is assigned to the key initial stage of making the data available on the Web. All the following stars are intended to make the data easier to discover, use, and understand. The second star is assigned to making the data available in a machine readable, structured way. The third star is for using non-proprietary formats. The fourth star is for using open W3C standards such as RDF to identify resources. Finally, the fifth star is for linking one's own data to other datasets [9].

Janowicz et al. [8] argue that Berners-Lee's 5 star Linked Data [9] is only a necessary precondition for what we really need: "Just converting a CSV file to a set of RDF triples and linking them to another set of triples does not necessarily make the data more (re-)usable to humans or machines".

Janowicz et al. [8] point out that "Interestingly, the original 5 star rating does not make any assumptions about the use of vocabularies" and argue that: "In practice, however, querying Linked Data that do not refer to a vocabulary is difficult and understanding whether the results reflect the intended query is almost impossible". The authors underline their views by introducing 5 stars for Linked (Open) Data vocabulary use to encourage data owners, engineers, and practitioners to publish and use vocabularies on the Web [8].

Scaling OpenStreetMap with Wikidata Knowledge. There are ongoing attempts by the Community to bring two of the largest open data projects closer together - OpenStreetMap and Wikidata. Wikidata is the free knowledge base that anyone can edit. It has grown to include over 24 million entries of structured data. By linking this data to the 18 million uniquely - named geographic features mapped with street-level accuracy for the OpenStreetMap project, with the assistance of volunteers, the result may be the largest spatial map of human knowledge ever created. (For example, to understand how we can leverage the combined data of both projects, [10] cites the example of Manali, a small town in the Himalayas. The properties shown for the feature on OpenStreetMap, describe it as a town with a small population, along with its name translated into five languages. When we link Manali to its corresponding Wikidata item, we are given access to much more information: its name in 28 languages, a collection of freely licensed images of the town from Wikimedia Commons, a travel guide from Wikivoyage and encyclopedia articles about the town in 26 languages from Wikipedia.) [10].

1.2 The Objective of the Present Study: A Contribution to Integrate "Crowd-Sourced" and "Official" Knowledge

The Objective of the Present Study. This research study seeks to enhance semantic interoperability between the contributions to authoritative knowledge made by repositories and community-based Wikimedia and OSM-platforms.

This is carried out through the reuse of terms and properties from existing, broadly accepted, standardized vocabularies like "W3C PROV-O" and "Schema.org" which are designed to represent and describe the respective areas of interest.

The work can be regarded as a "Citizen Science - oriented" response to the "Five Stars of Linked Data Vocabulary Use" system put forward by Janowicz et al. in [8], who point out that "Querying Linked Data that do not refer to a vocabulary is difficult and understanding whether the results reflect the intended query is almost impossible".

Contribution of the Present Work. Good et al. outline in a few of the opportunities and associated challenges that Wikidata provides to the broad biocuration community:

(Opportunity I-A:) As a fully open public knowledge graph;
(Opportunity I-B:) As a shared concept resource for information extraction;
(Challenge II-A:) Community ontology building;
(Challenge II-B:) Establishing computable trust;
(Challenge II-C:) Building up Wikidata through text mining;

This will be applied in our present work to improve interoperability between "official" and "community" based knowledge, by reusing of controlled vocabularies.

1.3 Structure and Organization of the Paper

The remainder of the paper is structured as follows. Section 1 discusses relevant related work, while Sect. 2 describes in detail the overall approaches and methodology that are employed. In an attempt to clarify the benefits of the proposal, Sect. 3 analyzes the results and provides illustrative examples from a case study. Section 4 discusses these together with the related benefits, challenges and limitations arising from the study. The paper provides a summary of its conclusions in Sect. 5.

2 Methodology

2.1 General Approaches

The following approaches are adopted:

(MA1) "The Five stars of Linked Data Vocabulary Use" System. In general terms, we are following the "Five stars of Linked Data vocabulary use" system of Janowicz et al. [6] who argue that "In practice, however, querying Linked Data that do not refer to a vocabulary is difficult and understanding whether the results reflect the intended query is almost impossible".

(MA2) Literature Review. We explored the literature which is summarized in the "Introduction" chapter;

(MA3) Wikidata Query Service (WDQS). We explored the Wikidata Query Service (WDQS); this is a software package and public service designed to provide a SPARQL endpoint which allows queries to the Wikidata dataset. WDQS operates on a dataset from Wikidata.org, represented in RDF, as described in the RDF dump file format. We explored this community resource as a means of strengthening the argument in this study (Chapter "Results").

(MA4) Mapping Against Opportunities and Challenges of Wikidata. Good et al. [6] suggested that Wikidata offers two opportunities and raises three challenges. We mapped our contributions in accordance with two opportunities (I-A, B, C) and three associated challenges (II-A, B, C) of Wikidata, suggested by Good et al. in [6] for the broad biocuration community: With regard to their suggestions, one of our objectives was to improve interoperability between "official" and "community"- based knowledge (contributions), by exploring the reuse of controlled vocabularies.

(MA5) We Explored the So-called Talk-Pages in Community-Wiki. Both the Wikidata and Wikipedia platforms are Wikimedia (Foundation) projects. Wikipedia is often described as a wiki, but it is in fact a collection of wikis. The Media Wiki software empowers OpenStreetMap as well as Wikidata and Wikipedia. We fully explored the corresponding "Talk" pages and their comments for the present work and discussion, e.g.:

- https://www.Wikidata.org/wiki/Property_talk:P107 and
- http://wiki.OpenStreetMap.org/wiki/Talk:Key:Wikidata

2.2 Delimitation of the Present Work

In this study, we explore opportunities and associated challenges that are provided by interlinking OpenStreetMap with Wikidata, that is frequently considered as a central integration facility since it is empowered by the reuse of terms and properties from commonly understood controlled vocabularies to represent and describe the respective domain of interest.

3 Results

3.1 Results: Opportunities and Challenges: Vocabulary-Related Factors When Interlinking Wikidata and the OpenStreetMap Application

In this chapter, we represent the results of our study by employing the methods, approaches and protocols listed in Sect. 2. In this way, we are able to apply our own results ("Vocabulary related factors when interlinking Wikidata and Open-StreetMap applications") to the Opportunities (I-A, B) and Challenges (II-A, B, C) of Wikidata suggested by Good et al. [6];

Table 1. Information gathering (queries) - interlinking Wikidata and OpenStreetMap

n	Platform (Interlinked)	Target (examples)	Query (SPARQL)	References
1.	Metadata Standards in WikiData	WHO's ICD10 classification (property P494) with language-specific labels (pt-br)	http://tinyurl.com/l6hyaz6	[6, 11, 12]
2.	Link (query) from WikiData to OSM	World-Map of hospitals	http://tinyurl.com/k46vxg2	[5, 13, 14], Table 1
3.	Link (query) from WikiData to OSM	Hospitals near you	http://tinyurl.com/lojdguk	[5, 13, 14]
4.	Link from OSM to WikiData	https://www.wikidata.org/wiki/Q50071?uselang=en-US https://www.openstreetmap.org/node/2703868060	https://overpass-turbo.eu/?w=%22wikidata%22=%22Q50071%22+and+type:node+global&R	[3, 14], Sect. 4

(Opportunity I-A). "Because of its truly open access status and its implementation of compliant standards, it could become the central component of the long promised Semantic Web in the life sciences".

As Regards "Wikidata": The fluent machine-to-machine communication in the Semantic Web depends on the ability of different IT systems to map different terms to a shared meaning (compare Table 1/n2 + n3; Fig. 1). Using a common vocabulary for representing and describing knowledge graph data allows to represent resources and relations between resources in a unified way. This increases the interoperability and allows a comfortable data integration [15]. Furthermore, the use of a URI instead of plain text, is particularly applicable to situations where the value of the element derives from a controlled vocabulary [16].

On the one hand, a comparative study of Färber et al. in [15] reveals a high extent to which external vocabulary is used in Wikidata.

On the other hand, considering the extent to which URIs of proprietary vocabulary are linked to external vocabulary via equivalence relations, the same study [15] reports that "only 0.1% of all Wikidata classes are linked to equivalent external classes" and "regarding the relations, Wikidata provides links in particular to FOAF and schema.org and achieves here a linking coverage of 2.1%." (e.g. the SPARQL query in Table 1/n1 uses Schema.org concepts).

Wikidata uses a uniform scheme for URIs of all entities, i.e., items and properties, such as Q42 or P184. These URIs follow linked data standards [5]. This IDs can also be used for connections between OSM and Wikidata.

As Regards "OpenStreetMap": OSM disposes over many different API's. The Overpass-API forms a particular powerful means of integrating OpenStreetMap and

Wikidata using OSM's literal KEY + VALUE pairs (compare Table 1/n4 and [17]): (TagInfo-API: https://taginfo.openstreetmap.org/tags/wikidata=Q50071?filter=nodes), and (OverPass-API: https://overpass-turbo.eu/?w="wikidata"="Q50071"+and+type:node+global&R).

It should be highlighted that both query examples, TagInfo and Overpass, are promoted (e.g. as regards precision and stability) by the fact, that Wikidata uses URIs, for all its items as per linked data standards (compare Sect. 3).

The following "Talk" page discusses similar questions such as WikData's use of "Controlled Vocabularies": (http://wiki.openstreetmap.org/wiki/Talk:Wikidata). There we can find very "significant" statements like "Adding these tags is already pretty much accepted in OSM, so you wouldn't have to do any persuasion on our side." (please compare [18]).

(Opportunity I-B). "Apart from its use as a knowledge graph, Wikidata could be of great value to the text-mining community as a multi-lingual collection of concept labels, descriptions, and links for encyclopedic text. The so-called 'items' in Wikidata are roughly analogous to the concepts in the Unified Medical Language System (UMLS)."

As Regards "Wikidata": Table 1/n1 shows a simple example how to explore controlled vocabularies like Unified Medical Language System (UMLS) and WHO's diseases classification ICD (compare https://en.wikipedia.org/wiki/ICD-10) by using Wikidata's Query Interface.

The problem with doing free text search with Wikidata is that it does not have yet a free text index. Without an index, the text search requires trying out a match for each label, which is not an efficient method. Currently, Wikidata's Query Service has limited capabilities to search through the data, due to the absence of support for the FTS engine (compare https://phabricator.wikimedia.org/T141813).

As Regards "OpenStreetMap": The OSM database can be accessed in various ways, including through the Overpass API. It provides a programmatic ¨read-only¨ access to the database and query possibilities by search criteria, e.g., location, type of objects, TAGs. (In any case, the process of text mining necessarily includes the mapping and linking of collection-specific vocabularies with a controlled vocabulary on the Web. Text mining will therefore always be facilitated through the use of controlled vocabularies.) [14].

(Challenge II-A). "In comparison with other attempts to build knowledge graphs on a large scale, the Wikidata approach is on the chaotic side. There is no strict application of an upper ontology, no automated reasoning to support class inference or quality control, and no over-arching plan to govern the evolutionary pattern of the system. So far, good progress has been made. That being said, there is a clear need for experienced ontologists to join the conversation and help to collaboratively guide this community forward if it is to reach its full potential."

These questions regarding volunteered citizen scientists and their motivation and preparation have already been broadly discussed in different channels like scientific literature, blogs, lists and forums, among others (e.g. [18]).

As Regards "Wikidata": The following "Talk" page discusses similar questions such as WikData's use of "Controlled Vocabularies": (https://www.Wikidata.org/wiki/Property_talk:P107). There we can find very "significant" statements like "Unfortunately the whole discussion (see above) is about how we can destroy this property and avoid controlled vocabulary".

As Regards "OpenStreetMap": In [18], the authors discuss very similar challenges. It is worth noting that some of the Talk-page-participants are volunteers from both communities: Wikidata and OpenStreetMap.

(Challenge II-B). "A key enabling feature of the Wikidata infrastructure is the capacity to provide provenance for its claims (the triples that compose the knowledge graph) through references. Each claim can be supported by any number of references that can back up sources of information. Unfortunately, many of the claims that are currently in Wikidata were not assigned references."

As Regards "Wikidata": The fulfillment of trustworthiness at a statement level, is determined by an assessment of whether a provenance vocabulary is used [15]. The source of the statements can be stored by means of a provenance vocabulary. The most widely used ontologies for storing provenance information are the Dublin Core Metadata terms and the W3C PROV-O ontology [15]. When relating statements to references, Wikidata use the property was Derived from the W3C PROV-O Ontology [5], (e.g. the following SPARQL-query from Wikidata's Query Service provides an example: http://tinyurl.com/z82uvou).

As Regards "OpenStreetMap": Again, in [18], the authors discuss very similar challenges with regard to OSM TAG's. Whereas volunteers discuss ¨properties¨ in Wikipedia, their colleagues discuss OpenStreetMap "TAGs" (please compare this with the discussion in Sect. 4).

(Challenge II-C). "The majority of the world's biomedical knowledge remains locked up in unstructured text. As text mining matures, it is increasingly possible to extract this knowledge automatically; however (1) most people, even within the bioinformatics community, do not have the skills and resources to undertake this work.".

The idea of "text mining" has a great potential because it is able to explore different opportunities at the same time. As explained by Good et al. [6], mature technologies are already available for "text mining" and most "knowledge remains locked up in unstructured text". But there are further reasons. For example, many controlled vocabularies are still being explored in legacy (e.g. text-based) representation and are in transition to LOD enablement (e.g. RDF-based representations [19]).

As Regards "Wikidata": Note-worthy is that the Wikidata SPARQL endpoint has a maximum execution time per query of 30 s. This might be a bottleneck for some (e.g. "text mining") queries. Please compare item (Opportunity B [4]).

As Regards "OpenStreetMap": Both OpenStreetMap and Wikidata are very free, with access to open data and have many free tools. This means that they can be "mined" at many "different" levels. A very attractive solution is the use of online queries with web-tools like TagInfo and OverPass (see Table 1/n4 and Sect. 4).

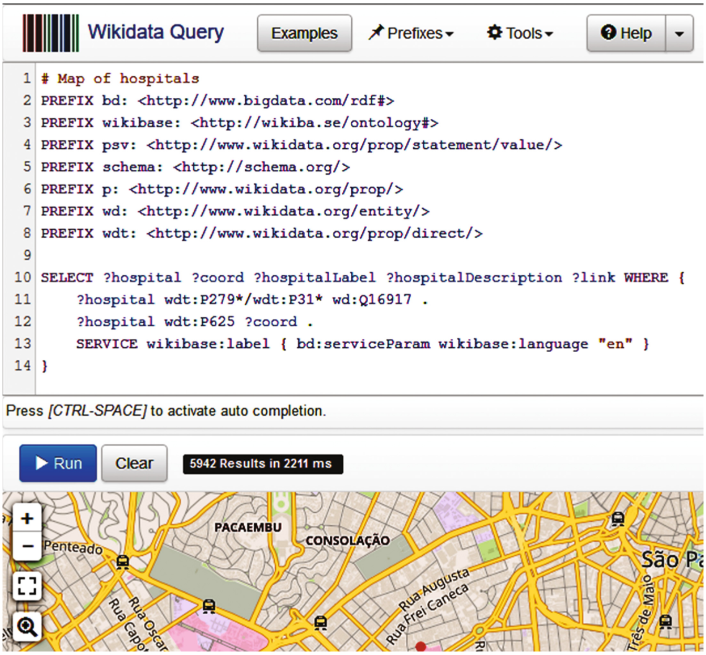


Fig. 1. Information gathering (queries) - interlinking Wikidata and Openstreetmap, Table 1/n2: World-Map of hospitals (http://tinyurl.com/k46vxg2)

4 Discussion

In addition to encyclopedia articles and software, peer production communities produce structured data, e.g., Wikidata and OpenStreetMap's metadata [18].

Peer production communities such as Wikipedia and OpenStreetMap (OSM) are among the most successful applications of social computing [18].

One major factor in the success of these communities is their widespread adoption of a "be bold" contribution principle. This principle encourages contributors to quickly "make changes as they see fit" and elevates contributor freedom to the status of a core community value [18].

However, this structured data is usable by applications only if it follows standards [18]. For Wikidata and OSM to support computing applications most effectively, their structured data must have a high degree of standardization [18].

Own of the mandatory design principles, advocated by inventor of the World Wide Web and the creator of the Semantic Web and Linked Data, Tim Berners-Lee is the use of "Use URI as names for things" [9]:

The Uniform Resource Identifier (URI) is a single global identification, a kind of unique ID, for all things linked, so that we can distinguish between those things, integrate them without confusion, or know that one thing from one dataset is the same as another in a different dataset because they have one and the same URI.

Promoting Quality Data. Wikidata IDs are also more stable then OSM ID's in the sense that they will not be reused for something else: Once an item has been deleted, that ID is not used again (which allows us to spot problematic links easier.

Connecting Wikidata and OpenStreetMap can be a very useful feature for users of both databases. In general, we can use Wikidata links to express and disambiguate information about OSM objects (such as the sculptor and subject of a statue).

One of the immediate benefits of this effort is that it is now feasible to create world maps with translated and transliterated place names in many more languages [10].

So if OSM-volunteers link all cities to Wikidata, anyone could do complex queries. As such it can have various attributes, like the name of the mayor (and a link to the entry about that person, if there is one), names in other languages, population, links to WikiVoyage and images on Wikimedia Commons, and much more besides.

OSM's Permanent ID Challenge. What if someone wants to bind (in Wikidata) to a specific object in OpenStreetMap? Linking from OpenStreetMap "directly" to Wikipedia articles is not future-proof, because articles are sometimes renamed, changing their URLs.

Nevertheless, as shown e.g. in Table 1/n2, n3, n4, Wikidata can already link to a point, using coordinates or OSM's wikidata tag (querying for OSM TAGs using Wiki-data keys).

5 Conclusion

This study shows how standardized Wikidata and OpenStreetMap interlinking can lead to LOD-integration, particularly when the capacity of Wikidata is explored as a powerful integrator with the Semantic Web. The study provides a description of conceptual considerations as well as the Wikidata-based integration of "primary community data" with the "official" Linked Open Data.

References

1. Rafes, K., Germain, C.: A platform for scientific data sharing to cite this version. In: Proceedings of the BDA 2015 Conference, Ile de Porquerolles, France, September 29–October 2, 2015 (2015)
2. Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M.M.: A review of volunteered geographic information quality assessment methods. Int. J. Geogr. Inf. Sci. 1–29 (2016)
3. Anelli, V.W., Calì, A., Di Noia, T., Palmonari, M., Ragone, A.: Exposing open street map in the linked data cloud. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 344–355 (2016)
4. Anelli, V.W., Di Noia, T., Galeone, P., Nocera, F., Rosati, J., Tomeo, P., Di Sciascio, E.: LOSM: a SPARQL endpoint to query Open Street Map. In: CEUR Workshop Proceedings, pp. 1–4 (2015)

5. Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., Vrandečić, D.: Introducing Wikidata to the linked data web. In: Mika, P., Tudorache, T., Bernstein, A., Welty, C., Knoblock, C., Vrandečić, D., Groth, P., Noy, N., Janowicz, K., Goble, C. (eds.) Proceedings of 13th International Semantic Web Conference, Riva del Garda, Italy, 19–23 October 2014, Part I, pp. 50–65 (2014). http://link.springer.com/book/10.1007/978-3-319-11964-9
6. Good, B., Burgstaller-Muehlbacher, S., Mitraka, E., Putman, T., Su, A., Waagmeester, A.: Opportunities and challenges presented by Wikidata in the context of biocuration. In: Jaiswal, P., Hoehndorf, R., Arighi, C., Meier, A. (eds.) Proceedings of the Joint International Conference on Biological Ontology and BioCreative, Corvallis, Oregon, United States, vol. 1747, 1–4 August 2016. CEUR-ws.org (2016)
7. Vrandečić, D., Krötzsch, M.: Wikidata. Commun. ACM **57**, 78–85 (2014)
8. Janowicz, K., Hitzler, P., Adams, B., Kolas, D., Vardeman, C.: Five stars of linked data vocabulary use (2014). http://dase.cs.wright.edu/publications/five-stars-linked-data-vocabulary-use
9. Berners-lee, T.: Linked data: design issue. https://www.w3.org/DesignIssues/LinkedData.html
10. Ganesh, A. (Mapbox): Scaling OpenStreetMap with Wikidata knowledge. https://www.mapbox.com/blog/scaling-openstreetmap-wikidata-knowledge/
11. Przybyła, P., Shardlow, M.: Interoperability Landscaping Report (EU, OpenMinTeD, D5.1) (2016)
12. Nielsen, F.: Danish resources. (2017)
13. Page, R.: Towards a biodiversity knowledge graph. Res. Ideas Outcomes **2**, e8767 (2016)
14. Meyer, M. (FOSSGIS e.V.): OpenStreetMap und Wikidata. In: FOSSGIS (eds.) Anwenderkonferenz für Freie und Open Source Software für Geoinformationssysteme Salzburg 04–06 Juli, Salzburg, p. 25 (2016). www.mapmedia.de
15. Färber, M., Ell, B., Menne, C., Rettinger, A., Bartscherer, F.: Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semant. Web J. (2016). (by IOS Press. 1366–2578, 01–53)
16. Godby, C.J., Smith-Yoshimura, K.: From records to things: managing the transition from legacy library metadata to linked data. Bull. Assoc. Inf. Sci. Technol. **43**, 18–23 (2017)
17. Openstreetmap Community: OSM's Overpass Application. http://wiki.openstreetmap.org/wiki/Overpass_API
18. Hall, A., Mcroberts, S., Thebault-spieker, J., Lin, A.Y., Sen, S., Hecht, B., Terveen, L.: Freedom versus standardization: structured data generation in a peer production community. In: Proceedings of the 35th Annual ACM Conference on Human Factors in Computing Systems (CHI 2017). (2017) https://chi2017.acm.org/organizers.html
19. Biodiversity Information Standards - TDWG: Darwin Core RDF Guide Introduction References Quick Reference Guide Simple Darwin Core RDF Guide Table of Contents Introduction Implementation Guide Term reference Text Guide XML Guide Namespace Policy Complete History Decision History Mapping t. http://rs.tdwg.org/dwc/terms/guides/rdf/

Empirical Studies in User Experience of an Emergency Management System

Mário Simões-Marques[1(✉)], Anacleto Correia[1],
M. Filomena Teodoro[1,2], and Isabel L. Nunes[3,4]

[1] CINAV – Portuguese Navy, Alfeite, 2810-001 Almada, Portugal
mj.simoes.marques@gmail.com,
{cortez.correia,maria.alves.teodoro}@marinha.pt
[2] CEMAT, Instituto Superior Técnico, Universidade de Lisboa,
1048-001 Lisbon, Portugal
[3] Faculty of Sciences and Technology, Universidade NOVA de Lisboa,
2829-516 Caparica, Portugal
imn@fct.unl.pt
[4] UNIDEMI, Department of Mechanical and Industrial Engineering,
Faculty of Sciences and Technology, Universidade NOVA de Lisboa,
2829-516 Caparica, Portugal

Abstract. A system designed following usability principles, and considering users preferences and expectations ensures a high quality of user experience, which is particularly important when the system is used to support complex tasks. That is the case of emergency response operations' coordination and execution. This work, conducted using an action research methodology, addresses challenges faced by the THEMIS project team during the conceptualization, design and implementation of an emergency management intelligent system, aiming to support disaster response operations and to improve the planning and execution efforts in disaster and crisis' scenarios.

Keywords: User Centered Design · Human-systems interaction · Personas · THEMIS

1 Introduction

Information systems, if well-conceived and designed, can help to more easily face and mitigate disaster and crisis' situations. An emergency management system is a specific kind of information system suited to support complex processes regarding the coordination of operations in response to several types of emergencies. The response to each of type emergency (e.g., fire, earthquake, flood, etc.) shares most of the coordination methods, but may require specific procedures and information to conduct the emergency operations.

This work, conducted using an action research methodology, addresses challenges faced by the THEMIS project team during the conceptualization, design and implementation of an emergency management intelligent system, aiming to support disaster response operations and to improve the planning and execution efforts in crisis'

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_10

scenarios. In the context of the emergency management intelligent system's life cycle, several surveys and empirical studies were planned in order to fine-tune the subsequent implementation considering the three factors that influence user experience: system, user and the context of use.

User Experience (UX) can be defined as the operator's perceptions and responses regarding the use of the system. According to the ISO 9241-210 definition, user experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after system use [1]. The goals of UX in the context of emergency management systems are conceptually similar to those of any other software application. However, facets such as the specificity and the complexity of the context of use, the variety of roles, or the criticality of the requirements, to name just a few, present some challenges in the analysis and assessment of UX that somehow differ from most applications addressed by literature (e.g., web portals).

This study helped to contextualize the emergency management intelligent system for each participant in the operations, since: (i) emergency operators, when carrying their tasks, rely on quick and easy system interactions, to gather up-to-date information, customized to their current situation; and (ii) the decision-makers require and intuitive, effective and efficient tool to support the complex tasks of operations' coordination.

2 THEMIS Project

THEMIS is the acronym for *disTributed Holistic Emergency Management Intelligent System*. This is an undergoing R&D project funded by the Portuguese Ministry of Defense aimed at supporting real time disaster management activities of decision-makers in coordination centers, and of responders in the field. A high-level conceptual perspective of the system is illustrated in Fig. 1 considering the scenario of a major disaster requiring international assistance, provided by multiple agencies. As shown THEMIS is meant as an intelligent system that accesses information from multiple sources (e.g., users, sensors, crowdsourcing), provides situational awareness based a georeferenced common picture, which is shared among system users. The information about incidents is analyzed to assess response priorities and advise on available resources' assignment. The system also provides support to the tasked response teams, for instance by means of situational awareness, positioning and navigation information, procedural guidance or the access to relevant hypermedia materials. The system functionalities are provided as a service, which will allow exploiting BYOD (bring your own device) and BYOA (bring your own application) approaches [2].

Despite not being a pure Group Decision Support System (GDSS), as defined in [3], the concept behind THEMIS shares many characteristics of the this type of systems, such as: (i) supports the parallel processing of information; (ii) the participation of large groups with more complete information, knowledge and skills; (iii) the use of structured and unstructured techniques and methods; (iv) rapid and easy access to external information; (v) helps participants frame the big picture; (vi) provides multiple ways to participate; (vii) contributes to keep the group on track; and (viii) enables several users to interact simultaneously.

The Portuguese Navy has been conducting operations in emergency scenarios, both as preparedness drills integrated in the operational training cycle (e.g. DISTEX – Disaster Exercise), and in actual humanitarian assistance missions (e.g. in landslides at Madeira island, in a volcanic eruption at Cape Verde islands, or in evacuation of nationals after a coup d'état at Guinea-Bissau). The THEMIS system, as a kind of GDSS, aims to add a decision-support layer to aid the operational activity in this domain.

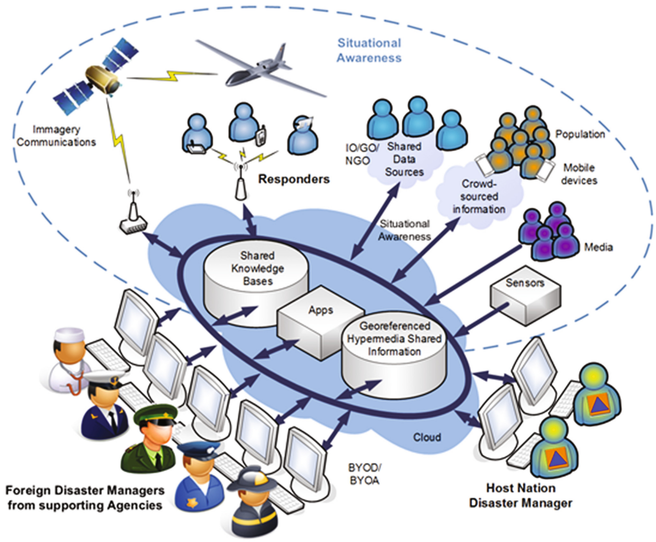


Fig. 1. THEMIS high-level conceptual view

3 Research Design Approach

There are a wide number of different approaches to design research. According to Crotty [4] there are four elements to consider in the research process:

Epistemology: the theory of knowledge which supports the theoretical perspective;
Theoretical perspective: the philosophical support;
Methodology: the plan of action, strategy, selection and application of methods;
Methods: the techniques or procedures to organize and analyze data.

Previous conceptualizations were reexamined by Creswell [5] particularly because this author recognizes a new type of inquiry approach, besides the traditionally assumed quantitative and qualitative (increasingly important in late 20th century). This new approach is the mixed approach, which was defined based on the observation that nowadays the research practices are somewhere between a quantitative and a qualitative approach.

Considering the three basic elements of research design referred by Creswell – knowledge claims, strategies of inquiry, and methods (data collection, treatment, analysis and compilation) – THEMIS follows a mixed methods pragmatic[1] research approach since this opens the door to multiple methods, different worldviews, and different assumptions, as well as different forms of data collection and analysis [5].

In fact, examples of methodologies and respective methods used in THEMIS implementation include:

- for **system design** – User Experience, User Centered Design and Scenario-based design (the last approach is further discussed in [6]);
- for **knowledge management** (e.g., knowledge acquisition, representation, inferencing) – ontologies, unified modeling language (UML), fuzzy inference methods (this issue is further discussed in [7]);
- for **simulation and validation** – probabilistic methods (this issue is further discussed in [8]).

4 User Experience

The intelligent system THEMIS is a very complex and interactive computer-based system. To insure a high quality of user experience when interacting with THEMIS, the design and development follows the principles and best practices underlying Usability and User Experience. These two concepts have overlapping aspects, while Usability refers to the quality of the task-based interactions, UX encompasses the quality of the overall experience perceived by the user before, during and after interacting with a system. In fact, UX expands the focus of Usability to other dimensions of the relationship user-system and considers also a different time span, not only focused on the interaction design solution development. Besides Usability concerns with effectiveness, efficiency and satisfaction of the user interaction, UX brings the perspectives of meaningfulness, added-value, emotional connection, and feelings towards the service conveyed by the application regarding the "business" domain, trying to understand who the users are and what they are hoping to accomplish using THEMIS.

UX addresses the entire life cycle of a system from the early stages, when a need or an innovative opportunity is identified and the system concept is born, through the development, entry in service, support and upgrade cycles until its removal. In fact, according to Kraft [9] the long-term UX is as important as the first impression if users are to use the system for a long time. This author refers that experience can be represented as a UX curve, and that an ideal curve shows an overall trend of positive growth in satisfaction over time, even if there are a few dips.

Along this process UX is concerned with users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments, as previously mentioned [1]. THEMIS development will follow a UX approach, since this

[1] Creswell groups the phylosophical worldviews (i.e. the knowledge claims) as being Postpositivist, Social Constructivist, Advocacy and Participatory, and Pragmatic [5].

is a process of enhancing user satisfaction, by improving the Usability, accessibility and pleasure in the interaction [10].

When developing a system, the UX objectives can be met by using different approaches (e.g. User-Centered Design (UCD), Design Thinking) if on the applied process they ensure means or methods to address both the pragmatic (e.g., effectiveness, efficiency) and hedonic (e.g., satisfaction, pleasure) quality of the system, while trying to cope with user goals regarding the system. Figure 2 offers a conceptual design perspective, as proposed by Bongard-Blanchy and Bouchard, which considers the relationship between user goals, system purpose and context of use, and also identifies several properties (functional, semantic, sensorial, behavioral) which affect the perceived quality of the system [11].

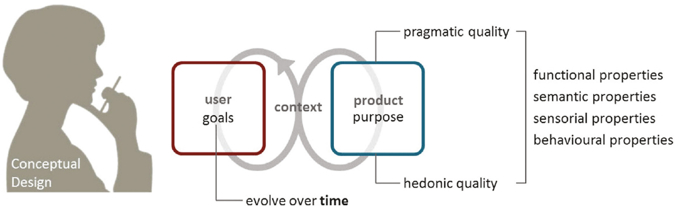


Fig. 2. Dimensions of conceptual design when designing for UX [11].

UCD is an approach consisting of several activities performed along the following main stages [1]: (a) context of use and user needs elicitation; (b) functional requirements specification; (c) system implementation (including prototyping its early steps); and (d) usability testing and evaluation.

4.1 Context of Use and Personas

As mentioned before the context of use for THEMIS are real time disaster management activities performed by decision-makers in coordination centers, and disaster response by responders in the field.

Personas were used to identify users' needs, wants and expectations in order to design best possible experience for them. Personas are fictional characters that represent archetypes[2] (i.e., real users' profiles) [12].

In order to define Personas a list of potential profiles of participants on ground emergency operations was built (see Table 1). The profiles are based on international guidelines defined by the International Search and Rescue Advisory Group (INSARAG) [13] (first and second columns of Table 1). Those 17 profiles were matched with the 17 current roles trained by Portuguese Navy crews in the DISTEX drills. However, the match was not on a one to one basis since there are several

[2] An archetype is a specific characterization of an individual which is representative of a group by opposition to a stereotype which is a very general characterization of the group.

INSARAG profiles filled by the same DISTEX profile, and conversely, some DISTEX profiles do not match with the ones defined by the INSARAG.

The characterization of each Persona was based on interviews to three individuals, chosen because of their previous experience of participation in humanitarian missions and Navy's emergency drills. The interviews, with a duration of about 1 h on average, took place within the users' working hours and at their working place.

An example of script used for a Persona (in this case representing the "Planning Officer/OSOCC Commander") characterization is:

> *"Lieutenant-Commander Silva is 40 years-old, and he is Executive Officer of the ship deployed to a disaster relief operation in a foreign country. He is a dynamic leader, very experienced in naval operations, but this is the first mission of this type in which he was engaged. Nevertheless, he recalls what he learned and the experience acquired in the DISTEX last year. He knows that his goals as Leader of the Coordination center are to get a clear picture about the incidents and the number and location of injured and dislodged people, so that he is able to direct the relief response to assist them. He is aware the resources are very scarce to face all the coming demand, therefore he needs to assess the incidents and set priorities. Another goal is to identify safe and adequate facilities to install a field hospital, to shelter the dislodge people, and to distribute food. Therefore, first he needs that the Search Team covers as much ground as possible, reporting victims, damages and suitable facilities. He observes the THEMIS screen and the first reports are showing up in the map. A bridge is collapsed. This is a problem with all the equipment to carry to the other side of the river. THEMIS advice is to deploy the Engineering Team to this incident, and he agrees. The Engineering Team has to figure a way of solving the problem. He turns to his assistant and orders him to task the team to the incident. Silva knows that in seconds his guys will receive the order in their THEMIS mobile device and start moving to the bridge. What's next?[...]"*

This script excerpt helps understanding the context, behaviors, attitudes, needs, challenges, goals and motivations of the real users.

In our approach the use of Personas is complemented with task scenarios which are addressed below but not described in detail.

Table 1. Comparison of INSARAG and THEMIS roles.

INSARAG role	Main responsibility	THEMIS role
Team Leader	Overall command of strategy, tactics and safety	Ship Captain (onboard)
Planning Officer	Coordination with local incident commander	OSOCC[a] Commander (ashore)
Operations Officer	Internal and external coordination	OSOCC Operations Officer
Structural Engineer	Knowledge of hazards in disaster environments	Hydrographic Team
Rigging Specialist	Knowledge of heavy construction equipment	Engineering Team
Liaison Officer	Participate on joint operations planning	OSOCC Liaison Officer

(*continued*)

Table 1. (*continued*)

INSARAG role	Main responsibility	THEMIS role
Safety Officer	Functional knowledge of hazards	SAR Team Leader
Logistics Manager	Operational planning pertaining to logistics	OSOCC Logistics Officer
Media Officer	Provide assistance to the media	OSOCC Public Affairs Officer
Comm./IT Spec.	Install, operate and maintain communications and IT	Communications Team Leader
Rescue Officer	Operational control of assigned area	SAR[b] Team Leader
Rescue Specialist	Operations in the environment	SAR Team
Hazardous Spec.	Identification of hazardous	-
Search Officer	Search ops including application of grid systems	SAR Team Leader
Search Specialist	Search ops including application of grid systems	SAR Team
Medical Manager	Provide medical input into the decision making	Advanced Medical Site
Medical Spec.	Basic first aid	First aid Team
-	-	Injured and Rescued Control Center
-	-	Displacement Center
-	-	Flight Operations Support Team
-	-	Driver

[a]OSOCC – On-Site Operations Coordination Center
[b]SAR – Search and Rescue

The method adopted to represent Personas' interaction with the system was Use Cases, a technique that is part of the UML standard [14]. In fact, Use Cases were drawn based on the individual Persona's script to highlight the interactions between such Persona (an actor in UML terminology) and the emergency system, to attain a certain functionality and achieve a relevant goal (see Fig. 3). Note that the emergency system was mainly based on manual procedures, sheet maps, and physical signs. Another part of the emergency system was supported by a previously existent software system (depicted in Fig. 4), with georeferenced functionalities. The procedures and signs although being standards in the organization were not aligned with the international guidelines. This hinders the interoperability of the naval emergency teams with rescue teams of other countries or organizations, when engaged in international disaster relief or humanitarian assistance operations.

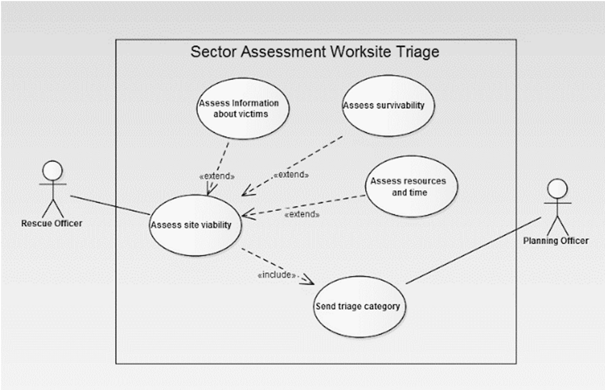


Fig. 3. Use Case diagram for the “Sector Assessment Worksite Triage”

(a) Desktop Interface (b) Mobile interface

Fig. 4. Samples of the user interfaces from the actual system for emergency management

4.2 Functional Requirements Specification

The Personas characterization covered all the INSARAG roles in Table 1. A list of functional requirements for THEMIS was derived from that work.

Observing the DISTEX exercises also helped the requirements team to get a sense of the coordination required among emergency teams, and how Personas relate with each other. This helped also on the building of coherent Use Case representations.

Besides the diagrammatical representation of the Use Case, shown above, a template with a complementary structured textual description of system functionalities was also used to gather the system requirements. In this template, for each Use Case, scenarios put interactions in context, in a kind of dialogue between a ‘Persona/UML actor’ and the system (see Fig. 5). Together with functional requirements, the textual specification of Use Cases also record non-functional requirements reported by the users, regarding namely, to system performance or usability and interface design.

Therefore, these tools help gathering user needs, as well as system functional and non-functional requirements, guiding how to design the interfaces (e.g. what information to present, and how to interact with devices connected to the system, depending on the specific characteristics of equipment interface).

1. Use Case Description
Assess the existence of survivors in a site's debries.
1.1. Name:
Assess site viability
1.2. Goal:
Decide the triage category for a worksite.
1.3. Use Case Team Leader/Members:
Rescue Officer, Rescue Specialist
1.4. Precondition:
Rescue Officer assigned to the sector of the site location.
1.5. Post-condition:
Location classified as Worksite (if there are survivors) or not.
1.6. Constraints/Issues/Risks
Site inaccessibility. No Communications.
1.7. Trigger Event(s);
The Rescue Officer chose a site of the assigned sector for triage.
1.8. Primary Actor;
Rescue Officer

Actor Action	**System Response**
2. Main Path:	
2.1 Rescue Officer provide user/password and submit a log on into system	System acknowledge user credentials
2.2 Rescue Officer select site on map	System acknowledge the geographical coordinates of the site
2.3 **Extension Point:** <<Assess Information about victims>>	Acknowledge of (no) victims is recorded
2.4 **Extension Point:** <<Assess survivability>>	Acknowledge existence of (no) small/big survivable voids is recorded
2.5 **Extension Point:** <<Assess resources and time>>	Record time and resources type and quantity needed
2.5 Include: <<Send triage category>>	Record worksite triage category
3. Alternative Path: No live victms confirmed	
3.1 Rescue Officer provide user/password and submit a log on into system	System acknowledge user credentials
3.2 Rescue Officer select site on map	System acknowledge the geographical coordinates of the site
3.3 Mark site as clear on map	Record site without triage category
9. Non-Functional Requirements	
Devices to be handled with exposure to sunlight.	
UI components manipulated by users wearing protection equipment	

Fig. 5. Textual information about the Use Case "Assess site viability"

This process supported the definition of the following general system requirements:

- provide a common platform for the compilation of emergency situations, integrating information from different sources (e.g., military C2 systems, civil systems, public sources);
- provide a platform for inventorying the resources available for the operation (categorized qualitatively, quantified and geo-referenced);
- provide situational knowledge and support coordination of the various levels; the volume of information and the scope of counseling is customized according to the user profile (e.g., the level and organization to which he/she belongs) and the type of hardware/communications infrastructure used (e.g., terminal capabilities, available bandwidth);
- provide intelligent decision support functions, adequate to the user profile, that assess the situation and recommend priorities of action, based on the current operational context (e.g., disaster type, security environment), and the optimum allocation of resources;

- act as an expert system, providing knowledge repositories and analysis tools relevant to support response to different types of disaster, providing contextual advice, according to the scenario, characteristics of resources and user profile; and
- provide a platform for simulations and training actions in humanitarian and emergency management operations.

5 Way Ahead

The system is at its early steps of the implementation stage, which involves developing and assessing prototype designs. The design of prototypes was preceded by the selection of a suitable prototyping software. For desktop and mobile equipment it was selected the *Justinmind Prototyper* software [15].

Prototypes will be mainly used for defining and testing the design layout and the navigation interactions, since they don't execute real system algorithms. As the solutions are implemented and Beta versions are released, it becomes possible to test the usability and UX associated to the system functionalities.

As referred before, usability looks at users' interaction with the system, i.e. how successful tasks are carried out (pragmatic quality aspects), while UX is a wider concept, describing the whole impact that the interaction with the system has on the user (before, while and after the use), namely in its thoughts, feelings and perceptions (hedonic quality aspects) [16–18].

Therefore, THEMIS prototype design proposals will be evaluated to identify the best possible solutions for later implementation. At the present stage, the project team is establishing the evaluation plan. The current plan considers the use of methods for usability and UX assessment that include inspection methods (Cognitive Walkthrough and Heuristic evaluation) and the User Experience Questionnaire (UEQ) [19], which will be used to evaluate the overall impression a user has once interacts with the system, since this questionnaire covers both pragmatic and hedonic quality aspects.

Later on, the assessment of the THEMIS Beta versions, besides using the above-mentioned methods, will also include some objective (i.e. quantitative) methods to evaluate effectiveness (e.g. number of user errors, number of tasks finished in a given period of time) and efficiency (e.g. number of actions performed, average, maximum and minimum time for performing the activity tasks).

6 Conclusions

A system designed following usability principles, and considering users' preferences and expectations ensures a high quality of UX, which is particularly important when the system is used to support a big variety of complex tasks, such as the ones related with the coordination and execution of emergency response operations.

THEMIS is an undergoing project funded by the Portuguese Ministry of Defense aimed at supporting real time disaster management activities of decision-makers in coordination centers, and of responders in the field. THEMIS is an intelligent system that accesses information from multiple sources providing situational awareness based on a georeferenced common picture, which is shared among system users. The information about incidents is analyzed to assess response priorities and advise on available resources' assignment.

The paper describes the approach adopted to apprehend the pragmatic and hedonic quality dimensions of the system design properties that ensure good UX, considering the context of use, users' goals and system's purpose. For this purpose, Personas were used to identify users' needs, wants and expectations in order to design best possible experience for them. UML Use-cases were used to represent the activity and interactions of the Personas with the system and among themselves. It also offers a perspective on the methods identified to perform the assessment of the prototypes and Beta versions UX quality.

The paper provides a bird's-eye view on some of the challenges faced by the THEMIS project team regarding its ongoing effort of conceptualization, design and implementation, considering the very demanding context of use experienced on disaster response operations.

THEMIS intelligent system will contribute for the preparedness and execution of disaster relief and humanitarian assistance operations.

Acknowledgments. The work was funded by the Portuguese Ministry of Defense and by the Portuguese Navy.

References

1. ISO 9241-210:2010: Ergonomics of Human–System Interaction—Part 210: Human-Centred Design for Interactive Systems. ISO (2010)
2. Simões-Marques, M.J.: Facing disasters—trends in applications to support disaster management. In: Proceedings of the Advances in Human Factors and System Interactions, AHFE 2016 Conference on Human Factors and System Interactions, pp. 203–215, Florida, USA, 27–31 July. Springer (2017)
3. Turban, E., Sharda, R., Delen, D.: Decision Support and Business Intelligence Systems, 9th edn. Prentice Hall Press, Upper Saddle River (2011)
4. Crotty, M.: The Foundation of Social Research: Meaning and Perspective in the Research Process. Sage Publications, Beverley Hills (1998)
5. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage Publications, Beverley Hills (2013)
6. Nunes, I.L., Lucas, R., Simões-Marques, M., Correia, N.: Augmented reality in support of disaster response. In: Proceedings of Advances in Human Factors and System Interactions, AHFE 2017 Conference on Human Factors and System Interactions, Los Angeles, California, USA, 17–21 July. Advances in Intelligent Systems and Computing, pp. 155–167. Springer (2018)
7. Correia, A., Severino, I., Nunes, I.L., Simões-Marques, M.: Knowledge management in the development of an intelligent system to support emergency response. In: Proceedings of Advances in Human Factors and System Interactions, AHFE 2017 Conference on Human Factors and System Interactions, Los Angeles, California, USA, 17–21 July. Advances in Intelligent Systems and Computing, pp. 109–120. Springer (2018)

8. Simões-Marques, M., Bica, J., Correia, A.: Modeling and simulation in support of disaster preparedness. In: Proceedings of Advances in Human Factors and System Interactions, AHFE 2017 Conference on Human Factors and System Interactions, Los Angeles, California, USA, 17–21 July. Advances in Intelligent Systems and Computing, pp. 121–132. Springer (2018)
9. Kraft, C.: User Experience Innovation: User Centered Design that Works. Apress, Berkely (2012)
10. Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., Sinnelä, A.: UX curve: a method for evaluating long-term user experience. Interact. Comput. **23**(5), 473–483 (2011)
11. Bongard-Blanchy, K., Bouchard, C.: Dimensions of user experience-from the product design perspective. Journal d'Interaction Personne-Système **3**(1) (2015). http://jips.episciences.org/1319
12. Reiss, E.: Usable Usability. Simple Steps for Making Stuff Better. Wiley, Indianapolis (2012)
13. UN, INSARAG Guidelines. Volume II: Preparedness and Response—Manual A: Capacity Building, U. Nations, The International Search and Rescue Advisory Group (2015)
14. Rumbaugh, J., Jacobson, I., Booch, G.: The Unified Modeling Language Reference Manual, 2nd edn. Pearson Higher Education, London (2004)
15. JustinMind. https://www.justinmind.com
16. Brade, J., Lorenz, M., Busch, M., Hammer, N., Tscheligi, M., Klimant, P.: Being there again—presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. Int. J. Hum. Comput. Stud. **101**, 76–87 (2007)
17. Tullis, T., Albert, W.: Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann Publishers, Burlington (2008)
18. Rauschenberger, M., Schrepp, M., Cota, M.P., Olschner, S., Thomaschewski, J.: Efficient measurement of the user experience of interactive products. Int. J. Artif. Intell. Interact. Multimed. **2**(1), 39–45 (2013)
19. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) USAB 2008, LNCS, vol. 5298, pp. 63–76 (2008)

Knowledge Management in the Development of an Intelligent System to Support Emergency Response

Anacleto Correia[1], Inês Severino[1], Isabel L. Nunes[2,3], and Mário Simões-Marques[1(✉)]

[1] CINAV – Portuguese Navy, Alfeite, 2810-001 Almada, Portugal
{cortez.correia, ines.sofia.severino}@marinha.pt, mj.simoes.marques@gmail.com
[2] Faculty of Sciences and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
imn@fct.unl.pt
[3] UNIDEMI, Department of Mechanical and Industrial Engineering, Faculty of Sciences and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

Abstract. Intelligent systems use applied artificial intelligence techniques with the aim of reaching, in a specific domain, a level of analysis and performance comparable to human experts. Intelligent systems are able to engage in complex inference processes, necessary for evaluating alternative options, presenting high quality conclusions and advice, and to offer explanations about the rationale that led to such conclusions. Knowledge Management is a key process associated to the development of intelligent systems, since it elicits, codifies, validates and relates the knowledge elements that are stored in the system's Knowledge Base. The paper focus on the issues involved in the design of the THEMIS intelligent system's Knowledge Base, and of the cooperative and participatory processes applied for knowledge elicitation, referring the usage of ontologies and UML use-cases. THEMIS project purpose is to develop an intelligent system to support complex and stressful Emergency Management activities. The outcome of the described Knowledge Management process can be determinant to quality of the user experience when exploiting the system, since most of the pragmatic and hedonic qualities of the interactions of an intelligent system are closely related with the characteristics of the underlying knowledge base and inferencing process.

Keywords: Knowledge acquisition · UML · Ontologies · THEMIS

1 Introduction

The variety and complexity of the situations faced and the stressful conditions in which emergency managers make decisions and first responders operate impose a comprehensive preparation for the response to emergencies. In fact, the preparedness phase of the Emergency Management cycle is key for the effectiveness and efficiency of the

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_11

response. Preparedness encompasses many activities aimed at saving lives and improving the success of response operations, including planning courses of action, stocking supplies and equipment, educating and training, or installing warning systems. Another key factor of success is the availability of an information sharing and decision support capacity (for instance, based on intelligent systems), which expedites the flow of information among the different emergency response actors and reduces the burden of fusing and analyzing high volumes of data. This is particularly true in disaster situations where decision-makers and responders are deployed to a different part of the world, eventually facing an environment with different languages and cultures, unknown operational conditions and, probably, interacting with international, governmental and non-governmental organizations for the first time. Note that, by definition, a disaster is an overwhelming and often sudden event that exceeds the capacity of response of the local organizations, calling for international support.

Intelligent systems use applied artificial intelligence techniques with the aim of reaching, in a specific domain, a level of analysis and a performance comparable or even better than that of a human expert. Despite their limitations, current intelligent systems are able to engage in complex inference processes, necessary for evaluating alternative options and offering high quality conclusions and advice, and to offer explanations about the rationale that led to such conclusions.

A key element of any intelligent system is the knowledge base (KB), since this is the component where the knowledge elements about a specific subject are stored. The process of transferring the expertise of humans into computers can be quite complex and challenging when the type of decision-making problem to replicate is unstructured or semi-structured, as the decisions are often based on human intuition. Therefore, developing KBs is usually a quite difficult task, since there is the need to figure out and code, among others, the knowledge elements, structure, context of use, composition and representation, relations and their respective importance. It is also crucial to capture the reasoning processes applied to solve the problems and code them as inference processes, which combine the inputs (coming from real world data) with the knowledge (from the KB) in order to present the desired outputs. The inference processes involved can range from structured (e.g., models) to unstructured (e.g., metaheuristic algorithms).

The paper reflects the experience gathered by the authors in the development of the KB for the intelligent system THEMIS[1] (*disTributed Holistic Emergency Management Intelligent System*), aimed at supporting real time activities of decision-makers in command posts, and of responders in the field. Further details on THEMIS project are available in [1, 2]. The development was conducted following a typical user-centered design approach, considering the context of use of the different system users, their needs and requirements, and, in particular, the choice of the most suitable formats for transferring and presenting the knowledge to the different system users, taking into account the types of tasks performed, the environment in which they operate and the characteristics of equipment used for interaction with the system. The paper addresses also the knowledge acquisition and representation processes, conducted in close

[1] THEMIS is an undergoing R&D project funded by the Portuguese Ministry of Defense.

cooperation with emergency management and emergency response subject matter experts (SME). Furthermore, the design of the KB structure and the knowledge coding considered the requirements for implementing an inference engine applying artificial intelligence techniques, such as fuzzy logics, and rule-based and case-based reasoning. The paper puts a particular focus on the issues involved in the design of the intelligent system's KB, and of the cooperative and participatory processes applied for knowledge representation, referring the definition of UML use-cases and of an ontology.

2 Knowledge Management

There is no straightforward definition for *Knowledge* and plenty of books were written on this issue. In March 2004, the European Committee for Standardization (CEN) proposed the following working definition [3]:

> *"Knowledge is the combination of data and information, to which is added expert opinion, skills and experience, to result in a valuable asset which can be used to aid decision making. Knowledge may be explicit*[2] *and/or tacit*[3]*, individual and/or collective."*

In [5] Simões-Marques and Nunes provided a thorough overview of the concepts related with the Knowledge Management (KM) process associated with the development of intelligent systems. The present paper builds on that, describing the approach that was adopted in the development of the intelligent system THEMIS, namely regarding the building of the KB.

From an intelligent systems' design perspective KM is the process aimed at transferring expertise from human experts to computers and back to humans. This process, which is also referred as Knowledge Engineering, involves five major activities [6, 7]:

- *Knowledge acquisition* (from experts and other sources) - involves the acquisition of knowledge from human experts, books, documents, sensors, or computer files. The knowledge may be specific to the problem domain or to the problem-solving procedures, This is the activity related with the externalization and combination of knowledge, since in this stage, it is required to make explicit knowledge that eventually is still tacit or to combine knowledge which is already explicit;
- *Knowledge representation* (in the computer and to the user) - previously acquired knowledge is organized so that it will be ready for use. This activity involves encoding of the knowledge in the KB and preparation of knowledge maps. It requires dealing with abstract concepts and their relations, which constitute the content of the KB;

[2] Explicit knowledge is the knowledge that can be expressed in words and numbers and readily shared (for instance as data, scientific formulae, specifications, manuals) between individuals in a formal and systematic way [4].

[3] Tacit knowledge is more difficult to transfer since is highly personal and hard to formalize, often resulting from experience [4].

- *Knowledge validation* - (or verification) involves validating and verifying the knowledge until its quality is acceptable. The results of the test cases are usually shown to subject matter experts to verify the accuracy of the intelligent system;
- *Inferencing* - is the reasoning capabilities that can build higher-level knowledge from existing heuristics. This reasoning consists of inferencing from facts and rules using heuristics or other search approaches;
- *Explanation and justification* (to the user) - This is the activity of delivering knowledge to users, namely to non-expert recipients, using adequate knowledge presentation and visualization formats, and involves an explanation capability with the ability to answer questions such as how a certain conclusion was derived by the computer.

The Knowledge Engineering activities are illustrated in Fig. 1 which was adapted from [6].

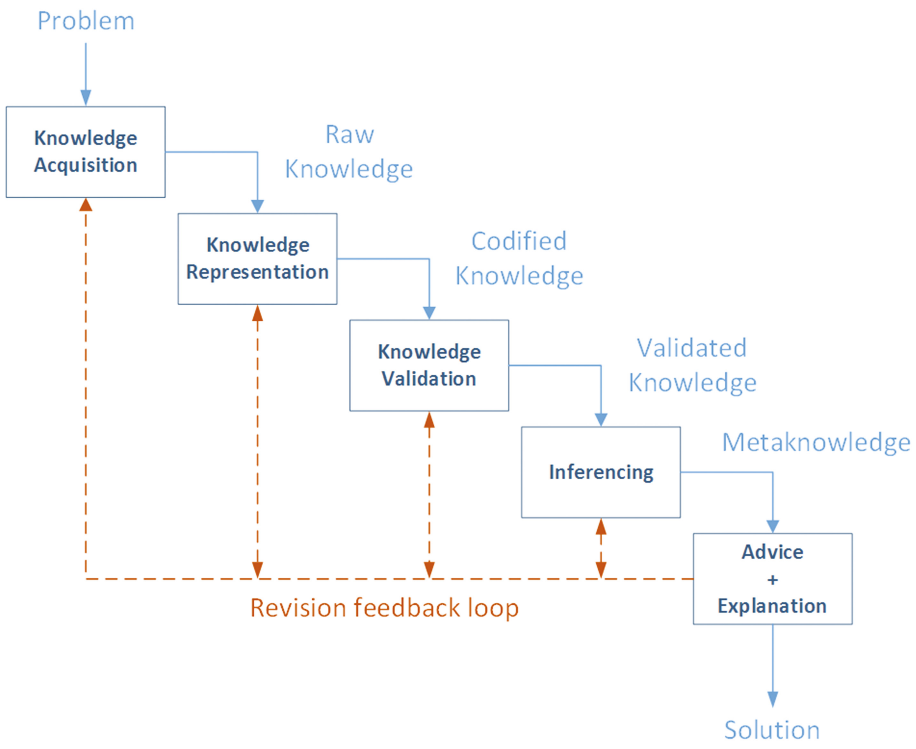


Fig. 1. The process of knowledge engineering (adapted from [6]).

Nevertheless, it is important to note that, from an organization's standpoint, the KM cycle encompasses other activities (create, capture, refine, store, manage and disseminate knowledge), with the purpose of preserving and leveraging its memory and intellectual capital usually involving initiatives meant to make knowledge visible (e.g., through maps, yellow pages, hypertext), to develop a knowledge-intensive culture, or to build a knowledge infrastructure [6].

3 Knowledge Acquisition Process

Considering that the domain of expertise of THEMIS regards the support to emergency response, the first stage of the knowledge collection process focused on identifying sources of knowledge and methods to capture it addressing this specific field of expertise. Several sources were exploited along this process, namely:

- *Documentation* - literature, handbooks, legislation, standards, procedures, statistic reviews, reports and any other forms of written documents are a good source of explicit knowledge. For this purpose a significant number of publications from relevant international organizations (*e.g.*, UNHCR, UNISDR, UNOCHA, International Federation of the Red Cross and Red Crescent Societies, European Commission), national organizations (*e.g.*, US FEMA, American Society for Public Administration, National Research Council, Portuguese Civil Protection National Authority (ANPC), British Standards Institution), cooperative organizations (*e.g.*, World Economic Forum, The Sphere Project), research institutions (*e.g.*, Centre for Research on the Epidemiology of Disasters, Asian Disaster Reduction Center), specific domain periodicals (*e.g.*, Emergency Management Magazine), books, scientific journals, conference proceedings, and legal was compiled and analyzed;
- *Subject Matter Experts* - the interaction with experts allows the access to tacit knowledge and, whenever possible, to make it explicit. In the project different approaches were used for this purpose, such as structured interviews, presentations by SME, workshops, discussion sessions. The SME involved were from Portuguese organizations with operational experience in Emergency Management (e.g., ANPC), in the deployment of disaster relief forces (e.g., Navy), and in teaching and training (e.g., National Fire Service School). Besides these institutions the project team incorporates academic and technical members with know-how in the subjects relevant to the development of the intelligent system, including the different stages of the KM process;
- *Observation* - the exposure to life situations (real or simulated) is another way of accessing to tacit and explicit knowledge. It was possible to observe the flow of information and decisions taken at ANPC's National Coordination Center, and during Disaster Relief Exercises (DISTEX) conducted by the crews of Portuguese Navy ships. In this case it was possible to observe the flow of information between the OSOCC (on-site operations coordination center) and the first responder teams that were training humanitarian assistance and disaster relief operations in an earthquake scenario. Pictures shown in Fig. 2 were taken in one of these observation sessions.

(a) Means used by the OSOCC in the DISTEX Command Post

(b) Relief operations performed by ship deployed teams during the DISTEX

Fig. 2. Pictures taken during an observation session for knowledge acquisition in a disaster relief exercise (DISTEX) conducted by the Portuguese Navy

4 Knowledge Representation

The knowledge acquired in the previous phase must be organized and codified so that it will be ready for use, by the intelligent system internal reasoning processes and by users interacting with the system. The first activity of this stage involves encoding the raw knowledge in the KB and preparing knowledge maps that support the subsequent stages related with Inferencing. This activity requires dealing with abstract concepts and their relations that will constitute the content of the KB. There is an increasing number of representation formats, emerging from applied theoretical and modeling disciplines (e.g., logics, ontologies, artificial intelligence, metaheuristics) and from the evolution of technical characteristics of computer hardware and software (e.g., programming languages, mixed reality, immersive technologies, multimodal interfaces). This section focus on the use of UML and ontologies in the process of knowledge representation.

4.1 Fuzzy Methodologies

Fuzzy Sets Theory is an example of theoretical framework available to code knowledge elements and relationships. Fuzzy Set Theory was introduced in 1965 by Lotfi Zadeh [8], based on the conviction that conventional quantitative techniques are not adequate to deal with humanistic systems neither with similar complex systems. Humanistic systems are the ones that deal with problems using an approach comparable to human reasoning; opposite to mechanistic systems, which reduce systems behavior to deterministic laws of mechanics, electromagnetism or thermodynamics. In fact, Fuzzy Sets Theory is a generalization of Classical Set Theory, better suited to handle problems in which the source of imprecision is the absence of sharply defined criteria of class membership. An example of such imprecise entities are the terms used in natural language, which can be captured by linguistic variables. Another advantage is that fuzzy sets can be seamlessly used in the knowledge inferencing stage, since Fuzzy Logics is particularly suited to implement approximate reasoning solutions, used in most intelligent systems designed to support complex decisions problems.

There is a host of scientific sources and practical applications of Fuzzy Logics, therefore this paper will not expand on this topic. There is also a significant amount of work developed in the field of emergency management, particularly related with decision support systems. For instance, [9] discusses examples on the usage of fuzzy methodologies in the context of emergency management.

4.2 UML

UML (Unified Modeling Language) is a modeling language that helps on the specification, visualization, and documentation of models of software systems, including their structure and design, in a way that meets all of these requirements.

According to language specifications [10] a UML model consists of three major categories of model elements, each of which may be used to make statements about different kinds of individuals within the system being modeled. These categories are:

- *Classifiers* - A classifier describes a set of objects. An object is an individual with a state and relationships to other objects;
- *Events* - An event describes a set of possible occurrences. An occurrence is something that happens that has some consequence with regard to the system;

- *Behaviors* - A behavior describes a set of possible executions. An execution is a performance of a set of actions over some period of time that may generate and respond to occurrences of events, including accessing and changing the state of objects.

Regarding semantics UML modeling construct encompasses two categories:

- *Structural Semantics* - defines the meaning of UML structural model elements about individuals in the domain being modeled, which may be true at some specific point in time (sometimes also called "static semantics.");
- *Behavioral Semantics* - defines the meaning of UML behavioral model elements that make statements about how individuals in the domain being modeled change over time (sometimes also called "dynamic semantics.")

In addition, there are some supplemental modeling constructs (i.e., use cases, deployments and information flows) that have both structural and behavioral aspects.

In software engineering, the UML use-case technique allows the graphical representation of interactions between actors (representing roles) and the system, to achieve their goals. Use-cases common representation as diagrams is supplemented by rich textual descriptions. Those descriptions follow a template that specifies structured functional requirements, illustrated in Fig. 3. Among the items covered by a use-case description, there are:

(1) the steps of a main scenario, narrating the interactions between actors and the system when everything goes according to plans and ending in successful tasks completion;
(2) alternative scenarios, to handle variations to the main scenario;
(3) pre-conditions, for the use-case instantiation;
(4) post-conditions, to be true after use-case execution;
(5) invariants, conditions that are true before and after running the use-case.

In addition, use-cases description can also include non-functional requirements, such as requirements regarding performance, security and user interfaces interaction.

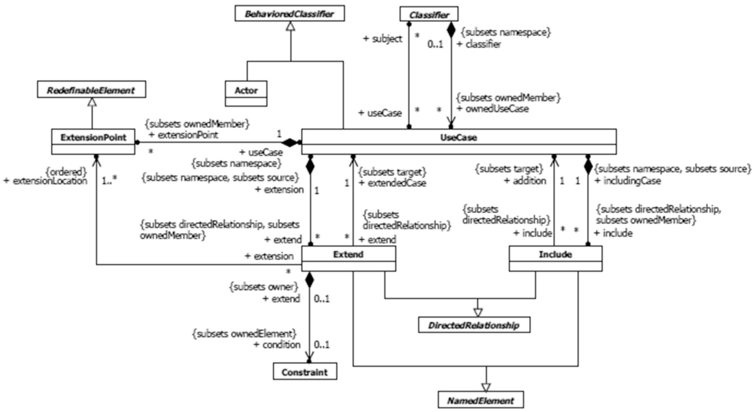


Fig. 3. Generic diagram for the design of UML use-cases (source [10])

In THEMIS' process of knowledge representation use-cases were used mainly to characterize actors, type of relations and interactions with the system, and functionalities to build in the system. Figure 4 offers an example of use-case related with simulation functionalities provided by the system. A detailed description of UML use-cases applied to the THEMIS context is offered in [2].

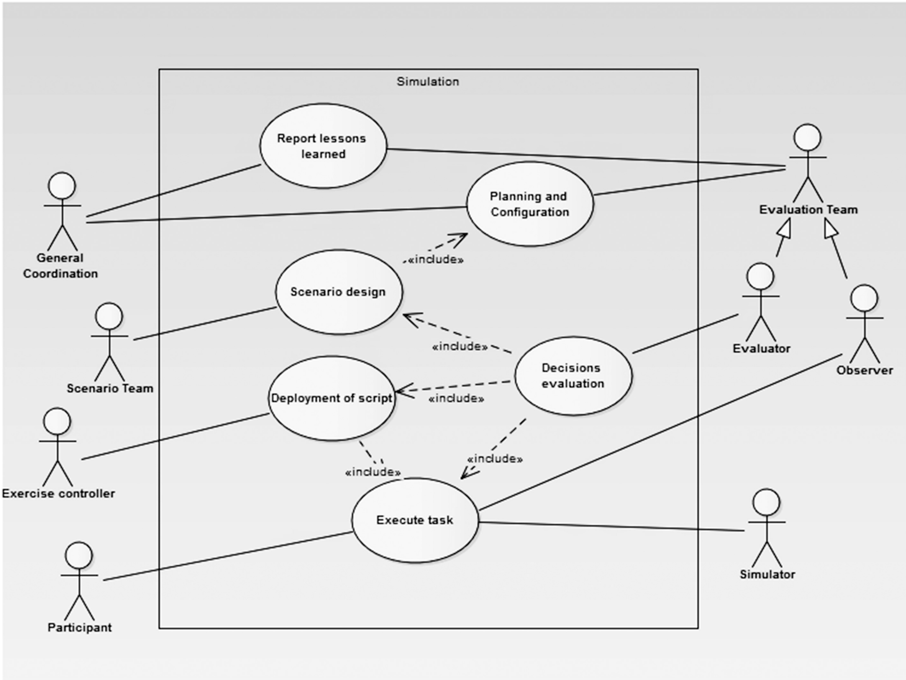


Fig. 4. UML use-case defined for the simulation functionalities of THEMIS (source: [2])

4.3 Ontologies

An Ontology is a formal naming and definition of the types, properties, and interrelationships of entities that really or fundamentally exist for a particular domain of discourse. A commonly cited definition is "an ontology is an explicit specification of a conceptualization", which was presented by Gruber, back in 1993 [11].

THEMIS project team created an ontology to represent the knowledge acquired regarding the Emergency Management process. The rational to define the ontology was based on a variety of knowledge domains, such as the standard taxonomy of disasters and the phases of the Emergency Management cycle [12], focusing particularly the Preparedness and the Response phases. The ontology also considered the DOTMLPF[4] referential, used in military planning to identify and bridge the capability gaps.

[4] Doctrine, Organization, Training, Materiel, Leadership, Personnel, Facilities.

There are many ontology editors and representation languages available. The ontology editor used in THEMIS for knowledge representation was the Protégé 5.1.0, which is a free, open-source ontology editor and framework, developed by the Stanford Center for Biomedical Informatics. Figure 5 presents a partially expanded view on the OntoGraph representation of the THEMIS ontology, where are represented a hierarchy of classes, instances of classes and relations. All these entities are used both to codify the raw knowledge acquired in the previous stage of the KM process, and to execute and support the subsequent validation and inferencing stages.

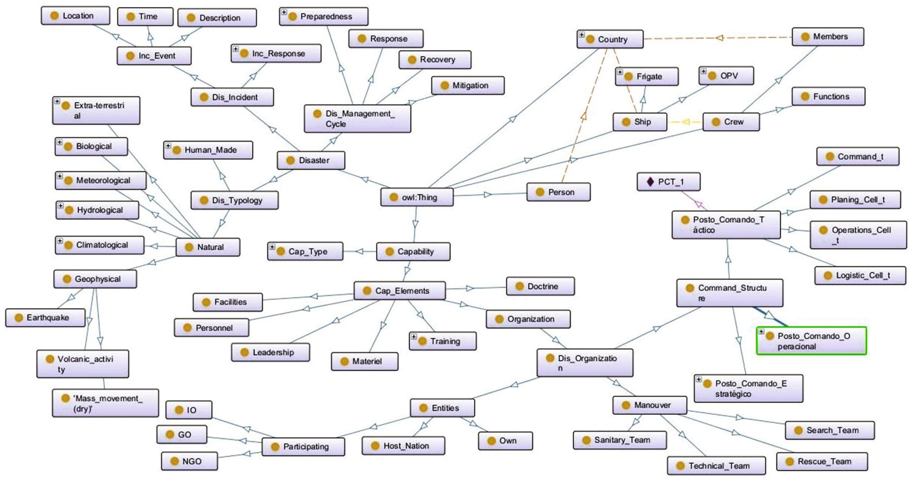


Fig. 5. Partial view of the graph representing the ontology used in THEMIS

5 Future Work

As discussed in [1] the THEMIS project follows a User Experience (UX) approach, in which the work here described fits. To attend the UX requirements, the KM process has to consider several other dimensions that were not addressed above, which have to be contemplated along the development process lying ahead.

In fact, besides the previously mentioned user centered design principles, user experience encompasses a more holistic approach including all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after system use [13]. Therefore, UX considers not only the pragmatic qualities of the human-system interactions (e.g., effectiveness, efficiency), but also the hedonic ones (i.e., satisfaction and pleasure) [14].

The next steps in the domain of KM relate with the definition of the inferencing process and the presentation of information, advice and explanations to users.

Ensuring high UX quality in the THEMIS implementation will be a big challenge, since this is an intelligent system that aims at supporting complex and stressful activities for a quite large variety of user roles. It is worth noting that the Knowledge

Management process plays a central role in achieving this goal, since the outcome of such process can severely affect the quality of the experience of users while exploiting the system. In fact, most of the pragmatic and hedonic qualities of the interactions of an intelligent system are closely related with the characteristics of the underlying knowledge base (e.g., knowledge coding, inferencing, knowledge presentation).

In THEMIS project the knowledge management activities will follow closely the interaction prototyping and the UX evaluation of the prototypes (refer to [1]), to make sure that any relevant findings is these domains are timely and duly considered.

6 Conclusions

The paper reflects the experience gathered by the authors in the development of the KB of the intelligent system THEMIS, aimed at supporting real time activities of decision-makers in command posts, and of responders in the field.

The KB development, which is one of the activities within the THEMIS development, followed a typical user-centered design approach, considering the context of use of the system; identified users' needs and system requirements; and addressed, in particular, different ways for acquiring and representing the knowledge, such as fuzzy logics, UML use-cases and ontologies. All the knowledge management activities performed took in account the types of tasks performed, the environment in which users operate and the characteristics of the interaction with the system. The paper focused the knowledge collection process, conducted in close cooperation with emergency management and emergency response subject matter experts. Furthermore, the design of the KB structure and the knowledge coding took into account the requirements for implementing an inference engine applying artificial intelligence techniques.

Finally, the paper addressed the challenge of ensuring high UX quality in the THEMIS implementation, since this is an intelligent system to support complex and stressful Emergency Management activities, noting the central role that the knowledge management process has in this matter. In fact the outcome of KM process can be determinant to quality of the user experience when exploiting the system, since most of the pragmatic and hedonic qualities of the interactions of an intelligent system are closely related with the characteristics of the underlying knowledge base and inferencing process.

Acknowledgments. The work was funded by the Portuguese Ministry of Defense and by the Portuguese Navy.

References

1. Simões-Marques, M., Correia, A., Teodoro, M., Nunes, I.L.: Empirical studies in user experience of an emergency management system. In: Proceedings of Advances in Human Factors and System Interactions, AHFE 2017 Conference on Human Factors and System Interactions, 17–21 July 2017. Advances in Intelligent Systems and Computing, pp. 97–108. Springer, Los Angeles (2018)

2. Simões-Marques, M., Bica, J., Correia, A.: Modeling and simulation in support of disaster preparedness. In: Proceedings of Advances in Human Factors and System Interactions, AHFE 2017 Conference on Human Factors and System Interactions, 17–21 July 2017. Advances in Intelligent Systems and Computing, pp. 121–132. Springer, Los Angeles (2018)
3. CEN: European Guide to Good Practice in Knowledge Management—Part 1: Knowledge Management Framework. European Committee for Standardization, Brussels (2004)
4. Nonaka, I., Konno, N.: The concept of 'Ba': building a foundation for knowledge creation. Calif. Manag. Rev. **40**(3), 40–54 (1998)
5. Simões-Marques, M., Nunes, I.L.: Contributions to the design of knowledge base development tools for intelligent systems. In: Rebelo, F., Soares, M. (eds.) Ergonomics in Design: Methods and Techniques, pp. 383–399. CRC Press – Taylor & Francis Group, London (2016)
6. Turban, E., Sharda, R., Delen, D.: Decision Support and Business Intelligence Systems, 9th edn. Prentice Hall Press, Upper Saddle River (2011)
7. Turban, E., Aronson, J.E., Liang, T.P.: Decision Support Systems and Intelligent Systems, 7th edn. Prentice-Hall, Upper Saddle River (2007)
8. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
9. Simões-Marques, M., Nunes, I.L.: A fuzzy multicriteria methodology to manage priorities and resource assignment in critical situations, Springer under Series ORCS. In: Zeimpekis, V., Ichoua, S., Minis, I. (eds.) Humanitarian and Relief Logistics: Research Issues, Case Studies and Future Trends, pp. 129–153. Springer, New York (2013)
10. OMG: OMG Unified Modeling Language TM (OMG UML)—version 2.5. Object Management Group, Inc., Needham (2015)
11. Gruber, T.R.: A translation approach to portable ontologies. Knowl. Acquis. **5**(2), 199–220 (1993)
12. FEMA: Principles of Emergency Management. Federal Emergency Management Agency (2007)
13. ISO 9241-210:2010: Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems. ISO (2010)
14. Bongard-Blanchy, K., Bouchard, C.: Dimensions of user experience-from the product design perspective. Journal d'Interaction Personne-Système (JIPS) **3**(1) (2015). http://jips.episciences.org/1319

Modeling and Simulation in Support of Disaster Preparedness

Mário Simões-Marques, João Bica, and Anacleto Correia(✉)

CINAV – Portuguese Navy, Alfeite, 2810-001 Almada, Portugal
mj.simoes.marques@gmail.com,
{santos.bica, cortez.correia}@marinha.pt

Abstract. Modeling and Simulation (M&S) plays a key role in exercising the decision-making process and in training the use of complex systems and procedures. Authorities, responsible to tackle complex processes in response to a plethora of types of natural and anthropogenic disasters, need tools that could support training and preparedness of teams, challenging them with scenarios as close as possible to reality. This work addresses challenges faced by the project team during the conceptualization, design and implementation of a simulation module of the THEMIS emergency management intelligent system. The aim of this intelligent system is to support disaster response operations and to improve the planning and execution efforts in disaster and crisis' scenarios. Because the M&S module provides the incident injects to use by THEMIS, it also constitutes a key element in the toolkit considered in the usability evaluation plan of THEMIS, namely regarding the user interactions in the different types of equipment adopted.

Keywords: Modeling · Simulation · Human-systems interaction · THEMIS

1 Introduction

Disaster Management (DM) is a complex process for dealing with several families of natural (e.g., geophysical, climatological, hydrological) and anthropogenic (e.g., non-intentional, intentional) disasters. A well-designed DM plan seeks to prevent the damage caused by disruptive effects of disasters and minimize the potential losses associated to known hazards, by ensuring a high level of preparedness for performing a timely and appropriate assistance to victims of catastrophes and recovery of the society's normal life, in a rapid and effective way. The DM cycle encompasses mitigation, preparedness, response and recovery stages. Through the activities of each of these stages authorities, organizations, and civil society make efforts to minimize the effects of future disasters, plan for appropriate response, react during and after a disaster, and take steps to return to normalcy as soon as possible. Naturally, only a careful selection of courses of action for the different phases of the DM cycle can lead to high preparedness, better warning systems, reduced vulnerability and the prevention of disasters. Therefore, the success of DM demands coordinated efforts from decision-makers, coordination center operators and first responders.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_12

Disaster relief preparedness requires addressing issues in different domains, such as doctrine, organization, training, materiel, leadership, personnel, facilities, or interoperability; with the goal of ensuring that participants are well equipped and organized, possess effective knowledge and experience of the tactics, techniques and procedures to implement, and are ready to be engaged in operations. This requires all relevant actors to follow a well-designed training plan in order to be acquainted with the typical tasks to perform, with the operation of the available emergency management systems and with any procedures to implement when emergencies arise.

Modeling and Simulation (M&S) plays a key role in exercising the decision-making process and in setting the stage for training the use of complex systems and procedures. The evolution and growing dimensions of this discipline on several domains (e.g., aerospace, telecommunications) are very well described in literature. M&S are representations of a system. 'Modeling' addresses the static representation of the system entities, while 'Simulation' addresses their behavior or dynamic over time. M&S plays also an important role in ensuring the success of man-machine interaction of socio-technical systems, since it can contribute to improve the quality of the user experience.

M&S can be used in all the stages of a system life cycle, and adopt many different approaches, as discussed by Simões-Marques in [1]. The current paper addresses the initial stages of the life cycle of an emergency system M&S module, from inception to preliminary design implementation. The M&S module is currently under development with the purpose of generating a correlated set of events, combining different types of emergency incidents (e.g. fires, traffic accidents, industrial accidents, oil spills and nuclear, biological, chemical or radiological events) using appropriate simulation techniques. The above mentioned events constitute injects for training sessions, ranging from dynamic near real-time table-top exercises (aimed at decision-makers) to the planning and execution of man-in-the-loop field exercises (aimed at first responders or at the entire disaster management structure).

The simulated events will be used mainly to assess the performance of the trainees involved in the emergency management actions, in accordance with the specific stage of their training process. However, the lessons learned from the execution of the simulations will be useful for other purposes. One, still related with training, is to fine-tune the training process, as well as the decision-making and response processes. Another, related with the user-centered design approach adopted in the THEMIS development process, is to use simulations to support the usability assessment, triggering events that result in interactions of users with the system, and also to support the validation of the accuracy and general quality of the support provided by the emergency management intelligent system.

The paper is structured as follows: the next section, presents the motivation for preparedness and the relevance of M&S in the DM cycle; Sect. 3, discusses the requirements an infrastructure must provide to support an M&S module, several approaches for M&S implementation, and the validation process of current the M&S approach; while the last section offers some conclusions.

2 Training for Disaster Management

2.1 The Preparedness Phase

An effective response to disasters requires the implementation of a preparedness and response plan, designed to enable organized and coordinated actions during the crisis. Plans are not theoretical exercises: they must be frequently tested so that they can be evaluated, adapted, and updated before and after an actual disaster situation [2].

Organizations used to deal with crises, such as the militaries, recognize the relevance of a regular training for improvement of individual, group and organizational skills, for worst-case scenarios. The aim in developing training techniques for disaster management is to improve the performance of decision-makers and field operatives. Training in emergency management also warrants the DM process with: a set of instruments that allows plan development for crises mitigation; knowledge about different kind of disasters to face; understanding on the several stages the different types of disasters may evolve, from inception to resolution; skills necessary for crises monitoring and control using key performance indicators.

2.2 Simulations

Simulations are useful for evaluating and testing the effectiveness of response plans. They can also provide an important contribution to improve preparedness by developing team work and inter- and intra-sectors coordination, through training and decision-making exercises, and the consolidation of procedures [1, 2].

Simulation techniques in the context of disaster management were developed based on the assumptions that [3]:

- It provides an active learning experience, as opposed to a passive role as learner;
- Offers an opportunity for participants practicing role-playing, getting insight on the actual difficulties and limitations of executing certain functions;
- Participants are equipped with interactive equipment that provide emergency's situation awareness;
- Increase participants' ability to cope with disasters, by exercising prescribed procedures, and exploring safely the available alternatives. Decision makers acquire information and confidence to select the best alternative when a crisis effectively occurs [4]. On the other hand, their decisions can be integrated in the process of simulation and thus influencing reactions taken by other participants;
- Derived scenarios provide realistic situations to participants, contributing to rise up preparedness. Noisy data can also be injected by both the exercise controller and by the participants adding new variables to the decision equation in emergency;
- The events happen in "compressed time";
- Considerable amounts of material and human resources can be saved;
- The simulation exercise should be evaluated and the results applied as lessons learned for adapting and improving preparedness plans.

2.3 The Simulation Script

A simulation is based on a scenario implemented by a script of events, some of them resulting from participants decisions, based on information that they receive during the exercise [2]. Those decisions, for being valid, must comply on existing procedures and be aware of existing resources.

Therefore, the script, as a key element of the simulation exercise, must be logical and realistic, by establishing the timeline of events and the involvement of the different roles in each of the described situations [2]. The script unfolds the sequence of messages and the actions expected by the participants in the situations they successively confront. The script generally also includes a description of the scenario (e.g. background of the locale; geopolitical characteristics, target population).

The Use Case diagram [5] in Fig. 1 depicts the main actors involved in the M&S process regarding script configuration, execution and evaluation, comprehending:

- *General Coordinator* - prepares the simulation model by defining the configuration parameters of the process, monitoring the simulation execution and reporting lessons learned, at the end of the process, in the evaluation phase;
- *Scenario Team* - designs the scenario script with the likely effects of the crisis being simulated, in collaboration with experts on the type of disaster;
- *Exercise Controller* - oversees the script unfolding through the occurrence of different events as time periods elapse during the exercise;
- *Participants* - perform assigned tasks and follow procedures according to specialties;
- *Simulators* - act as simulated characters according to the roles assigned;
- *Evaluation Team* - follows the decisions and actions taken by participants during the exercise and record the observations. This team integrates evaluators and observers;
- *Evaluators*- assess the decisions taken by participants during the exercise;
- *Observers* - evaluate processes or activities, based their own criteria or experience.

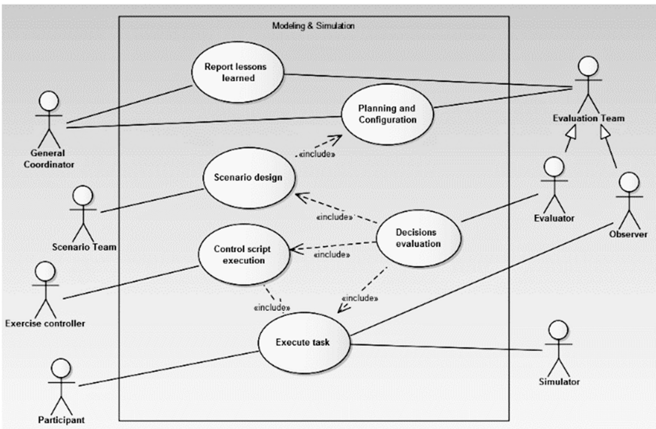


Fig. 1. Use case diagram of the M&S process

3 Modeling & Simulation in THEMIS

3.1 Modeling & Simulation Architecture

In the context of THEMIS project, M&S functionalities were added to the decision support system, by designing a module for scripts configuration, execution and evaluation.

The *Event* class is the core entity in the M&S system (see Fig. 2). This consists on the description of the features of an incident and its effects on the population, infrastructure, services, environment, and general impact on the affected area. The *Scenario* of a simulation is previously subject to a *Configuration* process, which gives place to a set of *Parameterized* events, reflecting a particular kind of incident. By unfolding the script, parameterized events are instantiated, given birth to *Simulated* events. The impact of an event and the responses it triggers from participants are depicted in Fig. 2 as unary association of the event class.

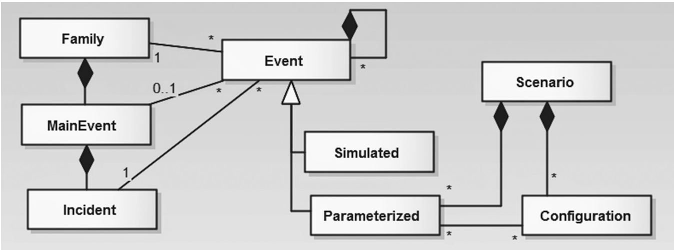


Fig. 2. UML class diagram of the M&S system

Following the IDRD classification system [6], the events characterization presented in Fig. 2 distinguishes three levels of information detail: *Family*, *MainEvent* and *Incident*. Each disaster family (e.g. geophysical, hydrological, etc.) can be further classified by a set of main events (e.g. earthquake, pandemics, flood). The geophysical family, for instance, can be further subdivided into earthquake, mass movement and volcanic activity. Whenever more detailed or specific information is available (e.g., damaged infrastructures data) this can be recorded at the incident level. An incident (e.g., lightning, tornado) is the specific cause of the victims and damage.

The UML class diagram was refined in a relational database (see the simplified relational data model in Fig. 3) to allow the M&S system implementation. This representation also supports the identification of the elements of information the users must interact with.

As shown in the class diagram, the *Event* entity is the heart of M&S implementation. For each event is stored the type of disaster (family, main event and incident), its status, time and date of incident, place of occurrence and other associated events generated. For simplicity, the field of event table *notes* stores detailed information such as, damage to public/private/critical networks and infrastructure/critical services facilities.

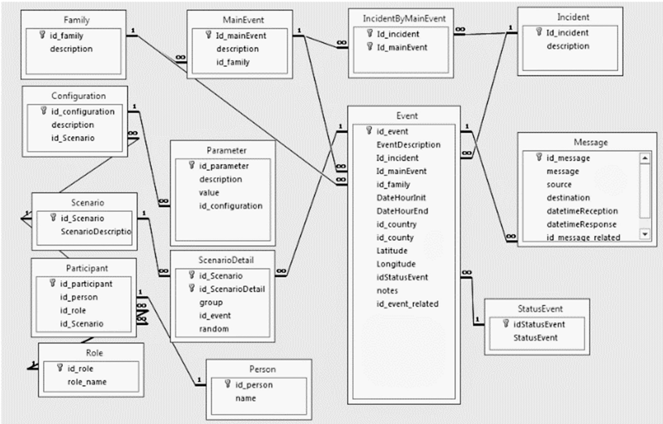


Fig. 3. Simplified relational data model of M&S system

Each participant must act according to the *Role* (i.e., a user profile) that he/she plays in the operational structure. The messages, on the other hand, inform participants of the development of the simulated events, pose problems, and provide instructions. They are delivered sequentially as established by the script.

It is worth mentioning that incidents are associated with one or several main event categories. For example, lightning could be associated with convective storms in the main event but could also be associated with tropical cyclones. In other words, there is not an exclusive one-to-one relationship between incidents and main events.

In terms of physical implementation, the simulation process involves two parts (see Fig. 4): (i) *M&S module* - the component that allows script configuration and generation based on the techniques referred in the next section; (ii) *Simulation interface* - the component that imports the generated events and present them, per role, to the simulation participants. The simulated messages can be delivered in different formats (e.g., media news, e-mails, direct inputs to THEMIS). At the current stage of the project there was a need to adopt an application that could provide an adequate interface environment, and the selection has fallen in the EXITO [7].

The EXITO system (see Fig. 5) was designed as a communication and coordination tool to keep in track the execution of complex exercises with large number of injects and players. EXITO developers identified the following set of features of the system: it helps the moderators of large scale, multi-party exercises and enables the exchange of events and reports among users by managing a Master Scenario Event List (MSEL); massively injecting events; and collecting feedback from the exercise players. EXITO is a web based system integrated with a Content Management System (CMS).

3.2 Event Generator

The M&S module event generator must address several factors that characterize the evolution of a simulated disaster, namely the regarding time, type and location.

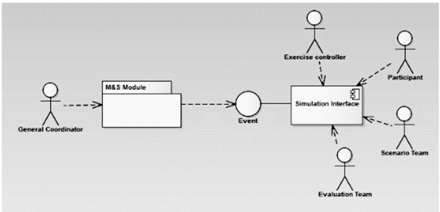


Fig. 4. Component diagram of the simulation process

Fig. 5. The EXITO interface

Time-wise. Regarding the timing of the simulated events, the M&S event-generator module implements probability functions, that were deemed as representative of the more relevant disaster categories (see Fig. 6).

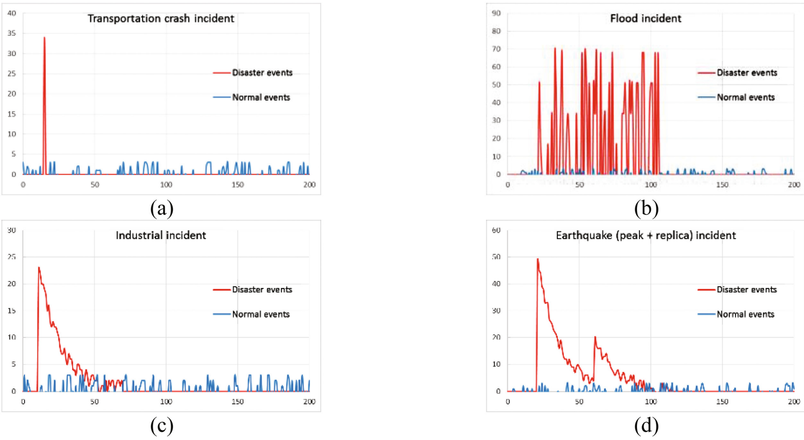


Fig. 6. Probability functions for different type of incidents: (a) transportation crash; (b) flood; (c) industrial/biochemical; (d) earthquake.

An incident such a crash involving transportation vehicles (e.g. airplane, train, bus), requires the mobilization of a great amount of emergency means for a narrow amount of time (Fig. 6a). This behavior is characterized by a function resembling the Dirac δ function. Therefore, the probability function (pdf) is zero everywhere except at one point p (the crash point):

$$f(x;p) = \begin{cases} 1, & x = p \\ 0, & x \neq p \end{cases} \tag{1}$$

A flood emergency crisis is characterized by a high level of requests (e.g. people and animal rescue) during the overflowing period, which may remain for a certain period (Fig. 6b). This phenomenon is mathematically modeled as a discrete uniform distribution where a known, finite number of incidents are equally likely to happen. Therefore, the probability function is zero everywhere except during the overflowing period, beginning and finishing, respectively, at moments a and b:

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & otherwise \end{cases} \tag{2}$$

In case of biochemical or industrial emergency, it is expected that, after the initial outbreak with a high number of reported incidents, the tendency will be for the deceleration on the number of reports (Fig. 6c). This behavior is modeled by the (negative) exponential distribution, since the events occur continuously and independently at a constant average rate λ (with $\lambda > 0$). It can also be seen as the distribution of the time between events when the number of events in anytime interval has a Poisson distribution. Therefore, supposing the outbreak occurs at moment p, the probability function is given by the expression:

$$f(x; \lambda, \mathrm{p}) = \begin{cases} \lambda \mathrm{e}^{-\lambda \mathrm{x}}, & x \geq p \\ 0, & x < p \end{cases} \tag{3}$$

Phenomena of type earthquake are expected to have a main peak with a high number of reported incidents, and possibly a replica, with a second burst of incidents, before the tendency for the deceleration on the number of incidents reported (Fig. 6d). This can be modeled by sum of two exponential distribution (one regarding the peak and the other the replica), with constant average rates of λ_1 and λ_2 (with λ_1, $\lambda_2 > 0$). Therefore, supposing the peak occurs at moment p_1 and the replica at moment p_2, the probability function is given by the expression:

$$f(x; \lambda_1, \lambda_2, \mathrm{p}_1, \mathrm{p}_2) = \begin{cases} 0, & x < \mathrm{p}_1 \\ \lambda_1 \mathrm{e}^{-\lambda_1 \mathrm{x}}, & \mathrm{p}_1 \leq x < \mathrm{p}_2 \\ \lambda_1 \mathrm{e}^{-\lambda_1 \mathrm{x}} + \lambda_2 \mathrm{e}^{-\lambda_2 \mathrm{x}}, & x \geq \mathrm{p}_2 \end{cases} \tag{4}$$

Therefore, and according to the purpose of the simulation, the MSEL receives injects that are generated with one of the above-mentioned distributions, function of the selection of disaster profile.

Type-wise. A second feature of the simulated events is the incident type (e.g., fires, traffic accidents, industrial accidents, oil spills and nuclear, biological, chemical or radiological events), considering the previously mentioned disaster system classification. The relational data model illustrated in Fig. 3 is an easily intelligible simplification of the real data model. In fact, the M&S event-generator module is linked to the THEMIS knowledge base which contains an ontology defining a set of relationships

Fig. 7. Types of incident location areas considered by the event-generator: urban, industrial, countryside, and maritime.

that allow to infer which types of events are adequate to simulate in a disaster scenario of a specified category (for further details see [8]). Ultimately, the type of incidents is randomly chosen from an adequate pool of archetype incidents recorded on the M&S database.

Location-wise. Besides incident time and type, the M&S event-generator module also must set the location of the simulated events. For that purpose, for each category of incident the database contains information regarding the geographical context in which the incident can occur, namely the feasibility of happening in urban, industrial, countryside or maritime areas, as illustrated in Fig. 7. Based on a classification on the nature of the areas present in the simulation field (Playground), the algorithm randomly positions the event and validates its plausibility to constitute an inject in the MSEL. The process is repeated until a plausible tuple type-location is reached.

3.3 Simulation Techniques

Modeling and simulation has entered the mainstream of methods available to help organizations increase their efficiency and effectiveness. The proved abilities of simulation to attack a wide range of problems and investigations rest on its abilities to accommodate stochastic variation, analyze discrete or continuous variables, or both, and provide visualization via animation [9].

Simulation analyses subdivide themselves into three major perspectives:

- *Continuous simulation* - processes to model are described in terms of differential and/or difference equations, exploiting the fact that relevant variables are continuous and differentiable functions of time. The analysis of the system then involves integration of these equations as time is driven forward. This perspective is well suited for modeling of physical systems, as is regularly done in structural analysis;

- *Discrete-event simulation* - studies processes in which time progresses in a series of steps, and significant variables within the model are integer-valued. In this approach, the concept of derivative is not applicable, since here discrete is opposite to continuous. For instance, the function *L(t)*, specifying an incidents' queue length as a function of time, is constant during intervals of time and changes value at arbitrary instants of time. Its derivative is zero on the intervals and undefined at the instants of change from one integer value to another; its integral is equal to the weighted average queue length. Under this perspective, differential or difference equations are inapplicable. There are three approaches to this perspective: (i) *activity-oriented approach* - the modeler develops characterizations of each significant activity within the system under study. These characterizations describe the system changes and quantify the passage of time during each activity; (ii) *event-oriented approach* - the modeler develops characterizations of each significant event within the system under study. Whereas activities span an interval of time, events occur at instants of time; (iii) *process-oriented approach* - the modeler develops a characterization of how items (e.g. emergency incidents) flow through the system under study. This orientation is most conveniently employed in the context of using either a computer language (e.g., general-purpose system simulator [GPSS]) or a simulation package expressly designed for this use (e.g., Arena [10]).
- *Mixed simulation* joins the discrete and continuous approaches. It is initiated by describing the system in terms of differential or difference equations, and then specifying threshold values beyond which each equation is no longer applicable. Hence, analysis should first consider the continuous aspects of the system, and then superimpose the discrete aspects, such as event triggers, on the continuous behavior.

Based on this categorization of simulation techniques, discrete-event simulation approach is the most appropriate for dealing with the context of emergency systems, where events are collected in a queue and dispatched according prioritization criteria.

3.4 M&S Module Evaluation

For the evaluation of the M&S module, two types of assessment are considered: (i) *Evaluator assessment* - a team of evaluators rates the logic and realism of the generated script; (ii) *Participant assessment* - a sample of participants rates how the script and the simulation help them to achieve their learning outcomes.

During the technical design of scripts' evaluation, the various execution phases are planned, including the events to be addressed by the participants, the tasks to be performed, the resources that will be available, and all other necessary items for the exercise.

The duration of the exercise (i.e. period between the script's starting and completion) should last between 4–6 h to achieve the full effect of an emergency for participants.

During the simulation, the evaluators will closely monitor the deliberations and actions of the participants without intervening in any way in the participants' dynamics. They use an evaluation form to record their assessments. At the end of the simulation, the coordinator of the evaluation team begins the evaluation period by asking the participants to express their perceptions regarding the timeline of the events generated, individually and by group, and their views on the quality and relevance of the entire scenario used. Following this, observers and controllers share the reasoning they use for evaluating the script performance, and finally the evaluation team communicates its views and findings on efficacy of the simulation. The evaluation team should meet to exchange notes, and analyze and gather the individual assessments to compile a comprehensive assessment. The evaluations are delivered to the general coordinator of the event for use by the M&S module team for algorithms and configuration tuning.

Documenting lessons learned from the script execution provides the feedback needed to update disaster and emergency scenario plans, and to correct any shortcomings to planning subsequent script generations. This requires analysis and compilation of all technical and administrative documentation generated at various script stages, including problems faced and how they were solved.

The evaluation of the scripts should consider issues such as:

- flow of information, overall dynamics and coordination triggered by the exercise;
- adequacy of the script to generate situations that test the ability of the participants to take appropriate and pertinent decisions under pressure;
- impact of the quality of decisions taken by the participants, on the script execution;
- logic and consistency of the script execution and what was anticipated in the plans, protocols, and procedures that were applied;
- coherence of resources made available by the script execution for the participants;
- impact of script execution in the specific performance of each of the participants.

4 Conclusion

To respond to disasters' impact decision makers must be prepared and aware of previous response plan, in order to be able of coordinating actions during the crisis. Plans should not be theoretical exercises, conversely, they must be frequently simulated so they can be evaluated, adapted, and updated before an actual disaster situation occurs.

This work addresses the initial stages of development of the M&S module of the THEMIS emergency system, from inception to preliminary design implementation. The M&S module is currently under development with the purpose of generating a correlated set of events, combining different types of emergency incidents using a discrete-event simulation approach. The above mentioned events constitute injects for training sessions, ranging from dynamic near real-time table-top exercises (aimed at decision-makers) to the planning and execution of man-in-the-loop field exercises (aimed at first responders or at the entire disaster management structure).

THEMIS is an undergoing project funded by the Portuguese Ministry of Defense aimed at supporting real time disaster management activities of decision-makers in coordination centers, and of responders in the field. THEMIS is an intelligent system that accesses information from multiple sources providing situational awareness based on a georeferenced common picture, which is shared among system users. The information about incidents is analyzed to assess response priorities and advise on available resources' assignment.

It was also noted that, because one of the M&S module purposes is to feed incident injects directly into the THEMIS intelligent system, the module also constitutes a key element in the toolkit that will be used to perform the usability evaluation of the THEMIS implementation, namely regarding the user interactions in the different types of equipment adopted.

Acknowledgments. The work was funded by the Portuguese Ministry of Defense and by the Portuguese Navy.

References

1. Simões-Marques, M.J.: Modeling and simulation in system life cycle. In: Procedia Manufacturing (Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, Las Vegas, USA). pp. 4373–4380. Elsevier (2015)
2. PAHO: Guidelines for Developing Emergency Simulations and Drills. Pan American Health Organization, World Health Organization, Washington, DC (2011)
3. Rosenthal, U., Pijnenburg, B.: Simulation-oriented scenarios. In: Rosenthal, U., Pijnenburg, B. (eds.) Crisis Management and Decision Making: Simulation Oriented Scenarios, pp. 1–7. Springer, Berlin (1991)
4. Booth, S.: Interactive simulation and crisis management training. In: Rosenthal, U., Pijnenburg, B. (eds.) Crisis Management and Decision Making: Simulation Oriented Scenarios, pp. 1–7. Springer, Berlin (1991)
5. Rumbaugh, J., Jacobson, I., Booch, G.: The Unified Modeling Language Reference Manual, 2nd edn. Pearson Higher Education, Upper Saddle River (2004)
6. IRDR: IRDR peril classification and hazard glossary. IRDR (2014)
7. Garcia, A.P., Siaterlis, C., Benoist, T.: Technical guidelines for preparing and executing cyber-exercises using EXITO, the EXercise event Injection TOolkit, European Commission (2012)
8. Correia, A., Severino, I., Nunes, I.L., Simões-Marques, M.: Knowledge management in the development of an intelligent system to support emergency response. In: Advances in Human Factors and System Interactions. Proceedings of the AHFE 2017 Conference on Human Factors and System Interactions, 17–21 July 2017. Advances in Intelligent Systems and Computing, 109–120. Springer, Los Angeles (2018)
9. Ülgen, O.M., Williams, E.J.: Simulation methodology, tools, and applications. In: Zandin, K.B. (eds.) Maynard's Industrial Engineering Handbook, 5th edn, pp. 11.101–11.119. McGraw-Hill, New York (2001)
10. R. Automation: Arena Simulation Software. https://www.arenasimulation.com/ (2017)

Visual Search Techniques and Performance in Electronic Map Display

Yan Shengyuan[1] and Jean Luc Habiyaremye[1,2](✉)

[1] School of Mechanical and Electrical Engineering, Harbin Engineering University, Nantong Street, No. 145-1, PO. Box 150001, Harbin City, China
yanshengyuan@hrbeu.edu.cn, habijealuc@yahoo.fr

[2] School of Science and Technology, University of Rwanda, PO. Box 3900, Kigali, Rwanda

Abstract. The appropriate visual search techniques and visual display methods lead to the most effective and quickest way of extracting information during visual search tasks. Visual search time and performance in visual display terminals can be influenced by various parameters, such as an amount of the clutter objects and the screen size. This paper investigated the visual search techniques and performance using radar interface with different symbols that vary in color and features. Thirty-six subjects participated in the study and experiments were performed on separate monitors whose screen size is 7, 15 and 21 in., the visual angles are experimentally set to be 45; 55; 65; 75 and 85 min of arc. This article took into considerations various performance variables such as completion time, fixation duration etc. to complete the visual search investigation. It was confirmed that the accuracy is highly affected by many clutter objects in the region surrounding a target.

Keywords: Display terminals · User interface · Icon size · Eye movements · MDT · MTD

1 Introduction

The appropriate use of the human-computer system (HCS) is central to achieving an efficient visual search performance for operators. Due to the complexity of user interfaces in computers and devices that complicate the retrieval of visual information, the users must try their best to achieve maximum accuracy and minimize search time. Tracking eye movement is one of the reliable methods to analyze the visual cognitive process and predict the users' visual attention and workload [1].

During the highly demanding visual tasks, pilots must fixate objects finely and quickly, scan the field terrain neatly for accurate search target identification [2]. The visual search and decisions making in the presence of too much information have become more demanding, and a growing information overload becomes a problem [3]. Therefore, the evaluation of visual search analysis was proven to be consistent for predicting future user interface design considerations, understanding the operator's visual workload and possible performance enhancing techniques [4].

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_13

1.1 Visual Search Parameters

The improved search efficiency depends on rapid control processes and automatic systems that keep visual attention active [5]. In visual search, saccades play a vital role in the assessment of complex visual situations and provide a maximum amount of relevant information for processing [6]. Besides, rapid eye movements are essential in selecting the locations to be fixated and determining the search time for each fixation in an environment where the user faces visual search challenges [7]. The visual search performance and the optimization of the search system can be achieved by analyzing fixation time length and saccadic amplitude data [8].

In visual search tasks, the completion time should be as short as possible to reduce mental workload and risk of missing a target. Any visual search task whose fixations appeared to exceed 500 ms increased participants' cognitive process when they perform multiple challenging tasks [9]. Some researchers confirmed that a combination of eye movements and high saccadic amplitude could be a source of visual strain in visual search task [10].

2 Factors Affecting Optical Search Efficiency

While working with complex user interfaces, operators have to receive substantial amounts of graphical information that needs to be interpreted, and so, the optimization of search times is adequately required in designing graphical information. Computer graphics transmit accurately informative communication faster than words because they are capable of presenting the meaning of the object in a condensed form and circumventing language limitations [11]. Also, in the graphical display area, factors such as stimulus density, object size, figure and background area ratio, the luminance contrast between the objects and background can unmistakably enhance visual search efficiency [12]. The technology has changed the traditional way of displaying information such as pictures, web pages, maps on the desktop and laptop users and switched on to the use of smaller display terminals such as tablets, smart phones and pocket PCs that are increasingly used to display users' mobile information [13]. The present survey will investigate the effect of various screen dimensions in displaying information and how graphical icon size can be adjusted to improve usability

2.1 Types of Map Background

The size of the local distracter region close to a search target strongly affects any chances of identifying the search target especially when it appears broad enough to mask the object and confuse the observer. Measuring clutter could subsequently help optimize visual displays or provide alerting system when the confusion level seems to be degrading visual search performance. Also, measuring interference help researchers penetrate search environment comprising more natural and sophisticated imagery during visual search activities [14].

This study took into considerations two types of topographic map presentation such as a map with displayed topography (MDT) and a map without a topographic display (MTD).

Maps are used to find the preferred route, carry out visual search tasks, locate a geographical area, evaluate visual search performance and workload, etc.

Some studies demonstrated that local clutter could influence target detection and affect human accuracy, visual search efficiency and performance [15]. Robert Lloyd analyzed the search processes used to find information on maps and confirmed that the nature of information and map background presented on a map could have an effect on search time and some bearing on decision-making [16].

Therefore, this paper carried out an assessment of the clutter characteristics and how it could affect the search performance in a map background. Although the map designer's job must be finished before the reader's begins, the designer must be aware of the cognitive processes used by map users to search for and interpret information, before selecting and representing it on the map background clutter.

2.2 Different Steps of the Analysis

Visual information is always acquired when people look at something. In this case, the cognitive processing load is distinguishable through the number and duration of fixation at a particular location [17]. Although several visual search studies with various techniques have been performed on the application of daily lives, little is known about the visual search accuracy and the design of the combat zone map display. This study intends to examine the interaction effect of visual search performance and fights zone-map display parameters.

3 Experimental Process

3.1 Participants

Thirty-six students from Harbin Engineering University aged between 24 and 29 years, participated in the study. They all have received a clinical proof to confirm if their vision is normal or otherwise corrected-to-normal as well as visual acuity. Two training sessions were performed to explain the meaning of the symbols to be used and objectives of the study and they were given a list of the field symbols. Participants had a half an hour to adapt to the trial room environment before the test starts.

3.2 System Setup

This study observed a distance of 50 cm between the screen and the participants and the presentation was performed on an HD screen resolution monitor. The SMI EyeLink II eye tracker measured the eye movements with a sampling rate of 250 Hz, binocular viewing and chin and forehead rests to minimize head movements. The illumination in the test room was around 500 lx.

There was no glare or stray image reflections in the trial room. This paper emphasized on display techniques of reading a map in a combat zone and investigating visual task performance. Each monitor was showing a combination of graphical icon sizes, and participants were assigned randomly to the various screens. It was imperative for participants to have their vision corrected to normal visual acuity with certified participation approval.

3.3 Experimental Stimulus

Icons are made of commonly encountered war zone symbols, and 20 combat zone images were used as experimental stimuli, and clearly distinct colors associated with a significant meaning [18].

Each trial contained 20 icons and arbitrarily comprised of five symbols from each color. The icons on the experimentally simulated radar interface include capital letters ranging from “A”, “B”, “C” ….to “T” for image identification (see Fig. 1).

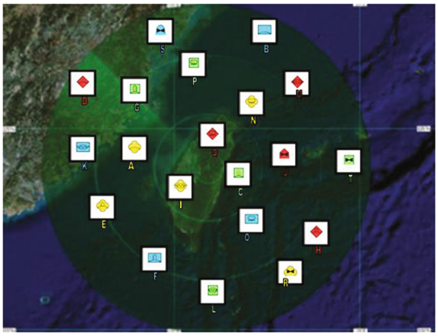


Fig. 1. Radar interface simulation

3.4 Assessment Procedures

Before starting a new search test, the system was routinely recalibrated to avoid the risk of inaccurate measurements. The ‘space button’ is pressed to display the stimuli from the radar interface simulation and participants start searching for the icon target. When the target image is found, participants were advised to verbally tell the corresponding letter of the figure and quickly press the “enter” button.

After the experiment is finished, the rate of accuracy, expressed in percentage was automatically recorded for further analysis. This study brought together eighty combinations obtained as follows: ten combinations resulted from an arrangement of five icons, and two types of map background clutter and eight distinct stimuli came from a single image size and one map background clutter in one particular stimulus.

4 Results and Discussion

The results obtained from the dependent measurement variables such as fixation, fixation duration, the amplitude of saccades, completion time and search accuracy with regards to the screen dimensions, icon size, and type of map background, are summarized in Table 1.

Table 1. Summary of the statistical results (mean values)

		Completion time (ms)	Search accuracy (%)	Fixation duration (ms)	Fixation count (times)	Saccadic amplitude (degrees)
Screen dimensions	7 in.	7374.43*	86.02	242.26	25.17*	2.31
	15 in.	8061.67	85.24	227.54	28.53	3.09
	21 in.	8612.2	91.75*	221.04	29.19	5.26
Icon size	45 min	9673.14	79.45	245.9	31.36	3.70*
	55 min	8607.21	87.87	235.01	29.23	3.82
	65 min	7649.9	88.21	226.32	27.31	3.91
	75 min	7463.54	91.69*	225.01	26.18	3.89
	85 min	6686.53*	90.99	219.18	24.08*	3.96
Map background	MTD	8006.04	86.18	230.62	27.5	N/A
	MDT	7818.54	88.82	228.65	27.42	N/A

Also, Table 1 contains the results coming from a combination of the screen dimension and icon sizes for each one of the five variants.

4.1 Time to Complete Tasks

Figure 2 shows the average completion time versus screen dimensions, types of map background and visual angle (icon size).

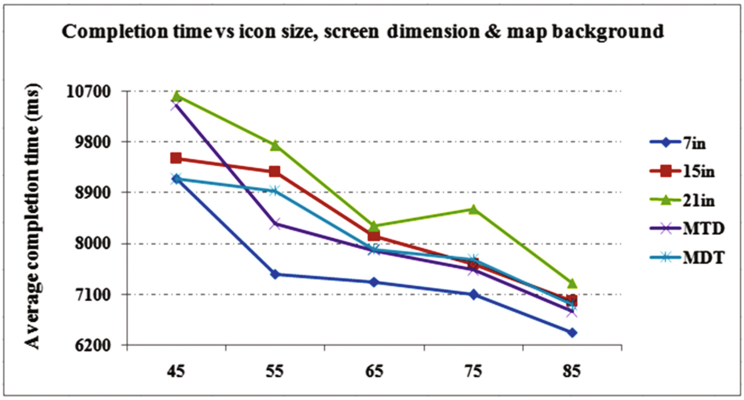


Fig. 2. Completion time vs icon size, screen dimensions & map background

The 3 × 5 × 2 ANOVA found a significant interaction between map background and sizes of screens and image ($F(8, 238) = 2.05$, $p < .05$). Besides, deeper analysis discovered a significant interaction ($F(4, 78) = 4.49$, $p < .01$) between the types of map background and the image size.

A 3 × 5 × 2 ANOVA analysis confirmed that the completion time with a 45-minute image on a 7-inch screen was significantly shorter ($F(2, 78) = 5.56$, $p < .01$), than the time recorded on 15 and 21-inch screens. For 55-minute icons, the completion time was also significantly short ($F(2, 78) = 14.45$, $p < .01$).

The completion time in MDT with 45-minute icon was significantly shorter than the one in MTD ($F(1, 78) = 9.02$, $p < .01$). There was a big difference between 21 and 15-inch screen ($p < .01$, $d = .13$) but the results remained unchanged between 7 and 15-inch screens ($p < .01$, $d = .16$).

The results revealed ($p < .01$, $d = .17$) at 75-minute icons and ($p < .01$, $d = .24$) at 65-minute icons. When images are set up to 75 and 65-minute, the time to complete tasks was not significantly different ($p > .05$). However, the completion time was significantly different between visual angles of 65 and 55 min ($p < .01$, $d = .19$) and ($p < .01$, $d = .24$) between visual angles of 55 and 45 min.

4.2 Visual Search Accuracy

A 3 × 5 × 2 ANOVA analysis revealed that the interaction effect for accuracy was significant in screen sizes ($F(2, 158) = 4.13$, $p < .05$). While combining visual angles (icon size) and types of map backgrounds, the effect of interaction on accuracy is also significant ($F(4, 78) = 4.31$, $p < .01$). Figure 3 shows that MDT presented a smaller accuracy effect at 45-minute icon ($F(1, 78) = 12.12$, $p < .01$).

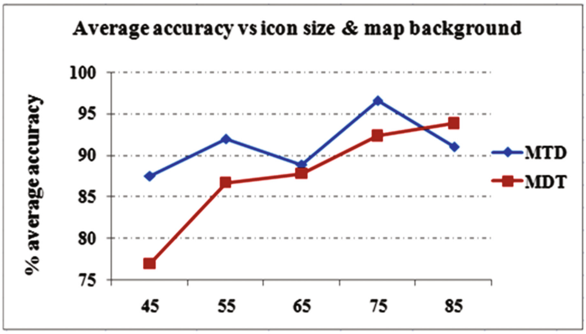


Fig. 3. Average accuracy vs icon size & map background

In Fig. 4, MDT presented a relatively lower accuracy effect when displayed on 7 and 15-inch screens than MTD. The average efficiency in the MTD was much better than the one in MDT ($F(1, 78) = 6.13$, $p < .05$) in 7-inch screen and ($F(1, 78) = 6.23$, $p < .05$) in 15-inch display.

On the other side, map backgrounds showed almost the same average accuracy on a 21-inch screen. ($F(1, 78) = 0.83$, $p > .05$). But, accuracy evaluation in the 21 and 7-inch screens was significantly different ($p < .01$, $d = 0.18$) and same scenario was observed between 21 and 15-inch screens ($p < .01$, $d = 0.22$).

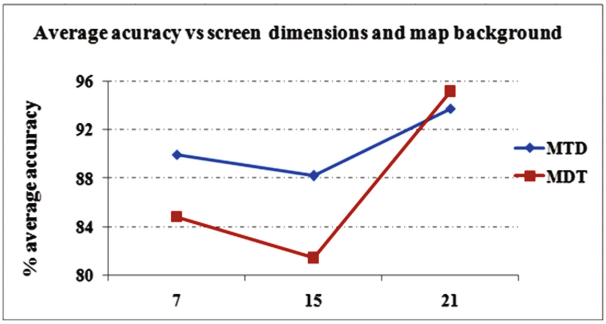


Fig. 4. Average accuracy vs screen dimensions & map background

Deeper accuracy tests confirmed there was a significant difference between 85 and 45-minute icons and that the accuracy was considerably lower with the smaller size than the one with the larger size ($p < .01$, $d = 0.31$).

4.3 Search Time and Fixation Duration

A 3 × 5 × 2 ANOVA showed that the interaction effect between screen dimensions and the size of icons was significant ($F(8, 238) = 3.48$, $p < .01$). Figure 5 shows the correlation between screen dimensions and map backgrounds for average fixation duration with regards to the icon sizes. Statistical analysis confirmed the fixation duration on 7-inch screen was significantly longer than on 15 and 21-inch screens and the results are as follows: at 45-minute icons ($F(2, 78) = 8.44$, $p < .01$), at 65-minute icons ($F(2, 78) = 26.28$, $p < .01$), at 75-minute icons ($F(2, 78) = 19.78$, $p < .01$) and at 85-minute icons ($F(2, 78) = 18.50$, $p < .01$).

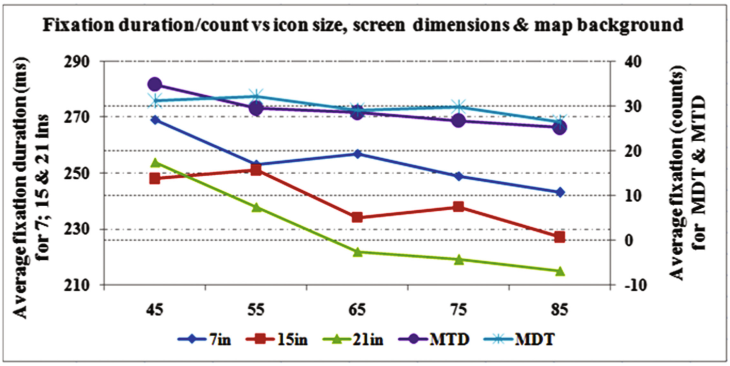


Fig. 5. Average fixation duration/count vs icon size, screen dimensions & map background

The average fixation duration only showed the significant difference at 55-minute icons with all three screen dimensions ($F(2, 78) = 2.02$, $p < .03$. For the 15-inch screen ($p < .01$, $d = 0.29$) and for a 21-inch screen, the average fixation duration ($p < .01$, $d = 0.44$) was significantly shorter than that for 7-inch screen (see Table 1).

It was also found that the fixation duration was significantly different ($p < .05$) for 85-minute images in comparison with a 55-minute icon ($p < .01$, $d = 0.22$). Also, the fixation duration was significantly different for 85-minute images as compared with 45-minute icons ($p < .01$, $d = 0.45$).

4.4 Visual Search and Fixation Count

In case of fixation count, the interaction effect between icon sizes and map background was significant ($F(4, 78) = 3.18$, $p < .01$). At 45-minute icons, the fixation count in MDT was significantly fewer than in MTD ($F(1, 78) = 5.49$, $p < .05$). Further analysis demonstrated that for 7-inch screen, the average fixation counts were significantly fewer than the fixation counts on 15-inch screen ($p < .01$, $d = 0.53$) and on 21-inch screen ($p < .01$, $d = 0.49$). The results demonstrated no significant difference between 85 and 75-minute icons ($p > .05$) or between 75 and 65-minute images ($p > .05$) and 65 and 55-minute icons ($p > .05$). However, the fixation count statistics shows a significant difference between 85 and 65-minute icons ($p < .01$, $d = 0.21$), between 75 and 55-minute icons ($p < .01$, $d = 0.21$) and between 55 and 45-minute icons ($p < .05$, $d = 0.15$).

4.5 Visual Search and Saccadic Amplitude

A $3 \times 5 \times 2$ ANOVA discovered that the interaction effect between icon sizes and screen dimensions was significant in the case of saccadic amplitude ($F(8, 238) = 3.19$, $p < .01$). The correlation between icon sizes versus screen dimensions for saccadic amplitude is displayed in Fig. 6. The saccadic amplitude increased with icon size on 15-inch display ($p < .01$, $d = 0.21$) and on a 21-inch monitor ($p < .01$, $d = 0.10$). The saccadic amplitude showed a significant difference between 65 and 45-minute images ($p < .01$, $d = 0.07$).

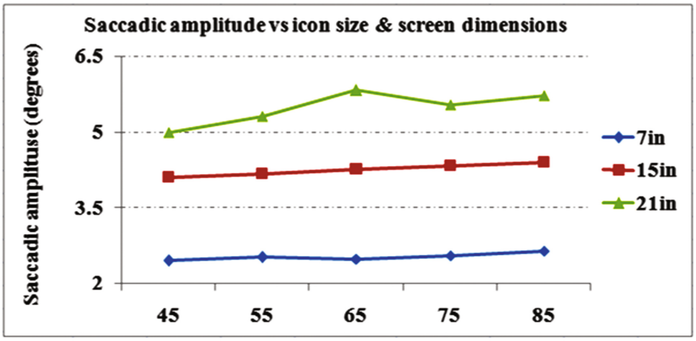


Fig. 6. Saccadic amplitude vs icon size & screen dimensions

5 Conclusion

This paper confirmed that the 85 and 75-minute images provided the best display results. Mainly, the 45-minute images performed poorly and ended up with inferior results in regards to task completion time, accuracy, fixation duration and fixation count. This study confirmed that the visual search performance highly improved with the bigger target size. Nevertheless, this article did not find any consistency associated with search time and accuracy for all screen dimensions. Besides, this paper revealed that the type of map background could influence the accuracy. The accuracy was more improved in MTD than in MDT.

Acknowledgements. The research team gratefully acknowledges the financial and academic support of the Chinese Scholarship Council in conjunction with Harbin Engineering University and the University of Rwanda.

References

1. Giordano, D., Maiorana, F., Leonardi, R.: Effects of monitor size on accuracy and time needed to detect cephalometric radiographs landmarks. Displays **33**(4), 206–213 (2012)
2. Yang, C., et al.: Analysis on eye movement indexes based on simulated flight task. In: International Conference on Engineering Psychology and Cognitive Ergonomics. Springer (2014)
3. Porathe, T.: A 3-D Nautical GIS targeting cognitive off-loading and decision making. In: Proceedings of 11th Scandinavian Research Conference on Geographical Information Science (2007)
4. Hess, E.H., Polt, J.M.: Pupil size in relation to mental activity during simple problem-solving. Science **143**(3611), 1190–1192 (1964)
5. Müller, H.J., Krummenacher, J.: Visual search and selective attention. Visual Cognition **14** (4–8), 389–410 (2006)
6. Saha, P.K., et al.: Analysis of visual search features. Int. J. Hum. Factors Model. Simul. **3**(1), 66–89 (2012)
7. Reingold, E.M., Glaholt, M.G.: Cognitive control of fixation duration in visual search: the role of extrafoveal processing. Vis. Cogn. **22**(3–4), 610–634 (2014)
8. Over, E., et al.: Coarse-to-fine eye movement strategy in visual search. Vis. Res. **47**(17), 2272–2280 (2007)
9. Einhäuser, W., Nuthmann, A.: Salient in space, salient in time: Fixation probability predicts fixation duration during natural scene viewing. J. Vis. **16**(11), 13 (2016)
10. Tseng, F.-Y., et al.: Effects of display modality on critical battlefield e-map search performance. Behav. Inf. Technol. **32**(9), 888–901 (2013)
11. Huang, K.-C.: Effects of computer icons and figure/background area ratios and color combinations on visual search performance on an LCD monitor. Displays **29**(3), 237–242 (2008)
12. Legge, G.E., Gu, Y., Luebker, A.: Efficiency of graphical perception. Percept. Psychophys. **46**(4), 365–374 (1989)
13. Burigat, S., Chittaro, L., Gabrielli, S.: Navigation techniques for small-screen devices: an evaluation on maps and web pages. Int. J. Hum Comput Stud. **66**(2), 78–97 (2008)

14. Rosenholtz, R., Li, Y., Nakano, L.: Measuring visual clutter. J. Vis. **7**(2), 17 (2007)
15. Henderson, J.M., Chanceaux, M., Smith, T.J.: The influence of clutter on real-world scene search: evidence from search efficiency and eye movements. J. Vis. **9**(1), 32 (2009)
16. Lloyd, R.: Visual search processes used in map reading. Cartogr.: Int. J. Geograph. Inf. Geovis. **34**(1), 11–32 (1997)
17. Just, M.A., Carpenter, P.A.: A theory of reading: from eye fixations to comprehension. Psychol. Rev. **87**(4), 329 (1980)
18. Thibault, D.: Commented APP-6A-military symbols for land based systems. Defence R&D Canada, September (2005)

Improving the Communication of Emergency and Disaster Information Using Visual Analytics

Chittayong Surakitbanharn(✉) and David S. Ebert

Potter Engineering Center, Purdue University, 500 Central Drive,
Room 127, West Lafayette 47901, USA
{csurakit, ebertd}@purdue.edu

Abstract. Technological advances in communication and the ubiquity of mobile devices have changed the role of emergency information during disasters. Information can now be easily shared between disaster managers, first responders, government agencies, and the public via websites and social media. However, is it unclear how the emergency information is accessed, understood, trusted, and used. This paper first reports on case studies of how disaster information is currently communicated in Queensland, Australia. Second, we propose concepts from visual analytics to improve the quality and effectiveness of communicating information. Third, we propose metrics and situation awareness experiments to measure the quality and effectiveness of new systems.

Keywords: Disaster management · Visual analytics · Systems interaction

1 Introduction

The omnipresence of smartphones and mobile internet connections has shifted the management of emergency and disaster information from a one-way distribution system into a two-way communication system. Traditional media formats – such as television, radio, and landline telephones – facilitated the one-way distribution of information. Disaster managers could broadcast updates about weather, road closures, evacuation centers, and other information, but with little feedback about the effectiveness of such measures. With little measure for effectiveness, massive search and rescue efforts were still required to ensure conformance to evacuation orders and the greater public safety [1].

The ubiquity of smartphones now allows the public not only to receive critical information but also share on-the-ground updates with disaster managers. As a result of this two-way communication, disaster managers and the public can gain situation awareness [2] and social capital (the quantity of social networks), factors which have both been attributed to improved outcomes during a disaster [3, 4].

Previous research and the common belief are that increased information flow will improve disaster response and mitigation [5]. This has led disaster agencies and local governments investing in software interfaces to communicate critical data. In Australia

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_14

and abroad, prior software interfaces were primarily targeted for to agency use [6]. However, interfaces such as the Emergency Management Dashboard (Fig. 1) are proactive steps intended to collate the breadth of data available to agencies (e.g., Bureau of Meteorology, Department of Main Roads, Energex, et cetera) and act as a visual analytic portal to communicate the information for public education and mobilization.

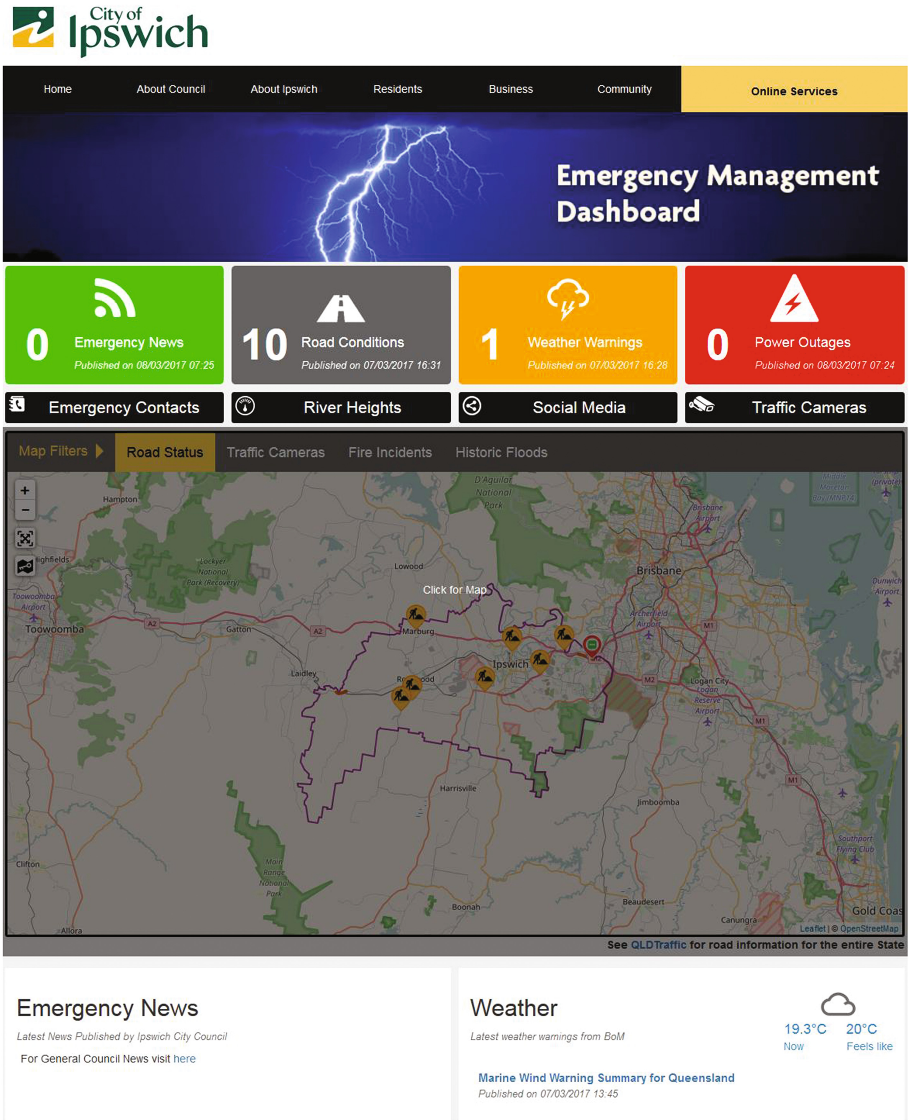


Fig. 1. Emergency Management Dashboard from Ipswich City Council on 8 March 2017.

Only in 2016, have local governments in Queensland, Australia widely adapted Emergency Dashboard systems. Due to their recent emergence, there has been limited testing in actual disaster situations about their use and effectiveness. Thus, there is currently limited knowledge about the public access, understanding, and perception of these new information systems, all factors that are vital the effective disaster management [7]. Rather than wait for next natural disaster to strike to understand and improve visual analytic Emergency Dashboards, we propose an experimental design to test the effectiveness of information systems for use in emergencies and natural disasters.

To develop our experiment, Sect. 2 of this paper surveys the roles of information and communication to increase situation awareness and build social capital at different temporal stages (mitigation and response) and describes case studies from two local governments in Queensland, Australia. Section 3 discusses Disaster Dashboard and visual analytic tools to convey information. Section 4 proposes an experimental design. Section 5 draws conclusions from this preliminary survey of disaster information.

2 The Roles of Disaster and Emergency Information

Australia's geography and sizes make regional communities particularly vulnerable to natural disasters; which are also expected to increase in frequency, intensity, and duration with ongoing climate change [8, 9]. As a recent example, in 2011 floods devasted 78% of the state of Queensland, which submerged an area greater than Germany and France combined, affected 2.5 million people in 210 towns and suburbs, damaged 29,000 homes and businesses, and resulted in 33 deaths [10].

Recent research has demonstrated the importance of mobile communications [2, 11] in providing updates about weather, road closures, evacuation routes, and other vital information to safely mobilize the public in response to a disaster. Additionally, an individual's perception of risk and danger is believed to affect how they prepare prior to disaster [12, 13] (e.g., creating an evacuation plan; compiling supplies like water, and canned food, and a torch; purchasing insurance), and decide on a course of action during a disaster (e.g., to evacuate or stay). Exposure to certain information may shape perception. For example, a household that has access to disaster frequency and severity data may make more risk-averse decisions.

There are several shortcomings with previous research in distributing disaster information. As previously mentioned, the first is a lack of user studies to determine the how the information is accessed, perceived, understood, and if it leads to the desired outcomes. The second is a lack research over the different temporal dimensions and functions (e.g., building situation awareness, creating social capital, et cetera) of disaster communications. In other words, different pieces of disaster information serve different functions at different times, but it is not clear when would be the most effective delivery of each component.

This research's overarching aim is to create a coordinated communication system that accounts for different temporal dimensions (mitigation versus response), the function of information (situation awareness versus social capital), as well as a means to measure effectiveness and outcomes. Figure 2 is a demonstrative map of the types of

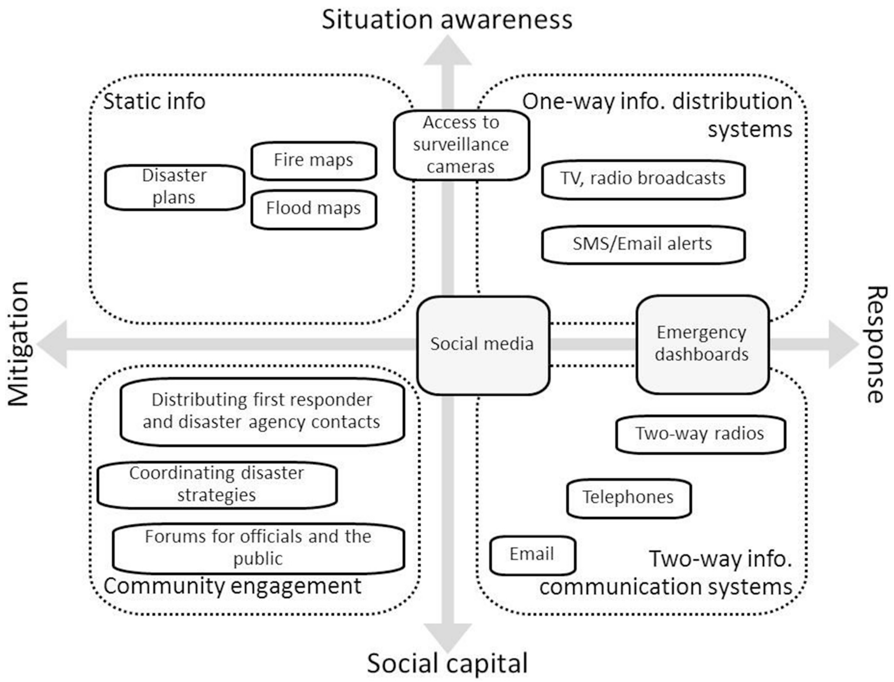


Fig. 2. Map of information types and communication strategies.

information, communication strategies, and relations to mitigation, response, situation awareness and social capital.

2.1 Situation Awareness

Situation awareness is defined by Endsley as a component for end-users to perceive and comprehend a current situation as well as project future states [14], all of which lead to a series of decisions and performance outcomes. Based on the notion of increasing situation awareness, counselors, city planners, and disaster managers believe that distributing information about can improve community decisions, and lead to the desired outcomes of greater public safety and lower financial exposure. In other words, city officials hope access to information will lead to proactive action to prepare for a disaster and better understanding by households of when they will be in danger during a disaster.

Information such as flood maps, fire maps, and disaster management plans may increase a person's situation awareness before a disaster but may have limited utility during a disaster (Fig. 2, "Static Info"). One such example is the flood maps provided by Bundaberg Regional Council (Fig. 3). The flood maps allow residents of the Bundaberg area visually explore the areas of inundation at different flood levels, and can make people aware of when they will be in danger well before a catastrophic event. By providing this information, city officials hope residents take proactive actions to mitigate the effects of natural disasters, such as purchasing flood insurances. The utility of information, such as the flood maps, is intended for use before a disaster. During a flood, actual reports and dynamic information of flooded areas will be necessary for effective response and evacuation (Fig. 2, "One-way info. distribution systems").

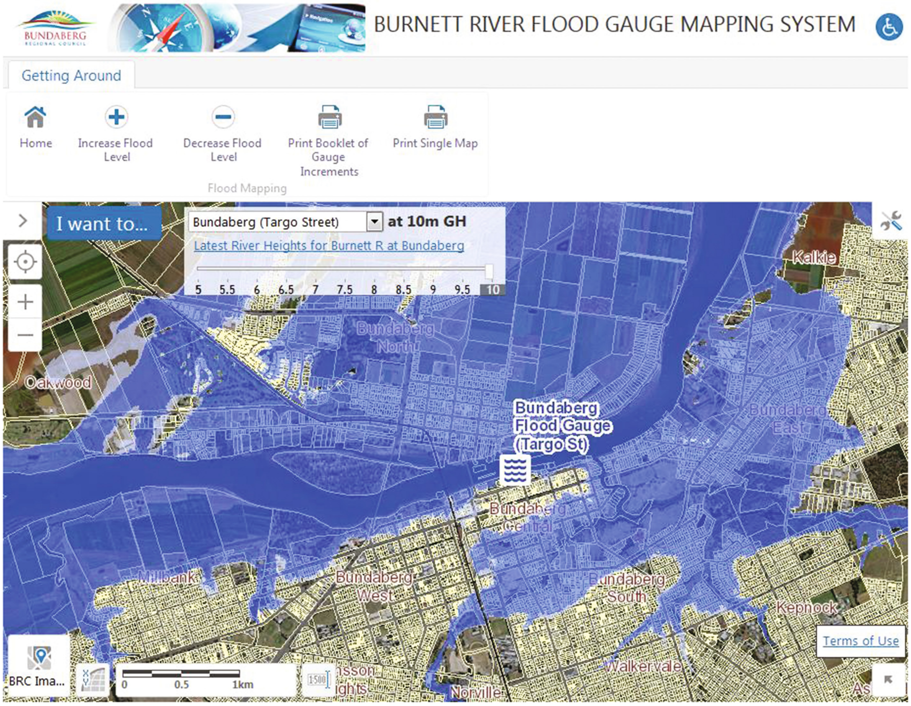


Fig. 3. Flood map of Bundaberg at 10 m of water inundation above the flood gauge. Accessed from the Bundaberg Regional Council as of 8 March 2017.

Traditional methods of distributing information as a natural disaster approaches or is occurring include TV and radio broadcasts. More recently, the same information about event intensity, evacuation routes and procedures, areas to avoid, and other imminent risks can be distributed via SMS and email alerts. Information in the response stage is constantly updating to account for changing conditions, whereas information used for mitigation generally does not incorporate an occurring event. The commonality between mitigation and response information intended to increase situation awareness is that the distribution channels are typically one way, and do not include a channel for the public to widely contribute information to disaster managers and first responders, or amongst each other.

2.2 Social Capital

Recent global research supports that social capital is the primary predictor to a community's disaster resilience [3, 4]. Social capital is a measurement of the number of social connections belonging to an individual; i.e. it is the number of people a person knows and can contact. The theory that higher social capital leads to greater disaster resilience stems from the idea that a community working together, or the public working with disaster managers, can more readily achieve a common goal compared to someone working alone.

For example, villages surrounding Mt. Merapi in Indonesia do not have access to the same technology advances and early warning systems as their Western counterparts [15]. To increase community safety, local and regional disaster managers created a sister village program to build social capital. In the program, villages in the path of Mt. Merapi's pyroclastic flow were partnered with villages residing in safe zones. The villagers in the safe zone volunteered to shelter those fleeing their homes during an eruption. Mt. Merapi is just one example of how increasing a community's social connections can increase its disaster resilience.

Australia and other Western countries also undertake activities to build social capital (Fig. 2, "Community engagement"). The mitigating stages of building social capital occurs before a disaster, such as outreach and engagement with community residents to raise disaster awareness. For example, Redland City Council in Queensland conducts workshops and designs curriculum for Grade 5 students. This future generation of residents is required to develop a disaster plan with their families as part of a class assignment. To complete the assignment, students must access council websites, disaster apps, and look up procedures in the disaster plan. Such activities attempt to put the public in direct contact with disaster managers and provide emergency contacts to the community as an attempt to lessen the distance between the public and first responders.

The response stages of building social capital include communicating via email, telephones and two-way radios. Although two-way radios and telephones aid the flow of information both to the public and back to disaster managers, they are limited to one communication channel at a time. Although these communication methods may increase situation awareness for the people receiving directly using them, they are limited in their use to increase situation awareness for the broader community. However, new technology, like social media and Emergency Dashboards, can assist in reaching a broader audience.

3 Disaster Dashboard and the Role of Visual Analytics

The Emergency Management Dashboards stemmed from prior technological needs of disaster agencies for efficient and effective emergency response, such as the Guardian software interface developed by QIT Plus. Guardian, which is used in nearly two-thirds of Queensland local governments as of 2017 [16], enabled disaster managers to collate multiple and varied sources of data, communicate among disaster agencies, and file reports to document decisions and events. Redland City Council disaster managers reported using the Guardian software interface to manage the response efforts to bushfires on Russell Island during December 2016 [17]. An interview with the disaster managers indicated that the use of Guardian enabled perceived higher levels of situation awareness and social capital, which led to a more coordinated response and efficient use of resources. The increased situation awareness was achieved through a collation of reports from first-responders on-the-ground and managing agencies that highlighted which personnel was needed at particular times to combat the bushfires and accommodate evacuees. The increased communication and connections among agencies, which may be interpreted as an increase social capital, resulted in disaster managers coordinating police, fire, and ambulances to share a transport ferry to Russell Island rather than each unit having to arrange their own transport. Since a software interface demonstrated utility in collating information and coordinating stakeholders at the agency level, the developers of Guardian aimed to test the utility of conveying similar information to the public, which led to the development of the Emergency Management Dashboards.

The Emergency Management Dashboard takes much of the information accessible by disaster managers and agencies and distributes it to the public with the intent of improving decision making and situational outcomes. Current Emergency Management Dashboards visualize data (Fig. 1) through reporting and mapping weather, road closures, and power outages. Although there are links to example to examine history flood heights, the data visualization may be most useful in increasing situation awareness in disaster response. There are two points in the Emergency Management Dashboard example that increase social capital in addition to situation awareness: emergency contacts and the social media feed.

However, a limitation with current social media feeds is difficulty for both disaster managers and users to verify information authenticity and accuracy. For example, one Tweet may report a flooded road, but there is no way for disaster managers or first responders to remotely verify this information without a picture or other visual indicator. A visual analytics tool like the Social Media Analytic Reporting Toolkit (SMART) (Fig. 4) may be one way to verify the accuracy of social media reports [11].

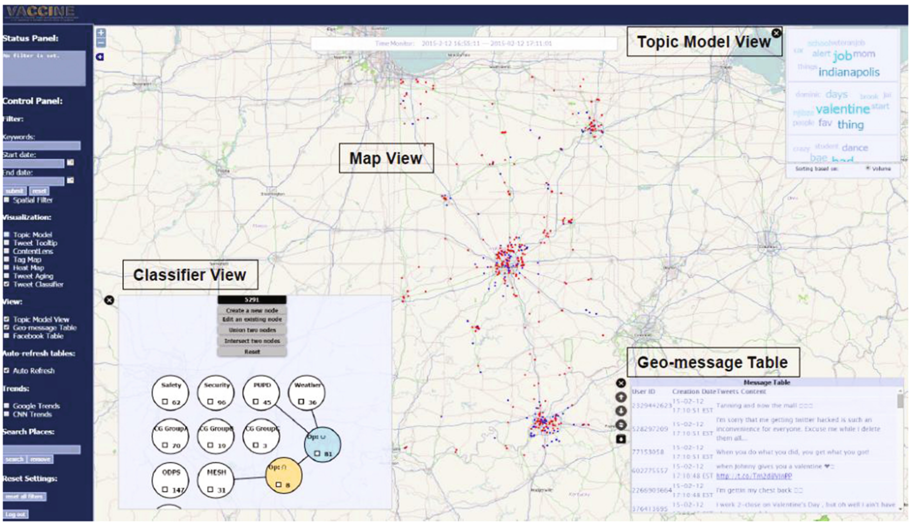


Fig. 4. Screenshot of Social Media Analytic Reporting Toolkit.

SMART geomaps the amalgam of Tweets, Instagram posts, and can incorporate other social media feeds displaying trending topics and keywords on a map, which enables several capabilities. First, the geomap can verify social media reports for disaster managers and agencies. Several reports about a flooded road may go unnoticed, but a series of Tweets or Instagram posts may enable a disaster manager's situation awareness and warrant action. Second, the geomap may reduce workload. Instead of having to sift through several social media feeds post-by-post manually, SMART can map the information and summarize major trends. Third, if a version of the SMART were to be provided to the public, people could use it to locate and learn from other people in similar situations from completely different backgrounds. Essentially, the SMART visual analytics may enable social capital.

There are several considerations for the Emergency Management Dashboard and other software interfaces to run and be effective. First, software interfaces, the Emergency Management Dashboard, and social media rely on a function network architecture, which may not always be available during a natural disaster. Second, there is room to improve the integration of both Guardian and Emergency Management Dashboard software interfaces with other data management systems used by local governments. Third, there is little widespread quantitative information about the effectiveness of such measures, which we aim to address through proposing an experimental design.

4 Experimental Design

Regardless of which system is implemented, the effects on the broader community remain unknown without proper testing. Emergency systems and disaster information may be improved through an understanding of who is using the Emergency Dashboard and other mediums of information, how the information is being accessed, and what measurable effects it has at the individual level and the community level. The following is an experimental design to test for these factors.

The first step is determining the probabilistic relationship between the desired actions of individuals (preparing for a disaster, evacuating at the right time) and exposure to disaster information. Economic theories can model the likelihood that an individual's perception and action changes based on exposure to information. These theories range from expected utility theory [18], where people are assumed to act in their best interest, to prospect theory [19, 20], where people are not always acting in their best interest. The first step attempts to answer: when people are given information, do they use it in the manner which disaster managers intend?

The second step is to test the relationship between how local governments and agencies currently distribute information and how the community accesses information. For example, web analytics data can show whether outreach on social media or in newspaper advertisements results in more people accessing the disaster information.

The third step is to test how users interact with, understand, and perceive the data. Engagement and use experience with Emergency Dashboard web pages and mobile apps can be measured with eye tracker studies using volunteers in a lab setting and from the local community. Querying methods such as Situation Awareness Global

Assessment Technique [14] can determine a user's comprehension of the information provided. The research findings and participant feedback can be incorporated into the development of a two-way communication system.

The user response and interaction testing may result in inconclusive differences in user experience or situation awareness for certain variables. For example, user tests may demonstrate no difference in displaying the Transport and Main Roads information before or after the Bureau of Meteorology information. In such cases, different formats of the Emergency Dashboard can be randomly assigned to different councils participating in the study. This random controlled test allows differences in how disaster information is distributed to be judged over time.

Finally, the contribution of the Emergency Dashboard to social changes can be measured using social capital, census, and other datasets. Longitudinal analysis of datasets that measure social capital and household exposure to disasters, such as the Household, Income and Labour Dynamics in Australia (HILDA) Survey and Australian National Development Index (ANDI), can changes to social capital over time. Additional performance metrics, such as community evacuation time, can be set before a disaster, measured during the event, and analyzed afterward. The combination of social capital and situation awareness measures can validate the probabilistic relationship to quantify the long-term societal impact of communicating crisis information.

5 Conclusion

This paper addressed four main points of current and future research for communicating disaster and emergency information. First, we introduced a map of how different information types and communication strategies relate to disaster mitigation, disaster response, situation awareness, and social capital. Second, the paper used case studies from Queensland to describe how current actions relate information types and communication strategies map. Third, we described how the use of visual analytics could increase the use of both situation awareness and social capital. Finally, the paper proposes an experimental design to test the effectiveness of current and future communication technologies. Through addressing these points, new visual analytics technologies can be integrated to current communication strategies to improve disaster management, mitigation, and response.

Acknowledgements. This work was partially funded by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0003. The authors would like to thank the disaster managers at Bundaberg Regional Council, Moreton Bay Regional Council, Redland City Council, and QIT Plus for their ongoing input and contribution towards research to improve the communication of disaster and emergency information.

References

1. Poteyeva, M., Denver, M., Barsky, L.E., Aguirre, B.E.: Search and rescue activities in disasters. Handbook of Disaster Research, pp. 200–216. Springer, New York (2007)
2. Yin, J., Lampert, A., Cameron, M., Robinson, B., Power, R.: Using social media to enhance emergency situation awareness. IEEE Intell. Syst. **27**(6), 52–59 (2012)
3. Aldrich, D.P., Meyer, M.A.: Social capital and community resilience. Am. Behav. Sci. **59**(2), 254–269 (2015)
4. Aldrich, D.P.: Building Resilience: Social Capital in Post-disaster Recovery. University of Chicago Press, Chicago (2012)
5. Kim, S.Y., Jang, Y., Mellema, A., Ebert, D.S., Collinss, T.: Visual analytics on mobile devices for emergency response. In: 2007 IEEE Symposium on Visual Analytics Science and Technology, pp. 35–42 (2007)
6. Saleem, K., Luis, S., Deng, Y., Chen, S.-C., Hristidis, V., Li, T.: Towards a business continuity information network for rapid disaster recovery. In: Proceedings of the 2008 International Conference on Digital Government Research, Montreal, Canada (2008)
7. Popoola, A., Krasnoshtan, D., Toth, A.-P., Naroditskiy, V., Castillo, C., Meier, P., Rahwan, I.: Information verification during natural disasters. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1029–1032. ACM (2013)
8. Howes, M., Tangney, P., Reis, K., Grant-Smith, D., Heazle, M., Bosomworth, K., Burton, P.: Towards networked governance: improving interagency communication and collaboration for disaster risk management and climate change adaptation in Australia. J. Environ. Plan. Manag. **58**(5), 757–776 (2015)
9. Aldrich, D.P.: Between market and state: directions in social science research on disaster. Perspect. Polit. **9**(1), 61–68 (2011)
10. Queensland floods commission of inquiry: Final report (2012)
11. Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D.S.: Public behavior response analysis in disaster events utilizing visual analytics of microblog data. Comput. Graph. **38**, 51–60 (2014)
12. Miceli, R., Sotgiu, I., Settanni, M.: Disaster preparedness and perception of flood risk: a study in an alpine valley in Italy. J. Environ. Psychol. **28**(2), 164–173 (2008)
13. Kunreuther, H.: Disaster mitigation and insurance: learning from Katrina. Ann. Am. Acad. Polit. Soc. Sci. **604**(1), 208–227 (2006)
14. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, vol. 783, pp. 789–795 (1988)
15. Surakitbanharn, C.: Soft mitigation in areas of reoccurring natural disasters: floods in Queensland, Australia and the volcanic eruptions of Mt. Merapi in Yogyakarta, Indonesia. In: Proceedings of Fourth International Conference on Sustainable Built Environment, Yogyakarta, Indonesia, 12–14 October 2016
16. Madsen, C.: Guardian Site List (2017)
17. Lollback, M., Tait, M.: Interview with Redland City Council Disaster Managers (2017)
18. Kunreuther, H.: Mitigating disaster losses through insurance. J. Risk Uncertain. **12**(2), 171–187 (1996)
19. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–291 (1979)
20. Asgary, A., Levy, J.: A review of the implications of prospect theory for natural hazards and disaster planning (2009)

System Interaction in Mixed Reality Environments and Simulation

Augmented Reality in Support of Disaster Response

Isabel L. Nunes[1,2(✉)], Raquel Lucas[1,3], Mário Simões-Marques[4], and Nuno Correia[1,3]

[1] Faculdade de Ciencias e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
{imn,nmc}@fct.unl.pt, rv.lucas@campus.fct.unl.pt
[2] Faculdade de Ciencias e Tecnologia, UNIDEMI, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
[3] NOVA-LINCS, Faculdade de Ciencias e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
[4] CINAV- Portuguese Navy, Alfeite, 2810-001 Almada, Portugal
mj.simoes.marques@gmail.com

Abstract. Disaster Management is a complex process, usually dealing with a large amount of uncertain, incomplete and vague information, which normally requires the coordination and collaboration among a variety of actors. THEMIS (*disTributed Holistic Emergency Management Intelligent System*) is designed as an intelligent system aimed at supporting real time disaster management activities of decision-makers in command posts, and responders in the field. It gathers information from multiple sources (e.g., users, sensors, crowdsourcing), and provides situational awareness based on a georeferenced common picture which is shared among system users This paper presents the preliminary work developed in the context of the THEMIS project addressing the use of Augmented Reality by first responders in a context of disasters relief operations.

Keywords: Mobile augmented reality · Disaster response · THEMIS

1 Introduction

Disaster Management is a complex process, usually dealing with a large amount of uncertain, incomplete and vague information, which normally requires the coordination and collaboration among a variety of actors. Disaster is defined as "a serious disruption of the functioning of a community or a society involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources."

Literature notes that the human society evolved creating the stage for novel types of crises, which also require different approaches when compared with traditional crisis management. Dealing with this novelty calls, among other capabilities, for: (i) expertise in crisis management; (ii) flexible and highly adaptive response teams, operating in multiagency multipurpose networks; (iii) standard operating procedures; and (iv) similar

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_15

tools and protocols, which could be utilized for interagency cooperation in disasters management considering multiple scenarios.

First responders, in particular, require support and guidance to perform their tasks in disaster response situations specially when operating in previously unknown locations. In large disasters, like in Haiti and Japan earthquakes is was observed that the mobile communications infrastructure often remains operational, opening room to use it in support of the disaster relief operations. However, first responders, have quite specific user needs and equipment operation limitations that must be attended while selecting the user terminal types and designing the interaction.

Augmented reality (AR) is a specific solution for reaching mediated reality environments where georeferenced and context specific computer-generated information elements (e.g., text, sound, graphics) complement the view and perception of the physical real-world. New examples of application of this enhancement of the real world's perception emerge every day, some serious (such as heads up displays to assist aircraft pilots or automobile drivers) and others for entertainment purposes (such as games and support to tourism). Among these examples, it is possible to find some recent applications in the field of emergency management and disaster relief operations.

The paper presents the preliminary work developed in the context of the THEMIS project addressing the use of AR by first responders in a context of disasters relief operations. THEMIS stands for *disTributed Holistic Emergency Management Intelligent System*, which is a project funded by the Portuguese Ministry of Defense whose components and goals are further detailed in [1]. The paper discusses the design options for the development of AR applications to be used by first responders in disaster relief operations, analyzing the context of use, user needs, organizational requirements and interaction limitations, addressing also the process planned for solution validation and evaluation, according to the User Centered Design methodology.

2 Augmented Reality

Augmented Reality allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. AR systems can be classified according to the type of display used [2], involving optical or video technology.

According to Kiner and Zorzal [3], it is possible to classify augmented reality into four types of systems:

Direct optical vision system - this system uses glasses or helmets with lenses that allow the direct reception of the real image, while allowing the projection of virtual images properly adjusted to the real scene.

Direct video vision system - this system uses helmets with coupled video cameras. The real environment is captured by the micro-cameras and is mixed with virtual elements generated by a computer. It is directly displayed in the user's eyes through small monitors mounted on the helmet.

Monitor based video vision system - this is the most commonly used system and it uses a webcam to capture the real environment. Once captured, the actual environment is mixed with the computer generated virtual elements and the result is displayed on the monitor. The view of the user usually depends on the positioning of the webcam.

Projection optical vision system - this system projects images of virtual objects onto surfaces of the real environment. The result is presented to the user without the need to use any type of device.

In the current system, we will use the direct video vision option mostly using a smartphone display.

In mobile AR applications such as the ones being developed, information visualization and interaction are key to enhance user experience. Due to the small screen, the limited angle and the short distance seen through the camera, many points of interest are off-screen, making it difficult to quickly find the information you need. It is then necessary to find ways to represent and search for off screen objects.

The following three techniques are the most commonly used to represent off-screen objects in AR applications on mobile devices:

Mini map - this technique consists of placing a representation of the user's location and points of interest around them so it is possible to see where objects are not appearing on the screen of the mobile device. This technique, although accurate, is difficult to interpret since it is necessary for the user to align with the reference systems.

3D Arrows - This technique consists of using 3D arrows to direct to points of interest. The size of the arrow is proportional to the distance the point of interest is. The main disadvantage of using this technique happens when there are many points of interest nearby as it can get too confusing.

Sidebars - This technique consists of having two sidebars to represent off-screen objects. It is also possible to integrate information about the distance, direction and type of the point of interest that is off the screen, allowing users to quickly identify relevant information.

All of these techniques are relevant and are under evaluation to identify which ones to include in the different interfaces made available for users' interaction.

3 Survey on AR Applications in Support of Emergency Management

A very brief survey identified several projects addressing the use of AR in support of emergency management related activities or presenting some features that may be useful considering the goals of the THEMIS project. Such studies/systems are described below.

Augview [4] – the authors describe this mobile application as an AR geographical information system that allows users to visualize and edit data in real time. The application uses AR for virtually visualizing 3D models of underground objects (e.g., cables, pipes) based on public data or on corporate data. The authors claim that the system provides a better understanding of the surrounding environment.

Augmented reality system for earthquake disaster response [5] - this prototype system aims to encourage the use of AR in the field of disaster management, and the target group are experts of Search and Rescue. The system was developed as a specialized equipment for supporting rescuers trying to find people trapped in the rubble of collapsed buildings using specialized equipment. The authors claim that the system is particularly suited to transmit knowledge between specialists from different areas who must work together. The disaster management tool can also be used as an experimental environment in which is possible to test different disaster prevention methods and plan what to do in case of disaster. The system is presented as allowing a more detailed vision, representing different disaster simulation information (e.g., people buried in debris, damage simulations and measures) and overlapping them on the real environment. The access to system information is meant to use either a display or a helmet mounted retinal display.

Augmented reality for fire and emergency services [6] – a system proposal to improve collaboration between different teams involved in a fire emergency, through AR wearable devices where the interactions would be done through gestures or touch. The envisaged system would be connected to control rooms and mobile users, allowing the centralized parts to have a better decision taking ability due to the access of important information. Using such AR system, field operators would be capable of identifying dangerous locations, areas that were already searched and areas that require immediate attention. It would also be possible to edit the 3D representation of buildings to identify damaged parts. The authors claim that using such system a control center could direct field operators to investigate certain areas by targeting locations on a 3D map which would be seen by the operators through the AR component of the system. The system would also permit an easy recording of data on the spot, and such recordings could be used to make simulations and to train people for emergency situations.

SAFE (*Smart Augmented Fields for Emergency)* [7] - the authors describe first explorations towards a distributed collaborative system based on the integration of wearable computing and AR technologies with intelligent agents and multi-agent systems, with the goal of helping rescuer teams and operators involved in victim rescue missions and improving interaction and coordination with other team members. The authors envisage the possibility of control room operators interacting with the application through a map that provides a field vision; and of rescuer actions (e.g., identify a patient and perform a screening) being assisted by intelligent agents. These agents would use context data (e.g., rescuer location, real time measure of patient vitals, environment state) to guide and ease the mission of the rescuer through an AR interface.

Training emergency responders through AR mobile interfaces [8] - The proposed system objective is to allow emergency response team members to explore AR interaction and familiarize with the mobile technology used in emergency response. The authors envisage two user categories: the designer of the simulation scenarios and the team members. The designer needs to have previous experience in emergency management and has the task to create the different stages of simulation. The team members will learn the system functionalities and the protocols needed to manage emergencies. The designer can insert spatial and temporal data describing the evolution of the events through time. During the simulation, team members can perceive the caused damage through the application that superimposes the simulation information to the real environment. The authors claim that such system would improve trainees' preparedness for emergency situations.

4 Augmented Reality on THEMIS Project

THEMIS is aimed at supporting real time disaster management activities of decision-makers in command posts, and responders in the field. It is designed as an intelligent system that gathers information from multiple sources (e.g., users, sensors, crowdsourcing), and provides situational awareness based on a georeferenced common picture which is shared among system users [1]. A high level conceptual perspective is presented in Fig. 1 considering the scenario of a major disaster requiring international humanitarian assistance, provided by multiple agencies. This study addresses the implementation of an AR solution to equip emergency responders, as highlighted on the referred figure.

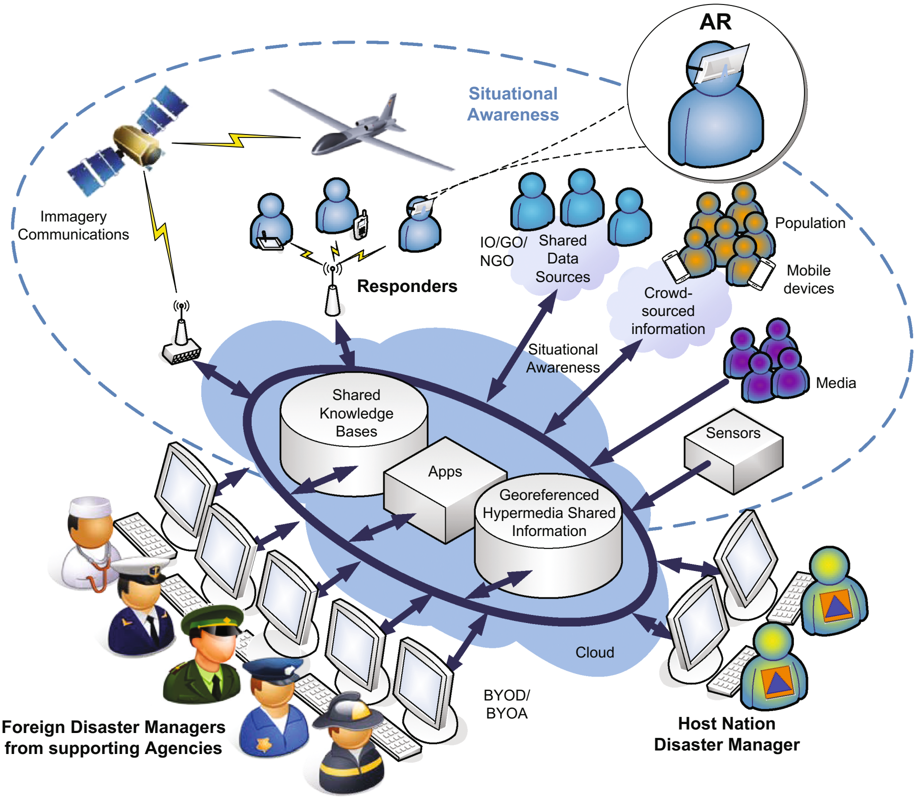


Fig. 1. THEMIS high level conceptual view (adapted from [1])

The process adopted for the developing the AR application within the THEMIS project is depicted on Fig. 2. The development process starts with the definition of the context of use, followed by the specification of the users, their needs and the organizational requirements to implement in the solution. The implementation (including the design and prototyping phases) is iteratively assessed in terms of compliance with the requirements and the quality of interactions, testing their usability. The identification of inadequacies implies the revision of design solutions and, eventually, of the requirements.

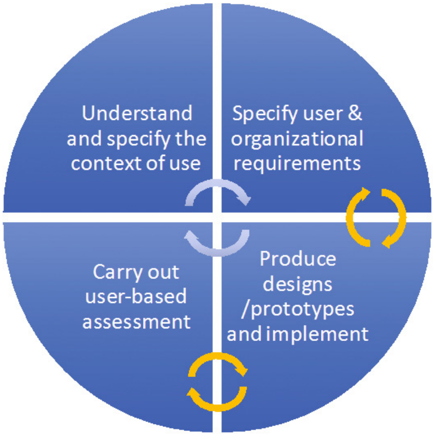


Fig. 2. Development process cycle

4.1 Context of Use, Users and Organizational Requirements

The context of use for the AR applications regards an environment of emergency response, where users are deployed integrated in a structured organization composed by teams (e.g., management, search, rescue, medical, logistics) with specific task profiles. There are guidelines established by international organizations (such as INSARAG Guidelines [9]) that set some level of standardization in this domain. For instance, according to these guidelines [9]: (i) the Management Team performs the coordination, planning and assessment of the operations of the other teams of its unit, and the coordination with the authorities and other units in the same area of operations; (ii) the Medical Team performs, among other tasks, casualty handover and transport, fatality management, and medical waste disposal, and also integrates with the Rescue Team to provide emergency medical care to entrapped victims and to ensure safe patient extrication; and (iii) the Rescue Team performs rescue (extraction and transport) in collapsed or failed structures in close operation with Search and Medical teams.

The picture in Fig. 3 illustrates the activity of a Management Team from a Portuguese Navy ship crew deployed ashore and engaged in a disaster relief training exercise (DISTEX - Disaster Exercise). The picture in Fig. 4 illustrates the cooperation among Medical and Rescue teams in the same disaster relief exercise. Comparing the activities presented in these pictures it is immediately perceptible that the ability to interact with information systems is quite different in the two contexts.

In fact, while management team members perform tasks that are compatible with the use of conventional computers (i.e., touch or keyboard-mouse interface hardware),

Fig. 3. Management team coordinating disaster relief operations in a DISTEX

Fig. 4. Medical and rescue teams performing disaster relief operations in a DISTEX

having no significant limitations on the interaction with these types of equipment, for the members of the other teams the reality is quite different. Due to the nature of their tasks most members of these teams need to use both hands, which makes difficult holding and manually interacting with information systems' equipment. Therefore, one possible solution is to exploit the use of AR devices that do not impair the vision but which provide support to dynamic situational awareness, guidance on actions and navigation.

Thus, this research intends to create an AR system based on mobile terminals (e.g., goggles, smart phones, mini-tablets) which are used:

- for data (and optionally voice) communications;
- to access georeferenced elements of information in the cloud (namely from a database fed by reports from different users and data sources of the emergency management system);
- to present general or filtered information according to the user profile, and to users' needs and/or preferences.

The AR interface must provide information to the user superimposing to the direct (optical or video) vision a combination of symbols (e.g., emergency related icons), alphanumeric information (e.g., incident identification, description, location) and spatial orientation support (e.g., direction arrows, scales, compass, "radar plots");

Any relevant georeferenced information must be presented synchronized/aligned with the user's field of view.

The interaction with objects (e.g., hyperlinks, buttons, menus) must be simple and easy to learn and (when applicable) involve standard gestures (e.g., tapping any number of taps), pinning in and out, panning or dragging, swiping (in any direction), rotating (fingers moving in opposite directions), long press (also known as "touch and hold");

The interface must allow the entry of data regarding relevant information elements (not necessarily related with the AR environment), and the configuration of parameters, for instance related with the user tasks (according to a list of predefined organizational profiles).

4.2 Implementation

The initial steps of this stage involve selecting the technological platform and implementation tools to use in designing, prototyping and developing the application, as well as detailing the objects, behaviors and interactions to implement.

AR Development Platform Selection. Regarding AR development platforms, there are many tools available to build AR applications for mobile devices. The choice for a specific development tool depends on the design requirements, since the features provided by the tools vary significantly. Therefore, a detailed analysis of alternative AR tools and platforms was necessary to identify the one that better complies with the above-mentioned THEMIS project requirements. Table 1 identifies the main platforms considered in this analysis, presenting their generic characteristics.

Besides compliance with the requirements, the selection process also considered the quality of the documentation and support available for the development of many functionalities that were identified for the project. After pondering all the factors (e.g., documentation and geolocation support) the platform selected as appropriate for the project development was Wikitude.

Table 1. AR tools considered in the selection process for THEMIS project.

Platform	Characterization	Ref.
BeyondAR	Is a platform that allows the geolocation of objects in an AR world, designed to simplify the creation of 3D objects and their visualization through a devices' camera	[10]
Vuforia	Is a software development kit (SDK) for AR that offers multiple resources useful in the creation of AR applications. Using this tool, it is possible to create virtual buttons, detect multiple targets, recognize and track planar images and simple 3D objects (e.g., boxes) in real time. Among other features, Vuforia supports multiple target detection of objects, images or text. This tool is mainly used for image recognition and allows the application to use local or cloud databases	[11]
Layar	Allows the creation of AR applications that are based on vision or location features	[12]
Wikitude	This SDK supports target recognition and tracking, model rendering in 3D and animations, video overlapping, localization based tracking, among other functionalities	[13]
ARToolKit	Is a library that provides tracking functionalities for AR applications. This tool supports video and optical AR, and allows calculating the position and orientation of the camera in relation to square shapes or flat surfaces, allowing the overlapping of several virtual objects	[14]

Design Studies. The early steps of the development process followed the generic principles of scenario-based design, which is a lightweight user-centered method that uses narrative descriptions for envisioning future use possibilities, for instance using sketches to capture the essence of interactions [15]. Figure 5 illustrates a scenario considering a typical disaster relief operation where a first responder is confronted with a situation where multiple entities were identified, signaled in the image by icons, for instance, of people (e.g., injured, dislodged, other responders), physical facilities that ensure some services (e.g., shelter, food, hospital), perils (e.g., NBCD), or resources (e.g., transportation). The first responder is assigned to deal with a specific incident, corresponding to assist an injured individual (marked with a circle). The scenario image considers that the responder has an AR device to support his actions, and the dotted arc represents his field of view. In this case, the assignment is out of the responder field of view. Based on scenarios of this nature, preliminary identification of the design and interaction requirements were drawn.

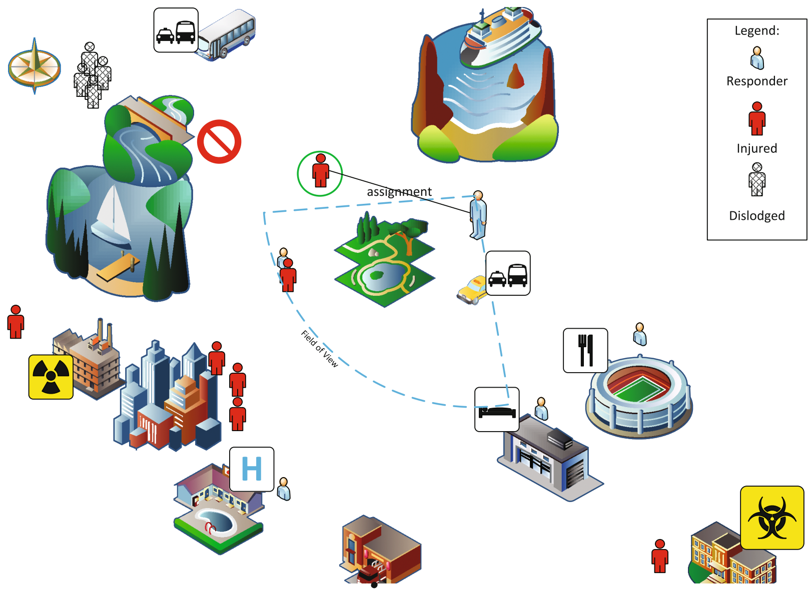


Fig. 5. Typical disaster relief operations scenario

Figure 6 presents a possible design for a simple AR interface where synthetic information is overlapped on the direct video from the camera. This information is composed by icons and textual labels that provide additional data (e.g., description, location). Besides this, the interface also includes geographical orientation information.

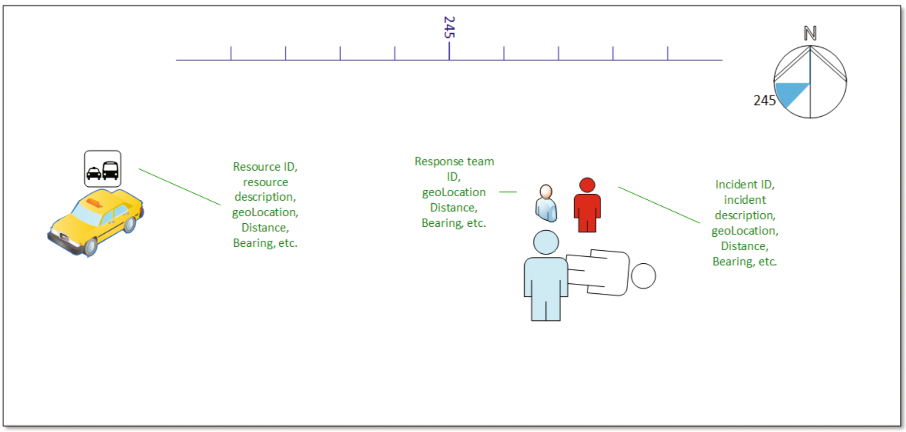


Fig. 6. Interface sketch superimposing icons and labels to an image acquired by the camera

Figure 7 illustrates a more elaborated AR interface which includes additional information besides the one considered in Fig. 6. A panel containing text instructions is shown on the left side. A radar inspired (distance-bearing) presentation of synthetic data (icons) offers situational awareness regarding the incidents in the neighborhood. The rings represent distance ranges. The dotted vector in the center bottom of the screen points the location of the assigned response, providing distance and bearing information regarding this target incident. The interface also provides symbolic and textual information regarding the target incident, as well as orientation using arrows on the left and right of the heading bar (top of screen) providing direction hints towards this incident.

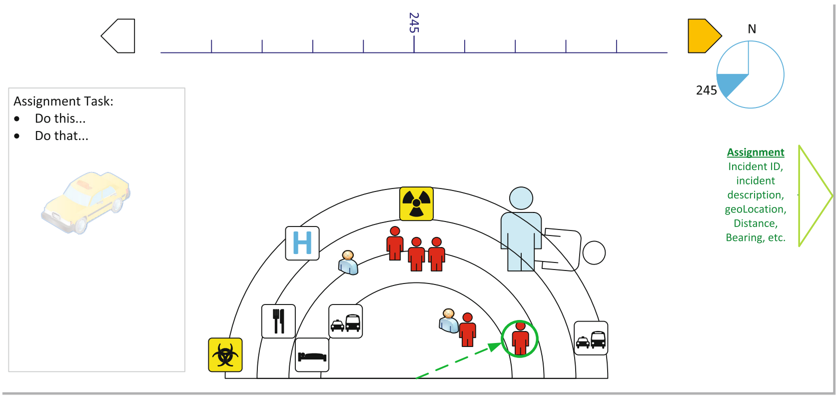


Fig. 7. Interface sketch superimposing additional context and navigation aids to the direct view

4.3 Implementation Assessment

As mentioned above the development and evaluation phases will be done in several iterations with the collaboration of users from the domain. From the initial requirements and brainstorming sessions, a first version of the user interface is being developed allowing for early feedback and assessment. The information system that supports this and other applications is also being developed in parallel, fact that is being taken into account for the AR component. The intent is to do formative evaluations with users based on early prototypes, and move towards qualitative evaluations when a more stabilized version of the user interface is reached and the integration with the underlying information system is achieved. More specific evaluations of AR behaviors will also be carried out to assess characteristics such as comfort of use and adequate behavior regarding information visualization. Whenever possible user evaluation will be done in context, as the environment plays an important role in defining the intended behavior of the application.

5 Conclusions

The paper addressed the implementation of an AR solution to equip emergency responders in the context of the THEMIS project, which is an intelligent system aimed at supporting real time disaster management activities. The system gathers information from multiple sources, provides situational awareness by means of a georeferenced common picture, and offers advice regarding alternative courses of action. The typical usage scenario is the response to major disasters requiring international humanitarian assistance, provided by multiple agencies. The paper presents a perspective on the work already developed, related with characterization of the context of use, user needs and organizational requirements, as well as the scenario-based design approach, which is a method that uses narrative descriptions for envisioning future use possibilities, for instance using sketches to capture the essence of interactions. It also offers some perspective on the way ahead, regarding the iterative process of implementation and assessment.

Acknowledgments. The work was funded by the Portuguese Ministry of Defense and by the Portuguese Navy.

References

1. Simões-Marques, M., Correia, A., Teodoro, M., Nunes, I.L.: Empirical studies in user experience of an emergency management system. In: Advances in Human Factors and System Interactions - Proceedings of the AHFE 2017 Conference on Human Factors and System Interactions, Springer, Los Angeles, 17–21 July 2017 (2017, to be published)
2. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graph. Appl. **21**(6), 34–47 (2001)
3. Kirner, C., Zorzal, E. R.: Aplicações educacionais em ambientes colaborativos com realidade aumentada. In: Anais do Simpósio Brasileiro de Informática na Educação, vol. 1, no. 1, pp. 114–124 (2005)
4. Augview. http://www.augview.net. Accessed Jan 2017
5. Leebmann J.: An augmented reality system for earthquake disaster response. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34, Part XXX (2004)
6. Wani, A.R., Shabir, S., Naaz, R.: Augmented reality for fire and emergency services. In: International Conference on Recent Trends in Communication and Computer Networks, Hyderabad, India (2013)
7. Brunetti, P., Croatti, A., Ricci, A., Viroli, M.: Smart augmented fields for emergency operations. Proc. Comput. Sci. **63**, 392–399 (2015)
8. Sebillo, M., Vitiello, G., Paolino, L., Ginige, A.: Training emergency responders through augmented reality mobile interfaces. In: Multimedia Tools and Applications, pp. 1–14 (2015)
9. UN OCHA: INSARAG guidelines. United Nations Office for the Coordination of Humanitarian Affairs (2015)
10. BeyondAR. http://beyondar.com/home. Accessed Jan 2017
11. Vuforia. https://www.vuforia.com. Accessed Jan 2017

12. Layar. https://www.layar.com. Accessed Jan 2017
13. Wikitude. http://www.wikitude.com. Accessed Jan 2017
14. Artoolkit. http://artoolkit.org. Accessed Jan 2017
15. Rosson, M.B., Carroll, J.M.: Scenario-based design. In: Jacko, J., Sears, A. (eds.) The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 1032–1050. Lawrence Erlbaum Associates, Mahwah (2002)

Virtual Environments Integrative Design – From Human-in-the-Loop to Bio-Cyber-Physical Systems

Didier Fass[1(✉)] and Franck Gechter[2]

[1] ICN Business School and Mosel Loria UMR CNRS 7503, University of Lorraine, Nancy, France
didier.fass@loria.fr

[2] LE2I-UTBM, Univ. Bourgogne Franche-Comté, Belfort, France
franck.gechter@utbm.fr

Abstract. The use of technologies such as Virtual Reality (VR) and Augmented Reality (AR) is more and more widespread in our everyday life and is now mature enough to be used in critical medical or aerospace applications. The main issue is then to provide more adaptive abilities to VR/AR applications in developing them using human concerns as primary element. To that way, beyond the classical HITL concept this paper proposes an integrative point of view for virtual environment design based on the concept of Bio-CPS. The goal of this paper is to present Bio-CPS related concepts and to sketch some guidelines for future virtual environments design focusing on specific application fields.

Keywords: Virtual-environment · Virtual Reality · Augmented reality integrative design · Bio-Cyber-Physical Systems

1 Introduction

A virtual revolution is underway in the use of technologies such as Virtual Reality (VR) and Augmented Reality (AR) in our everyday life. Besides the obvious interest of these technologies for entertainment, they also have a huge potential for medical, aerospace, military and education purposes.

From now on, the next challenge is to provide more adaptive abilities to VR/AR applications and improve the skills of practitioners/designers for users' benefits. This progress expected progress requires the use of specific sensors/effectors aimed at observing and stimulating the user. The addition of these requires a rethinking of the design/engineering of human/machine systems with an integrative point of view. Indeed, as opposed to the classical human-in-the-loop (HITL) concept where the artificial system, considered as a Cyber-Physical System (CPS), is developed separately and relies on the sole adaptation ability of the human, the work presented in this paper is aimed at developing new solutions for human system integration (HIS). These will include considerations on the different natures and organizations of the involved

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_16

systems (human and artificial systems) avoiding the use of a reductionist approach, which is no longer suitable for these kinds of complex co-adaptable systems.

Beyond the focus on VR/AR, these applications address the question of the relation-ship between the human and the artificial system such as the principles developed along the Bio-Cyber-Physical System (Bio-CPS) especially when user centered sensors are introduced. This brings some important issues related to the interaction/integration between the system and the human. Bio-CPS are considered as an integration of computation elements within biological systems. Beyond the classical HITL concept, which consider interaction as only information exchanges, designing and developing human/machine systems cannot be reduced to this simple scheme. Using interactive and artificial technologies requires integrating artificial elements and structural design, usually by artificial or artifactual functional interactions and their dynamics. Designing Bio-CPS is a real intellectual challenge at the intersection between the biology and the cyber science. The development of new interactive systems involves the necessity to develop theoretical basis for designing Bio-CPS. In addition to the classical issues of CPS, Bio-CPS has added some other linked the differences in the nature of the interactions between system components. Another important issue is linked to the numerous hierarchical levels involved in biological systems from the cell level to the body level, the physical or the cyber systems dealing at most with two levels. Moreover, the complexities of these two systems make interoperability hard to design. Finally, challenging human-machine system - biocompatible and bio-integrative medical devices, design and organization involves modeling a heterogeneous system of systems, which are different by nature. This requires a proven and valid epistemic framework fitted to hybrid systems and domain specific engineering and challenges the question of the nature of human/machine integration, correctness-by-construction and ensuring human systems integration reliability.

In the literature, there are few papers dealing with this issue. However, some research work start to deal with Cyber-Physical Systems with human-in-the-loop, the human interaction being able to occur at cyber level, physical level or both. Thus, it is necessary to develop a new engineering perspective in the design of VR/AR application through the Bio-CPS design point of view. The goal of this paper is to present Bio-CPS related concepts and to sketch some guidelines for future virtual environments (VEs) design focusing on specific application fields. In addition to these Bio-CPS considerations, the paper will also take into consideration ethical and philosophical aspects (Philosophy of technics) of human use of such systems.

The paper is structured as follows. Section 2 explains the related concepts through detailed definitions. Section 3 explains what are Bio-CPS. Section 4 is giving some target applications and, finally, Sect. 5 concludes the paper by drawing some perspectives.

2 Concepts and Definitions

Safe design and correctness by construction of human-machine systems, as virtual environment, is a major challenge for both consumer and professional applications, especially in the areas of critical systems such as health, aerospace and defense.

2.1 Virtual Environments Design

As a scientific challenge, Virtual Environment (VE) design is about the conception of a theoretical framework dedicated to human-machine integration and its modeling, from epistemology to formal and experimental methods. As a practical challenge, VE design is about correct design for enhancement and reliable engineering of human systems integration, from human-machine interaction to sociotechnical system levels, their behavior and their performances.

2.2 Virtual Environment and Epistemological Consideration

Current methods and tools of virtual environments design and human-machine interaction engineering are derived from cybernetics, automation, computer science, cognitive sciences and human factors. According that classical paradigm, virtual environments (VR and AR) are modeled as “human-in-the-loop” systems, in which a human is interacting with a less or more immersive cyber-physical system. It models the system using a technical approach, describing interaction links and interfaces as a physical or computational process. In this paradigm, interactions are considered as a process of communication between components of the system reduced to each other in symmetric loops “input, information processing, output”.

This neo-behaviorist, metaphorical and reductionist conception is not sufficient to model integrated systems in their whole and their relativity of space and time scale. Understanding and describing the organization of virtual environment complex systems, whether organic or organized, requires a renewed conceptual framework of knowledge, representation and modeling principles.

2.3 Cyber-Physical Systems

A Cyber-Physical system is defined as a deep intricate system made of both computational elements (cyber part) and physical elements. It is generally considered as an extension of classical embedded systems which are based on feedback loop paradigm. The main difference between feedback based embedded systems and CPS is the number of interactions between components which cannot be reduced, in CPS, to a simple bidirectional data exchange (cf. Fig. 1).

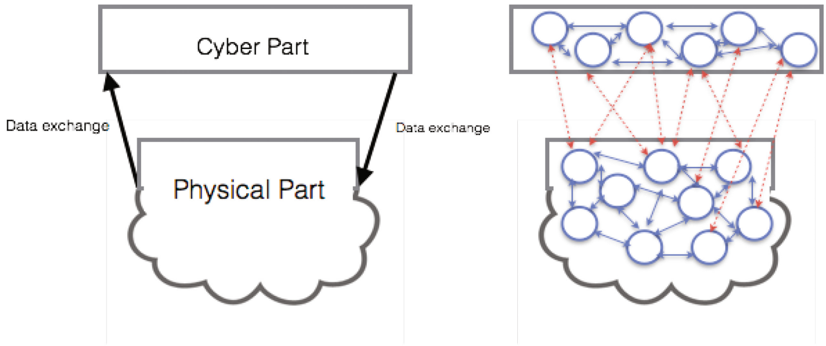


Fig. 1. Embedded systems based on feedback loop (left) vs. CPS (right)

Indeed, the number and the different nature in interactions make a CPS a complex system in which emergent phenomena can occur [1]. This emerging property make the system non-deterministic and then somehow unpredictable. Generally, the emergence is tied to the existence of an external observer aimed at noticing and analyzing phenomena that appear at a higher level of abstraction. This concept, which has been first introduced in biology and in philosophy as opposed to reductionism, is now wide-spread in literature especially in computer science models such as in artificial neural networks [2], in multi-agent systems [3] or in sociology.

We may consider that the increase in complexity in the system will lead to a de-crease the domain of stability. If it is generally the case in engineering, it has been proven that, in biological system, the increase of complexity can lead to an increase of stability domain [4].

Two conceptual methods are generally used for designing CPS. The first method is focusing on the physical part and is based on the introduction of computer elements into physical processes. This is linked to hybrid system theory and is based on a common reference of time spread all over the systems components whether they are computer or physical parts. The second method focuses on how to integrate component from Physics into computer models or algorithms. The main difficulty is to re-think the abstract model used in computer science so as to integrate the physical constraints into the model conception.

The system evaluation is also different from classical embedded systems to CPS. Indeed, where the performance evaluation of the firsts is exclusively associated to time response efficiency, the CPS evaluation is more related to the adaptability of the system to the physical part dynamic evolution.

2.4 Interactions: Different by Nature

An interaction is generally defined as a dynamical link between several elements using a set of actions that can be reciprocal or not. Obviously, several forms of interaction can be defined from the high-level language exchange to Physics interactions based on force concept such as gravitation or electromagnetic forces. Despite the numerous types of forces, one can sort them out depending on their nature. Thus, interaction can be local or non-local, direct or indirect.

An interaction is considered as local when the realization of the interaction is made within a short distance between elements. By contrast, they are defined as non-local when they occur between entities separated in space with neither noticeable intermediate medium nor process. As examples, spring-damper physical functions are local whereas exchanges between organs in animal bodies are typically non-local. If, in these examples it seems obvious to determine whether an interaction is local or not, it is not so simple. Even if one can generally say that biological interactions are mainly non-local and that physical interaction are local, some Physics interaction have been proved to be non-local such as the Aharonov-Bohm effect [5] or the quantum entanglement [6].

By the same, interactions can be direct if the related entities actions have a direct influence on their mates (a collision between two particles). They can also be indirect when the action is passing through a medium for transmitting the associated influence (data transmission using a network, data exchanges using pheromones in ant colonies…) Generally, the medium is a, sometimes non-active, part of the system shared by all its components.

Another important characteristic of interactions is their symmetrical properties. If most of the physical interactions are symmetrical such as the electromagnetic or the gravitation forces for instance, the symmetry in interaction is not as common in Bio-logical systems [7].

2.5 Integration and Virtual Environment Design

Virtual environments are classically described as a HITL artificial technical system in which the interacting participant is immersed. The relation between the human and the VEs technical devices is represented as a cybernetic closed control loop produced by the symmetrical process of interaction. The quality of that interactive process and the feeling of being there, inside the artificial environment, is described by the metaphysical concept of presence. According that classical paradigm, VEs designers try to attempt that ideal quality developing human centered design methods.

But HITL metaphor is based on the belief of a symmetricity of the interactive process and a reductionist equivalence of nature - physical and computational. Moreover, in that way physiology is reduced as the way to measure and obtain computable data, that could be use to regulate the artificial system states. That classical approach is grounded on a fundamental epistemological error.

Thus there are two factors that limit Human System Integration: The first factor is a misunderstanding of the autonomy of a living system. The second is the underestimation of the nature of interaction and integration as a biological process involved in the dynamics of knowledge as a physiological function of the living body.

To avoid that trap we propose to slip from interaction design to integrative design and introducing the concept of Bio-CPS.

3 Towards Bio-CPS Concept and Virtual Environment Integrative Design

3.1 What Are Bio-CPS?

Bio-CPS can be defined as an integration of computational and physical elements within biological systems. The main goal of designing Bio-CPS is to make Biological system working with Computational and Physical parts to perform jointly one specific task that no part of the system can accomplish alone. The main issue of Bio-CPS is tied to the difference in nature of the implied systems. The human nature is biological when the machine nature is artificial, computational and physical.

Thus, making all these elements work together asks the question of the nature of the relationships between elements. If we consider that the interaction can be reduced to a simple coupling such as in the Human in the loop concept, we place automatically the design process into a reductionist point of view which is not is not adapted to the complex nature of this system of systems. Besides, this state of mind may make us miss some interesting properties of the Human/Machine coupling. By contrast, if we want to adopt an integrative point of view, we are considering as primary element the complex nature of the whole system.

3.2 The Challenges of Integrative Design

One of the main challenges in Bio-CPS is to think this system of systems as a whole, which must be considered at a higher level of organization as the Human body. In this integrative point of view, the CPS part of the system may be designed considering the biological nature and associated constraints. This will lead eventually to the design of CPS as anatomical and physiological extensions of the biological system. To answer to this issue and to shift from the HITL paradigm to the Bio-CPS concept, we proposed an isomorphic framework for modeling this kind of systems. This conceptual framework proposed in [8] (cf. Fig. 2) is composed of a three-dimensional referential frame, the axis of which represent the structural elements, the shapes and the dynamics dimensions and requirements. Each couple of dimension describes a specification plan. Thus, the plan described by the structural elements and shapes corresponds to the architecture, the one described by shapes and dynamics specifies the behavior of the system and finally, the plan described by structural elements and dynamics corresponds to system evolution through three main types of functional interactions on physical (Φ), logical (Λ) and biological (Ψ) levels.

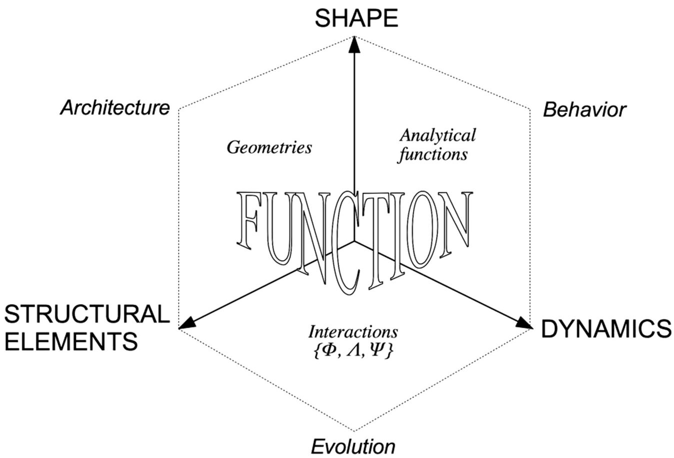


Fig. 2. The Bio-CPS conceptual framework [8, 9]

4 Target Applications

4.1 Medical Applications

As previously said, these technologies are triggering a new step in the use of simulation technology for clinical purposes. After an early incursion in the field twenty years ago which suffered from a lack of maturity from the devices, it is now mature enough to consider more usable and affordable high fidelity VR scenarios for the conduct of human research and clinical applications. Thus, evolving behavioral health applications can now usefully leverage the interactive and immersive assets that VR and AR afford as the technology continues to get. Among the already existing applications, Virtual Human (VH) concept and especially Virtual Standardized Patient (VSP) such as Justin, Justina or Sgt. Castillo [10, 11], is well adapt to the specificity of Psychology. The extension of VR and AR to other medical fields such as gesture training, assisted surgery [12] or functional reeducation [13] needs also some improvements from human-machine bio-engineering and artificial intelligence. Indeed, it seems now obvious to introduce more (self-)adaptability to the artificial behavioral models to reflect the intrinsic complexity of the reality.

4.2 New Kinds of Applications for Education

Beyond the medical applications presented in previous section, educational applications using VR and AR have also a marvelous potential for future education practices. The main underlying goal of using AR/VR in education is to provide a better understanding to the student. The key idea is then to cross the next step in distant education currently massively made by MOOC whether it is teacher centered (xMOOC) or student centered with collaborative features (cMOOC). This new step that can be called "VRMooc" will then benefits from both immersive and interactive features. Using suitable artificial intelligence algorithms, it could also bring adaptation to the learning progression and to the needs of each student. Moreover, as opposed to current MOOC, which is devoted to theoretical learning, the VR/AR version can also explore gesture learning taking inspiration from application used for physical therapies. This will bring a wide range of new educational purposes even for manual learning provided it is associated to haptic devices and/or sensors for tracking user/student behaviors. As opposed to other massive distant learning, which are relying on model based learning (general sciences, engineering…), this can also be used for evidence based learning (medicine, surgery…). Another challenge related to the integration of VR/AR into education is the possibility to obtain a full bi-directional interactivity using, for instance, similar principles as those we can find out in the Microsoft "holoportation" application associated to HoloLens device [14].

5 Conclusion and Perspectives

Virtual environments safe design, correctness by construction and reliable development are main issues for both personal and professional applications. With current interactive or cognitive systems, boundaries between the human and the artifacts fade. This disappearance of the boundaries between the biological and social human and the interactive and informational (cyber-physical-system) questions current methods of VES design, risk management and the legal framework defining good practices.

Nowadays “bio-psycho-social” human is reduced to a desembody computo-logico-symbolic cognitive being. The behavioral and cognitive heuristic metaphors, which remain productive for the design of automatic systems, generate new questions and problems and lead us to question the theoretical and experimental conceptual foundations of VES design.

For us, it is a question of designing an integrative framework ranging from man-machine systems to the human-to-human systems with its artefacts as a question of integrative coupling in multi-scale dimensions of two systems of different natures which can not be reduced to one another: the biological and anthropological human and the CPS. Understanding this synthetic hybridization requires a new conceptual apparatus and a new knowledge system of human systems integration, capable of thinking the human machine and its applications, modeling them and testing them as a structural and functional integrated whole.

In this conceptual and practical framework, inspired by theoretical integrative biology and CPS theoretical analysis, we define the integrative design of virtual environments and their future education and medical applications

References

1. Moncion, T., Amar, P., Hutzler, G.: Automatic characterization of emergent phenomena in complex systems. J. Biol. Phys. Chem. **10**, 16–23 (2010)
2. Hanson, S.J., Negishi, M.: On the emergence of rules in neural networks. Neural Comput. **14** (9), 2245–2268 (2002). doi:10.1162/0899766023
3. Noël, V., Zambonelli, F.: Engineering emergence in multi-agent systems: following the problem organisation. In: 2014 International Conference on High Performance Computing and Simulation (HPCS), Bologna, pp. 444–451 (2014). doi:10.1109/HPCSim.2014.6903719
4. Chauvet, G.A.: Hierarchical functional organization of a formal biological system: a dynamical approach. I. The increase of complexity by self-association increases the domain of stability of a biological system. Philos. Trans. R. Soc. Lond. B. **339**, 425–444 (1993)
5. Figielski, T., Wosinski, T.: On the quantum interference transistor based on the electrostatic Aharonov-Bohm effect. In: Second International Conference on Advanced Semiconductor Devices and Microsystems, pp. 27, 30, 5–7 October 1998. doi:10.1109/AS-DAM.1998.730158
6. Rechtsman, M.C., Lumer, Y., Plotnik, Y., Perez-Leija, A., Szameit, A., Segev, M.: Topo-logical protection of photonic quantum entanglement. In: 2016 Progress in Electromag-netic Research Symposium (PIERS), Shanghai, pp. 974–974 (2016). doi:10.1109/PIERS.2016.7734548

7. Chauvet, G.A.: Hierarchical functional organization of formal biological systems: a dynamical approach II. The concept of non-symmetry leads to a criterion of evolution deduced from an optimum principle of the (O-FBS) sub-system. Philos. Trans. R. Soc. Lond. B Biol. Sci. **339**(1290), 445–461 (1993). doi:10.1098/rstb.1993.0041
8. Fass, D., Gechter, F.: Towards a theory for bio-cyber physical systems modelling. In: International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. LNCS, vol. 9184–9185, pp. 245–255. Springer International Publishing, August 2015
9. Fass, D.: Augmented human engineering: a theoretical and experimental approach to human systems integration. In: Cogan, B. (ed.) System Engineering Practice and Theory, Intech - Open Access Publisher, Rijeka, Croatia, pp. 257–276 (2012). ISBN 979-953-307-410-7
10. Virtual Patient with Voice Recognition. https://www.youtube.com/watch?v=CQTEcJJ_RhY&index=44&list=UUQrbzaW3x9wWoZPl4-l4GSA
11. What is SimCoach.wmv. https://www.youtube.com/watch?v=PGYUqTvE6Jo&list=UUQrbzaW3x9wWoZPl4-l4GSA&index=47
12. Delorme, S., Laroche, D., DiRaddo, R., Del Maestro, F.: NeuroTouch. Oper. Neurosurg. **71**, ons32–ons42 (2012). doi:10.1227/neu.0b013e318249c744
13. A Brief Introduction to How the Primesense Camera will Revolutionize Virtual Rehabilitation. https://www.youtube.com/watch?v=geyIvG4uKxY
14. Holoportation – Microsoft Research. https://www.microsoft.com/en-us/research/project/holoportation-3/

A Thermal Manikin Crus with Thermal Regulation Characteristics Research

Lei Yun[1,2], Tian Yisheng[2], Tanqiu Li[3], Jiang Ke[1], Wen Lei[4], and Ding Li[2(✉)]

[1] Beijing Information Science and Technology University, Beijing, China
[2] School of Biological Science and Medical Engineering, Beihang University, Beijing, China
ding1971316@buaa.edu.cn
[3] National Key Laboratory of Human Factors Engineering, Astronaut Center of China, Beijing, China
[4] Chongqing Qingping Machinery CO., LTD., Chongqing, China

Abstract. The Manikin is a kind of standard measuring equipment that has gradually become the clothing ergonomics and environmental thermal comfort evaluation field indispensable tool. The Manikin is a heat exchange device of stable environment design, so the control model mainly by constant temperature and constant heat flux and the CE model can not reflect the characteristics of human thermal regulation. Therefore, it is difficult to apply to individual protective equipment (PPE) thermal function test. In order to realize the continuous time manikin to the changes of the environment stress adjustment process. This article carries on the transformation to a dry state manikin crus model, the human bioheat equation simplified improved prediction model, and combined with thermal manikin. In order to validate the model, we selected four subjects in a continuous heating environment (27–33 °C) in each of the two groups of tests, and compared with the predicted results. And the results showed that the average temperature change curve of human skin was consistent with the calculated results.

Keywords: Thermal manikin · Electric heating · Thermal control · Equivalent temperature

1 Introduction

The manikin can simulate the steady state heat production and human physiological changes. In this way, the manikin can replace the human body to carry on the thermal comfort evaluation. It is widely used in the field of transportation, spaceflight, clothing and so on. But for the personal protective equipment (PPE), such as Chemical protective clothing, pressured spacesuit, etc. It is difficult to maintain the heat balance of human body for a long time. There is a small space inside clothing, so heat capacity is small. The micro environment will be fluctuating even a little heat is input. Therefore, the common manikin productions are very hard to be used in these cases. The mainly reason is these manikin control models can't simulate the human thermal regulation function.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_17

Thermal manikin control mode can be divided into three categories: Constant skin surface temperature (CST) mode. This mode according to setting a skin temperature as measure standard and adjusting the heat under different conditions to maintain the temperature stability. But this kind of mode can't be used in some extreme environments. Because of limited heating, it's hard for a manikin to maintain skin temperature. Constant heat flux (CHF) mode. The heat flux of this model is constant. It can be used to simulate the metabolism of the human body. But as we know it, the manikin is made by copper. There are larger differences in heat dissipation between copper and human skin. It'll cause huge errors in data. The comfort equation (CE) mode [1]. The mode set the skin surface temperature under the neutral conditions (Maximum temperature not exceeding 36 °C, Minimum temperature not exceeding 10 °C). But it can't be used in some extreme circumstances.

The three control modes mentioned above had been used in the evaluation of clothing [2], calculations of local heat transfer coefficients at different body segments [3]. But all the above models can't be used for functional testing of some special clothing, such as liquid cooled suit and fire-protection suit, etc. And all the above models can only be used in collecting data at a certain time.

In addition, the heat transfer factors of thermal manikin is an indispensable besides the thermal control model. However, According to the division of human body temperature field. The early research is limited to calculation conditions. The initial mathematical model is very simple. Until Burton [4] is applied to the human body heat conduction equation for the first time. The theoretical solution of the model shows that the temperature distribution in the human body is a parabola. Then Machle and Hatch [5] improve the energy balance equation. The temperature of the human body in the human body temperature model can be represented by two typical temperatures: the core and the shell temperature. The establishment of this model is the beginning of the study of human body temperature distribution. The human body heat regulation has the very large impact to the human body temperature distribution. But none-of them consider the human body heat regulation function at that time. Until Crosbie et al. established a closed loop control system model of human body temperature regulation considering the physiological regulation function of human body. This model laid the foundation of the mathematical simulation of the human body heat regulation system. The model can be used to predict the steady skin temperature and core temperature, and can also be used to predict the dynamic changes of skin temperature. Then Stolwijk [7] further improved the mathematical model of neural regulation function. He believes that the human body sweating, evaporative cooling, muscle contraction, vasodilatation chills were affected by the hypothalamus and the mean skin temperature control. He described the mathematical description of the "fixed point" principle in physiology. Although in the Stolwijk model, the control equations of thermal regulation activities lack of adequate physiological evidence from the engineering point of view. However, these control equations are simple and reflect the control mechanism of human body thermal regulation.

After perfecting the mathematical model of human body heat regulation. Foda combined with the Pierce model [8], made a dry thermal manikin control model in 2011 [1]. He made a new prediction model by adjusting the setting point and calculating parameters of the skin under neutral conditions. The average absolute temperature difference of thermal manikin for a range of 0.3–0.8 °C. This kind of prediction model in thermal manikin comes up for the first time in history. But even this manikin also has a problem. The model does not introduce the concept of time. We can only get the temperature at a certain time, but we can't predict the change of temperature within a period of time.

Therefore, this experiment will study the thermal regulation of the crus segments. Establishing a body segment with adaptive thermal regulation characteristics. And use it to study on the simulation of human body thermal regulation. In addition to predicting the temperature change of a certain time, the model can predict the change of human body temperature in a short time.

In order to confirm the experimental results, we conducted several experiments to 4 people in the same room. Finally, the experimental data and the predicted results are compared.

2 Methods

The manikin crus segment tightly wound heating sheet, the aluminum foil is laid on the heating sheet to make the heating quantity even. Then place the temperature sensor in the PT100 manikin crus segment skin surface. If the experiment requires multiple PT100 temperature sensors, then we take the average of the measured temperature. Comparison of manikin crus segment surface temperature measurement value and the value of the bio heat equation. If the absolute value of the error is less than 0.5 °C. It proves that the result of temperature control is achieved.

Bio heat equation:

$$\rho c \frac{\partial t}{\partial \tau} = \lambda \left(\frac{\partial^2 t}{\partial x^2} + \frac{\partial^2 t}{\partial y^2} \right) + \mathrm{q}_m + q_{res} + m_b c_b (\mathrm{t}_{ar,i} - \mathrm{t}) \tag{1}$$

- ρ is human tissue density, kg/m^3. Because the manikin segments crus using internal sponge, we consider only the changes of skin temperature. ρ skin = 1085 kg/m^3.
- c is human tissue specific heat. Consider only the skin surface. C skin = 3680 J/(kg°C).
- t is human tissue temperature
- λ is thermal conductivity of human tissue
- τ is time
- x, y is space coordinate
- q_m is metabolic heat production per unit volume
- q_{res} respiratory heat exchange per unit volume
- m_b is unit volume blood flow
- c_b is the specific heat of the blood
- $t_{ar,i}$ is segmentally arterial blood temperature (37 °C)

The temperature is uniform in the circumferential direction. So the results calculated by y = 0 are the same as those of other coordinates. So y = 0, this model will be reduced to a one-dimensional coordinate system. In this formula, X is considered to be a manikin crus segment skin thickness. In addition, the exchange of heat exchange in the unit volume is only in the head and chest segments, there is no respiratory heat exchange in the crus segment, which can be considered as 0.

Simplified equation:

$$4400\Delta t^2 - 85179.73\Delta t + c = 0 \quad (2)$$

The constant C in the formula is two unknown quantities MB and QM, so only these two unknown quantities are needed. According to the quadratic equation with one unknown, can obtain the variation of temperature Δt.

Connect the temperature control system of thermal manikin crus segments hardware part of the temperature control box and computer. First, the user enters the skin surface of the computer to measure the initial value of the temperature and the ambient temperature and wind speed are used as the boundary conditions Δt. t + Δt as the next input value through the cycle to continue operations, and according to the set time interval in output. Then, the variation of the calculated value is simulated by the manual control temperature control box, so that the hardware is heated and stable. We can find that the human's crus temperature change is the same as temperature change of the manikin crus segments at the same time. So we reached the design significance of the manikin crus segment.

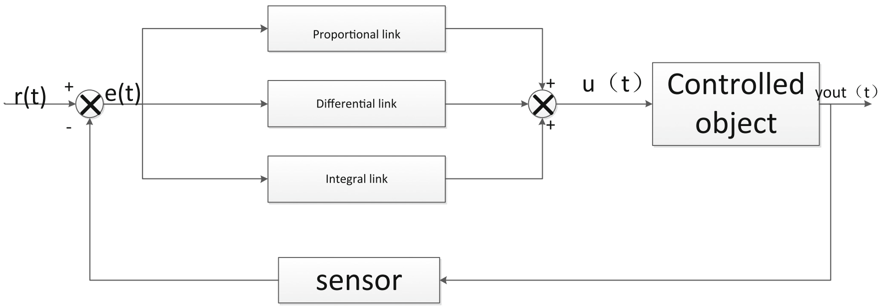


Fig. 1. The basic control principle of thermal manikin

The closed loop control system is a kind of feedback system which gives a signal to the system by the input quantity (Fig. 1). Closed loop control system not only positive feedback but also negative feedback. The human body temperature system is a closed loop control system with negative feedback. Human skin, eyes, nose and so on are sensors. They will signal to the body temperature of the hypothalamus, the hypothalamus to the human body to complete the work instructions. The PID control system used in the design is a closed loop. When the setting temperature SV is higher than the measured temperature PV, the temperature control box run and is supplied to the silicon rubber heating sheet for heating. When the temperature reaches the set temperature, the system automatically cut off the power supply to stop heating.

3 Thermal Manikin Experiments

The use of temperature control laboratory control boundary conditions that the ambient temperature and wind speed. Then, Invite different subjects to conduct test. Take four loci of subjects' crus digitorum longus as the temperature measuring point. The four PT100 temperature sensor attached to the crus site. The senso is connected to the SWP-ASR-M multi-channel inspection instrument and connected with the PC through the RS232-RS485 converter. In the PC side there is a good sub VI can be read to the human body every 60 s of temperature data and collection. Then the initial value of the data for the initial temperature of the human body temperature T input to another connected VI. It then automatically calculates a theoretical value and loops. The human experiment compared with calculated temperatures and presented in tabular form. The test proves that the dry manikin can simulate the human adaptive adjustment (Fig. 2).

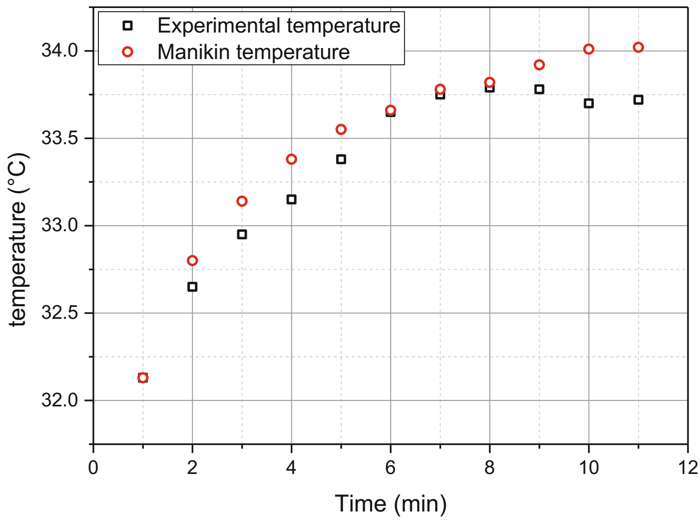


Fig. 2. First test

4 Result

Measure the first human body temperature. And take the initial human body temperature as the initial temperature. The human body temperature is recorded every 60 s to get the experimental temperature. The experimental temperature is input into the two equation of one dollar and calculate Δt. Add Δt and the experimental temperature can get the second temperature. After 10 calculation, we can get 11 sets of experimental data (Fig. 2). The difference of Δt is obtained by subtracting the calculated temperature from the experimental temperature. Through the experiment, it can be seen that the calculated temperature line from the calculation is in an upward trend from 31 to 34 °C, and it tends to be stable after reaching the temperature of 34 °C. The calculated results of the model agree well with the experimental temperature, and the absolute value of the error is less than 0.5 °C. It is proved that the two factor model of the human biological heat equation can simulate the adaptive temperature and thermal regulation of the human body within a certain range (Fig. 3).

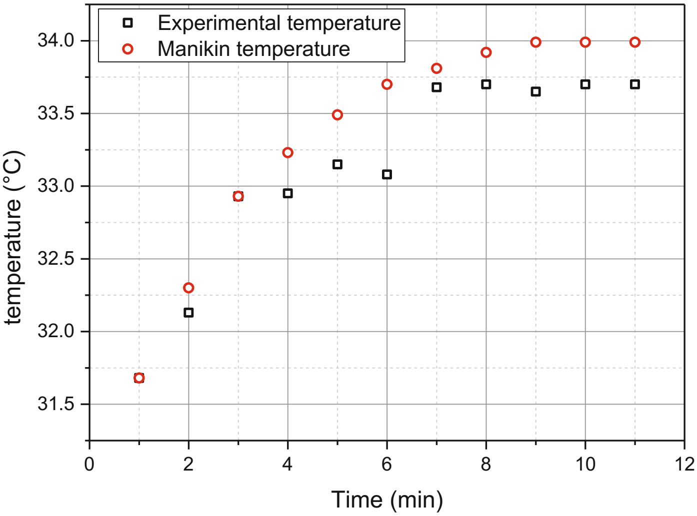


Fig. 3. Second test

5 Discussion

The shape of the human crus is irregular. Therefore, the processing of the silicon rubber electric heating plate and can't completely cover the surface of the crus. It's easy to produce voids. And leads to uneven temperature distribution (Fig. 4). The heating processing is not perfect in the silica gel and the manikin crus segments close to the degree of deficiency of many voids. Directly affect the control of manikin temperature, cause electrical loss.

It is very difficult to simulate the change of the electric heating in the sudden encounter with the cold environment (Fig. 5). All the cooling principle depends on its own heat to simulate. There are design defects in this area.

PID temperature control is used when the pulse temperature control, which may occur in the acquisition of data when unnecessary errors. However, the error the error can be calculated by calculating the heating rate and cooling rate, and the real data can be taken.

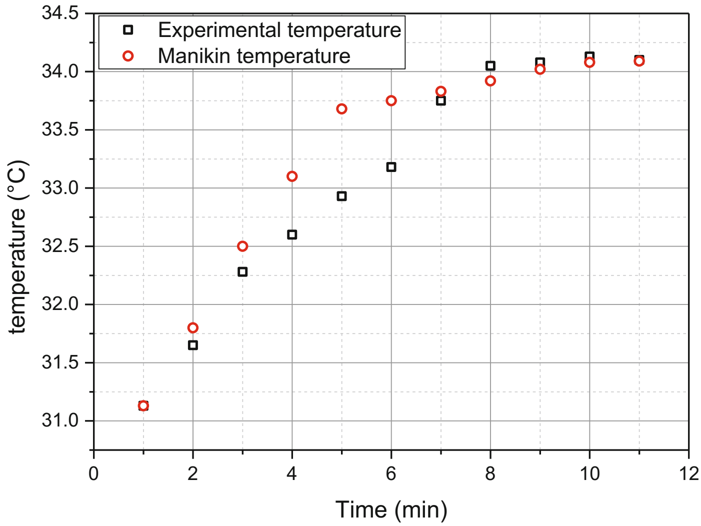


Fig. 4. Third test

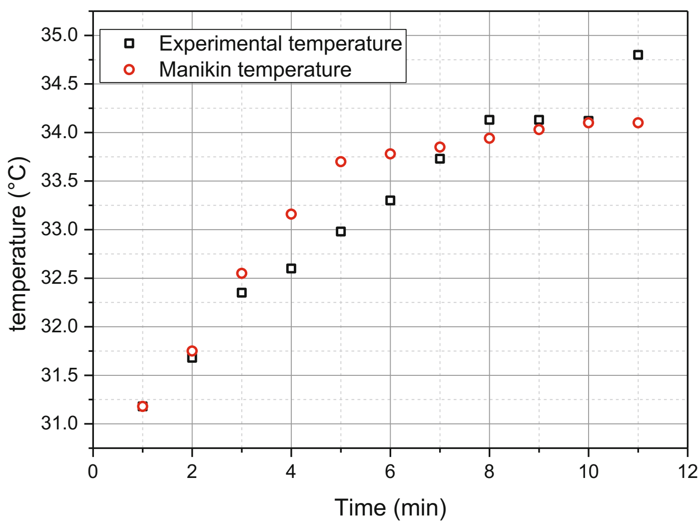


Fig. 5. Fourth test

6 Summary and Conclusions

1. We made a thermal control model with adaptive function of thermal manikin.
2. The data is similar. It means the physiological changes of manikin can be simulated in a certain extent. The calculated curves are similar to those obtained by the human body under the same conditions.
3. Each part of the manikin is independent. There is no interaction problem between them. Also this research does not consider the respiratory heat transfer situation. But the manikin crus segment system theory can be extended to various segments of the human body. Simulation of changes of various temperature so as to form a complete thermal manikin system, this is also a great prospect for future.

Acknowledgments. This research was supported by National Key Technology R&D Program (2014BAK01B05), National Natural Science Foundation of China (51175021) and Defense Industrial Technology Development Program (JCKY2016601B009).

References

1. Foda, E., Sirén, K.: A thermal manikin with human thermoregulatory control: implementation and validation. Int. J. Biometeorol. **56**(5), 959–971 (2011)
2. McCullough, E.: The use of thermal manikins to evaluate clothing and environmental factors. Elsevier Ergon. B. Ser. **3**, 403–407 (2005)
3. Quintela, D.A., Gaspar, A.R., Borges, C.: Analysis of sensible heat exchanges from a thermal manikin. Eur. J. Appl. Physiol. **92**, 663–668 (2004)
4. Burton, A.C.: Human calorimetry: the average temperature of the tissue of the body. J. Nutr. **9** (8), 261–280 (1935)
5. Machle, W., Hatch, T.F.: Heat: man's exchanges and physiological responses. Physiol. Rev. **27**, 200–227 (1947)
6. Crosbie, R.J.: Electrical analog simulation of temperature regulation in man. Trans. Biol. Med. Electron. **8**, 245–252 (1961)
7. Stolwijk, J.A., Hardy, J.D.: Temperature regulation in man-theoretical study. Pflugers Archiv. Eur. J. Physiol. **291** (1966)
8. Gagge, A.P., Stolwijk, J.A.J., Nishi, Y.: An effective temperature scale based on a simple model of human physiological regulatory response. ASHRAE Trans. **77**(1), 247–262 (1971)

Mid-Air Hangeul Input System

Kyungjin Han(✉) and Jongweon Lee

Department of Digital Contents, Sejong University, 209, Neungdong-ro, Gwangjin-gu, Seoul, South Korea
kjinn.han@gmail.com, jwlee@sejong.ac.kr

Abstract. Some of the traditional keyboards and their layouts are inadequate when it comes to the usability aspect of various responding devices. Even though users anticipate various interaction experiences, the ordinary mutual interaction method has lagged far behind the expectation of users. Thus, this study proposes the Mid-air Hangeul input system available in a large screen environment which direct contact by applying gesture recognition is difficult. User experiment has been conducted based on ten participants. Even though the typing speed was slower than other studies, this system is easy to use and learn. This study proposes the usability in a limited situation which users are facing from the text input viewpoint and the possibility of flexible combination with other devices.

Keywords: Text input · Natural User Interface · Virtual reality · Gesture recognition · Usability

1 Introduction

People living in the modern society use countless number of devices every day. Together with the development of devices, the studies on User Interface are also continuously taking place. User Interface (UI) refers to the physical and postulated medium which has been created for the purpose of temporary or permanent approaches enabling communication between users & objects or between systems, especially machines and computer programs. User Interface - the term comprehensively referring to the devices and technology connecting computer system and users - is the system enabling smooth mutual interaction between users and systems and includes physical hardware element and logical software element. User-friendly Interface, the technology close to the conversation style of humans, unlike the existing Interface, is the technology through which intuitive and natural interaction with systems is possible by using a body part without a mouse, a keyboard, or a pen. Users can control the device and systems by using the gestures and voices based on the familiar motions used in daily life without the burden of having to learn how to use the device. Natural User Interface (NUI) is expanding the consumer class of the existing IT by shortening the gap between users and IT devices. For a smooth mutual interaction between users and systems, text input is essential. Text input is one of the primitive functions of interaction identified by Foley et al. [1]. Even if the interactive system for general purposes is not used for lengthy uninterrupted text input, text input must be supported by the

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_18

system. Touch style text input and voice recognition using capacitive touch technology are widely used in mobile devices including wearable devices and a number of related studies have been carried out. However, there has been few studies on the noncontact style text input method when there is a physical distance between user and device. The existing studies are mostly about the system which is able to type texts by using handheld button-equipped pointing devices. Thus, this study suggests the intuitive Mid-air Hangeul input system which reflected the inherent characteristics of Korean alphabets based on the user-friendly input method based on the gesture recognition without handheld devices.

2 Related Work

Input techniques are mainly divided into touch input techniques and non-touch input techniques. The touch-based input techniques - the leading touch input techniques - are the most widely used in the mobile computing environment. Touch techniques - the proximity sensing touch-based recognition not free from physical distance - are the methods which respond to the multiple touches on screen and capacitive method is the universal method on smartphone tablets. The popularization of wearable devices and the fat-finger problem in the interacting process on small touch screens have been pointed out as new factors and the studies to resolve these problems are being conducted. Relevant researchers have developed the QWERTY layout keyboard style virtual text input system targeting smart watch devices [2–6] and the input system which introduced a new style layout keyboard has been suggested recently [7, 8].

Non-touch input techniques are mainly classified into the motion or gesture recognition which recognizes motions by instructions, the voice recognition which recognizes voice or speech by instructions, and the vision trace-based recognition. Voice recognition - the technique understanding and responding to voice or speech by automatically converting into text and analyzing them - is the fastest among all input techniques but has the disadvantage of markedly low level of recognition accuracy since it is highly influenced by the macro-environment and users' vocalization and pronunciation. Moreover, owing to the social issues such as the exposure of personal information and privacy, a number of problems are involved when it comes to being widely used from the perspective of social receptiveness. As of now, the users of the mobile computing environment partially use input through voice recognition in their daily lives [9]. Also, the technique of typing texts by tracing users' vision is the optimal method through which users can interact independently. However, there are limits for it to be widely distributed such as expensive equipment for vision trace and eye fatigue.

Gesture recognition - the technology which the system directly understands and responds to users' intentions by recognizing and analyzing humans' motions through sensors - is suitable for long distance interaction since it has less limitation when it comes to physical distance than touch recognition. The number of studies on text input system based on gesture recognition is small but Anna et al. proposed eyes-free multi-finger gesture text entry [10]. The study reports performance and individuation characteristics of fingers and deploys them to the design of a mid-air text entry method using multi-objective optimization. Similarly, Sridhar and Srinath et al. proposed the dexterity of using multiple fingers for mid-air input [11]. The mid-air input allows text entry on any device where direct touch is not practicable such as smartwatches or TV screens. The aforementioned two studies [10, 11] are the gesture input techniques that trace users' hands.

When it comes to the text input system targeting big screens, Garth et al. proposed the system through which users can input text by the Wii remote controller. The study designed three layouts - Circle, QWERTY, and Cube - and the QWERTY layout, the most familiar to users at short distances, was the fastest. However, as a result of conducting experiments by setting different long distances, short distances, vision-dependent condition, and non vision - dependent condition, even the QWERTY keyboard, the familiar layout in long distances, had low performance status [12]. Aoki et al. proposed the non vision - dependent text input system suitable for large screens [13]. The study used the up, down, left, and right rudders of the Wii remote controller and the final text input is possible through the movements of wrists. The previously mentioned two studies were able to input texts through an additional handheld device whereas Gang et al. proposed text entry using freehand gestures captured with a low-cost sensor system [14]. Freehand Gestural Text Entry designed two keyboard virtual layouts - QWERTY and Dual-circle - and three text selection techniques - timeout, reach, and timeout & reach. Markussen et al. proposed Vulture - a word-gesture keyboard for mid-air operation [15]. Vulture adapts touch based word-gesture algorithms to work in midair, projects users' movement onto the display, and uses pinch as a word delimiter. However, the above-mentioned studies have the burdens of having to use additional devices or having to learn the gestures of the letters which users want to input. Thus, in order to minimize preliminary learning, this study proposes the non-touch input system reflecting the characteristics of Korean alphabets on large screen by choosing the swipe gestures easy to be used by users.

3 Proposed System

This study is a Mid-air Hangeul input system which is available in a large screen environment. Figure 1 is the system structural design proposed by this study.

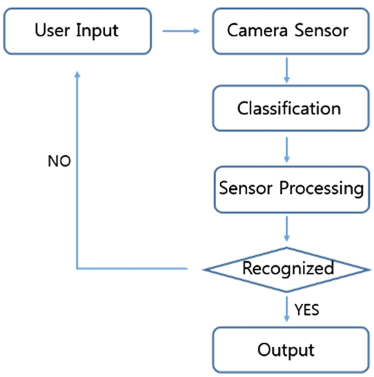


Fig. 1. The developing process

We obtained motion sensor data by using user's joint information through the low cost sensor of Microsoft Kinect. Recognize and analyze the motions through classifications based on the coordinate data of user's fingertips. Draw results according to gestures. For an effective alphabet input system, the experiment to decrease interruptions felt from the standpoint of the user was conducted. In order to minimize the surface occupied by alphabet input layout on the screen, two or three alphabets were included in one button. In addition, the buttons divided into consonants and vowels were placed at both sides of the user so as to reflect the characteristics of Korean alphabets. This has been done so as to induce natural movements using both hands when typing Korean alphabets.

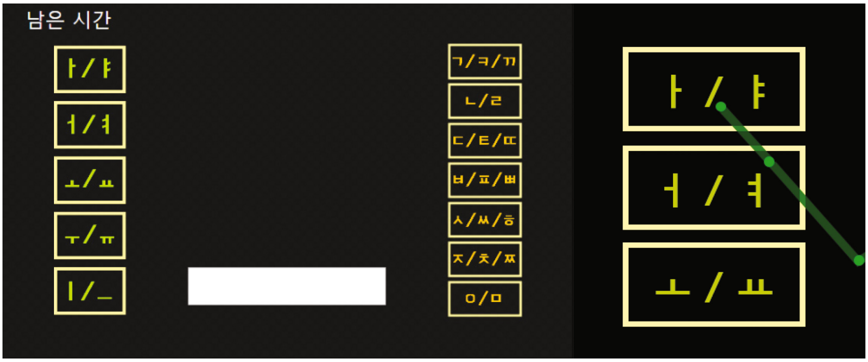


Fig. 2. Preliminary system

As in Fig. 2, the initial version of this system was a layout in a straight line and alphabets could be typed only when the fingertip of user stayed on a certain button for several seconds. Experiments were conducted with the initial version but the phenomenon of users choosing two buttons at the same time took place. Moreover, it was difficult to choose the buttons placed on the lower part when the body of users were fixed and they could only move their arms. After taking into account the problems of deviating from the purpose of the input system considering the convenience of users, the form was adjusted to the waveform as in Fig. 3.

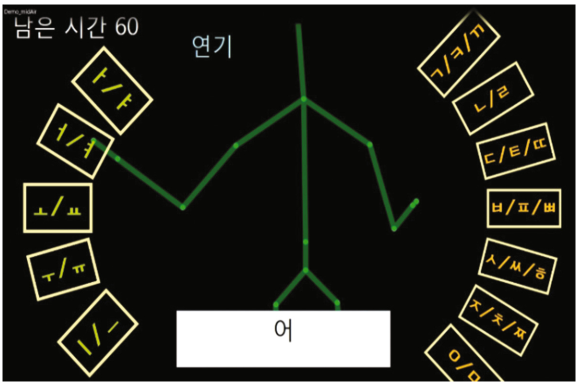


Fig. 3. Mid-Air Hangeul input system

After making adjustments toward the waveform layout, choosing buttons in all sectors including the upper part and the lower part became easier than before.

The following is the instructions of this Korean alphabet input system. Users can stay on a certain button containing multiple alphabets or choose a button through a swipe gesture and a certain alphabet can be typed by repeating the same gesture twice when one wishes to type the second alphabet of the button. Through this method, users can choose the alphabets they wish.

4 Evaluation

The experiment toward the usability of the Mid-air Hangeul input system on large screen was conducted. Measurements were based on characters per minute (CPM), ease of use, ease of learning, and satisfaction. For this experiment, ten participants (8 females and 2 males) had been gathered within university campus and the average age of the participants was 22.8 years old and they were familiar with computers. Also, 4 out of 10 participants have experienced playing a game using gestures.

The experiment was conducted by randomly setting the Korean alphabets widely used on the Hangeul Typing Certification Program and the contents of the experiment abide by the following.

Firstly, after the participants listened to a brief explanation of the instructions of the Korean alphabet input system, the typing speed was measured without preliminary practice. Secondly, after five minutes of practice for those who had completed the previous experiment, the typing speed was measured based on three one-minute tests which is the same as the first stage. Lastly, the participants filled out the questionnaire regarding three types of user satisfaction levels. The following is the questionnaire (Table 1).

Table 1. Mid-Air Hangeul input system Questionnaire.

Mid-air Hangeul input system questionnaire		
1	It is easy to use the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
2	It is easy to remember the instructions of the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
3	It was easy to learn the instructions of the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
4	It was fast to learn the instructions of the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
5	Using the Mid-air Hangeul input system does not require a lot of effort	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
6	Mistakes can be undone fast while using the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
7	It is interesting to use the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree

(*continued*)

Table 1. (*continued*)

Mid-air Hangeul input system questionnaire		
8	It was easy to type the Korean alphabets of my choosing	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
9	The Mid-air Hangeul input system was generally satisfactory	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
10	I will recommend the Mid-air Hangeul input system to my friends	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
11	It was difficult to use the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
12	It was difficult to remember the instructions of the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
13	It was difficult to learn how to use the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
14	It took a long time to learn how to use the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
15	I invested a lot of effort into using the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
16	It was difficult to undo the mistakes while using the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
17	It was boring to use the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
18	It was cumbersome to input the Korean alphabets of my choosing	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
19	I was generally dissatisfied with the Mid-air Hangeul input system	Strongly agree 1 2 3 4 5 6 7 Strongly disagree
20	I would not recommend the Mid-air Hangeul input system to my friends	Strongly agree 1 2 3 4 5 6 7 Strongly disagree

The questionnaire abided by the Likert 7-points scale with regard to the twenty questions which applied the determination of meaning difference method and included ten positive questions and ten negative questions. The questionnaire was assessed based on ease of use, ease of learning, and satisfaction.

4.1 Result

The following graphs - Figs. 4 and 5 - display the input speed and questionnaire results of the Mid-air Hangeul input system. Among the three types - ease of use, ease of learning, and satisfaction - ease of learning had the highest score and it shows the comparisons of typing speed before and after learning.

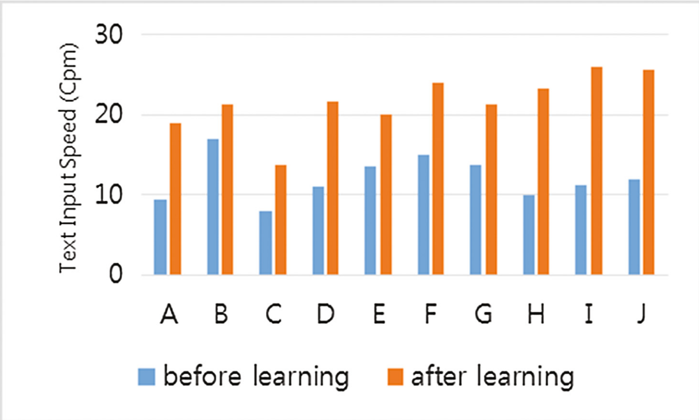


Fig. 4. Typing speed

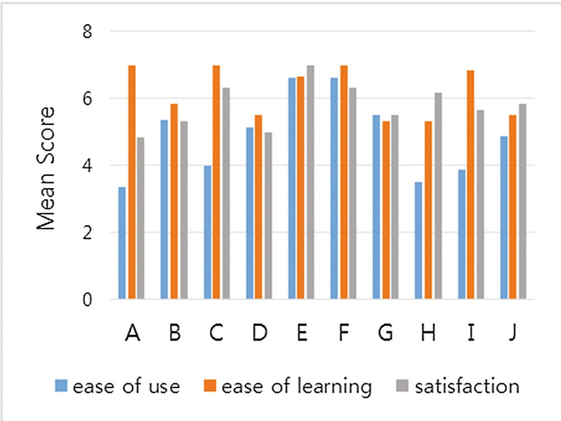


Fig. 5. Result of evaluation

Even though the input speed of all participants increased to the minimum level of 4.34 cpm to the maximum level of 14.78 cpm five minutes after learning, the typing speed of this system is still very low when compared to the experiment results of the traditional QWERTY layout keyboard form. The new form of layout is not only unfamiliar but also the typing efficiency declined due to the increase of keystrokes per character (KSPC) [16]. Furthermore, this is a comprehensive outcome which resulted from the distance from location of user to buttons in a virtual environment, the physical distance from device to user, etc.

However, when making comparisons between the measurement results of typing speed before and after learning and the relevance between user assessment results, this system, as the easy and intuitive input method, is expected to have a positive influence on the increase of user performance status.

5 Conclusions

This study is the Mid-air Hangeul input system which is available without any additional device except for human body part in a virtual environment which making direct contacts is difficult by users. Even though the typing speed of this system was markedly lower than the initial expectation setting when compared with those of the QWERTY layout or relevant studies, it is expected to bring about a positive effect through users' preference toward the system self-satisfaction in the available interaction method without the burden of learning. Based on the design of this system, in response to the rapidly changing computing environment, the objective is aimed at flexibly applicable system and future studies need to be conducted and expanded to achieve an effective system which faithfully carries out the roles - high input speed and accuracy - as an input system.

Acknowledgements. "This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-R2718-16-0017) supervised by the IITP (Institute for Information & communications Technology Promotion)".

References

1. Foley, J.D., Wallace, V.L., Chan, P.: The human factors of computer graphics interaction techniques. IEEE Comput. Graph. Appl. **4**(11), 13–48 (1984)
2. Oney, S., Harrison, C., Ogan, A., Wiese, J.: ZoomBoard: a diminutive qwerty soft keyboard using iterative zooming for ultra-small devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2799–2802. ACM, April 2013
3. Cha, J.M., Choi, E., Lim, J.: Virtual sliding QWERTY: A new text entry method for smartwatches using Tap-N-Drag. Appl. Ergon. **51**, 263–272 (2015)
4. Shibata, T., Afergan, D., Kong, D., Yuksel, B.F., MacKenzie, I.S., Jacob, R. J.: DriftBoard: a panning-based text entry technique for ultra-small touchscreens. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 575–582. ACM, October 2016
5. Hong, J., Heo, S., Isokoski, P., Lee, G.: SplitBoard: a simple split soft keyboard for wristwatch-sized touch screens. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1233–1236. ACM, April 2015
6. Gordon, M., Ouyang, T., Zhai, S.: WatchWriter: Tap and gesture typing on a smartwatch miniature keyboard with statistical decoding. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 3817–3821. ACM, May 2016
7. Cho, H., Kim, M., Seo, K.: A text entry technique for wrist-worn watches with tiny touchscreens. In Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology, pp. 79–80. ACM, October 2014
8. Komninos, A., Dunlop, M.: Text input on a smart watch. IEEE Pervasive Comput. **13**(4), 50–58 (2014)
9. Shneiderman, B.: The limits of speech recognition. Commun. ACM **43**(9), 63–65 (2000)

10. Sridhar, S., Feit, A.M., Theobalt, C., Oulasvirta, A.: Investigating the dexterity of multi-finger input for mid-air text entry. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3643–3652. ACM, April 2015
11. Feit, A.M., Sridhar, S., Theobalt, C., Oulasvirta, A.: Investigating multi-finger gestures for mid-air text entry, April 2015
12. Shoemaker, G., Findlater, L., Dawson, J.Q., Booth, K. S.: Mid-air text input techniques for very large wall displays. In: Proceedings of Graphics interface 2009, pp. 231–238, May 2009
13. Aoki, R., Maeda, A., Watanabe, T., Kobayashi, M., Abe, M.: Twist&tap: text entry for TV remotes using easy-to-learn wrist motion and key operation. IEEE Trans. Consum. Electron. **56**(1) (2010)
14. Ren, G., O'Neill, E.: Freehand gestural text entry for interactive TV. In: Proceedings of the 11th European Conference on Interactive TV and Video, pp. 121–130. ACM, June 2013
15. Markussen, A., Jakobsen, M.R., Hornbæk, K.: Vulture: a mid-air word-gesture keyboard. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 1073–1082. ACM, April 2014
16. MacKenzie, I.S.: KSPC (keystrokes per character) as a characteristic of text entry techniques. In: International Conference on Mobile Human-Computer Interaction, pp. 195–210. Springer, Heidelberg, September 2002

A Brain Wave Research on VR (Virtual Reality) Usage: Comparison Between VR and 2D Video in EEG Measurement

Sang Hee Kweon[1(✉)], Hea Ji Kweon[2], Se-jin Kim[3], Xuena Li[1], Xin Liu[4], and Hea Lin Kweon[5]

[1] Graduate School of Journalism and Mass Communication, SKKU, Seoul, Korea
skweon@skku.edu, 444814892@qq.com
[2] Graduate School of Business, Seoul National University, Seoul, Korea
hazzyjk@snu.ac.kr
[3] Graduate School of Journalism and Mass Communication, Korea University, Seoul, Korea
oosejinoo@naver.com
[4] Department of Journalism and Mass Communication, SKKU, Seoul, Korea
liuxinscda2011@163.com
[5] Department of Journalism and Mass Communication, Kyung Hee University, Seoul, Korea
healin123@naver.com

Abstract. Focusing on virtual reality (VR) and human factors, this study is designed to observe the differences between the receivers' cognition of VR and 2D videos through brainwave tests. The study will examine human computer interaction (HCI) trying to understand the characteristics and effects of VR by comparing the brainwave patterns of receivers when exposed to 2D and VR contents. The main purpose of this study is to grasp the differences in the human factors that affect the cognition of stereoscopic and flat images regarding their dimensional distinction by gathering concrete empirical data and comparing the alpha (α) and beta (β) wave patterns observed when watching VR and 2D videos. The study adopts a 2×3 experimental research design, statistically processed from 20 subjects. The 2D and VR materials used in this study is divided into three categories: sports, news, and advertisements. The participants viewed the videos in random order and the differences in brainwaves according to the genre and dimension were recorded. The results showed that β-wave vibrations observed when subjects were exposed to VR videos were statistically significant compared to when subjects watched 2D videos. The study also tested HCI factors on subjects that showed stronger in the frontal lobe when watching images with higher degree of VR stereoscopic effect and dynamic feeling. Electroencephalography (EEG) oscillations indicated that β-waves when watching VR were statistically significant compared to the brainwaves of viewing 2D videos. It was found that there were differences in β-waves according to program genre. Videos that were fast pace and more dynamic were more likely to show stronger

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_19

β-waves. There was especially high β-wave activity in the frontal lobe for the sports VR category.

Keywords: VR · VRTV · VR content · Laboratory study · Human factors · Brain waves · Alpha (α) wave · Beta (β) wave

1 Introduction

This study explores to test the differences in brain waves during visual image fusion acceptance or viewing of VR (virtual reality) and 2D (two dimensional "flat image"). The study follows a 2(2D and VR) × 3(genre) research design of the brain wave responses. Brain waves were measured using EEG responses by recording alpha (α) waves (8–12.99 Hz) and beta (β) waves (13–29.99 Hz). Different parts of the brain (frontal lobe, temporal lobe, occipital lobe) were also measured in order to compare the response difference to stimulus.

The experimental study was based on a statistical analysis on the EEG responses obtained through a total of 20 subjects. The brain wave difference between the 2D and VR contents were measured first. Then, responses were measured by 2 × 3 experimental design in order to measure the difference in brain waves according to the program type (genre). The subjects' brain waves were measured after viewing 2D and VR images. Image contents fall into the following categories: sports, news and advertisements (promotional videos).

2 Research Purpose

Our research mainly focuses on the human centric approach and responses in the consumption of VR contents by conducting a study on the viewer's perception on dimensional reality and genre. This study was designed to measure and verify the brainwave activity – the α waves (8–12.99 Hz) and β waves (13–29.99 Hz) – when watching VR and 2D videos and how the reactions differ in the frontal, temporal, occipital lobes of the brain according to the type of content. In detail, this study measures the size of the brain wave activity from both flat 2D images and VR dimensional images and compare the measurements according to genre recognition from the recipient.

3 Cognitive Research on VR Video

3.1 VR Videos and Human Factors

It is one thing to understand the engineering level of VR, but it is much more to understand the experience level of its audience. Beyond the visual dimension, there is much to research about the interaction and interests at the audience level of VR imagery. Studying how people view and understand VR contents provides not only industrial but also cultural significance of videos. Thus, basic research on effective measurement of VR videos' messages is required.

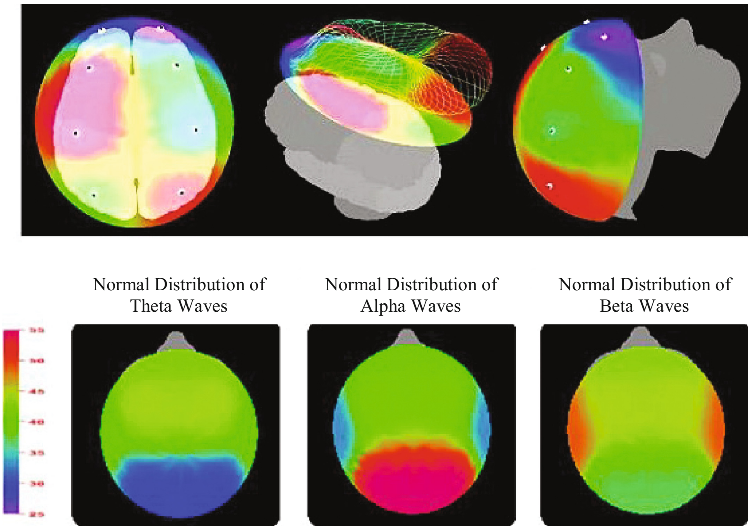


Fig. 1. Normal distribution of theta (θ), alpha (α) and beta (β) brainwaves

3.2 EEG as a Measurement of Human Factor and Awareness

Generally, EEG waves are divided depending on the range of the oscillation frequency (Fig. 1): delta (δ) waves (0.2–3.99 Hz), theta (θ) waves (4–7.99 Hz), alpha (α) waves (8–12.99 Hz), beta (β) waves (13–29.99 Hz), gamma (ζ) waves (30–50 Hz). All brain wave frequencies can be measured; however, for research purposes, waves that are more significant should be extracted and analyzed. Thus, only alpha (α) waves and beta (β) waves were used in this study.

Alpha waves are detected when the brain is at relaxed mental states. At this state, the brain feels more stability – more comfort – to increase the amplitude. Overall, regular waves do continue to appear the least but most significantly in the frontal, parietal, and occipital lobes. Alpha waves are often interpreted as relaxation, imagination, visualization, memory, learning and concentration.

Beta waves are most active in the frontal region of the brain, during conscious activities such as waking up and speaking. Beta waves also appear when doing complex calculations or feeling nervous. Often beta waves are associated with as logic, critical thinking, stress, anxiety and restlessness.

Since the head surface of the cerebral cortex frontal lobe, parietal lobe, temporal lobe, occipital, including each part divided into two different roles appears. The cerebral cortex is divided into four major sections, called “lobes”: the frontal lobe, the parietal lobe, the occipital lobe and the temporal lobe. The occipital lobe corresponds to the primary visual cortex. Thus, the occipital lobe is responsible for primary visual information processing. The parietal lobe is located near the parietal somatosensory cortex and is responsible for movement or sensory information processing.

Table 1. Brain wave and channels' function

CH 1	CH 2	CH 3	CH 4	REF
Right				
Red	Yellow	Green	Blue	Vibrant colors
Pre Frontal Lobe (β)	Frontal Lobe (β)	Temporal Lobe	Occipital Lobe (α)	Ear Ball
Complex Think		Hearing	View + Sensory Think	Standard Pole
CH 5	CH 6	CH 7	CH 8	GND
Left				
Orange	Purple	Grey	White	Black
Pre Frontal Lobe (β)	Frontal Lobe (β)	Temporal Lobe	Occipital Lobe (α)	Ear Ball
Complex Think		Hearing	View + Sensory Think	Standard Pole

By varying the type of video visible to the subject, through a channel wavelength change, it can confirm the structure and development of the brain function accordingly (Table 1). In particular, the characteristics that constitute alpha waves and beta waves inverse. If you watch footage with relatively high α-wave characteristics, there is intensive stimulus for emotional activity. On the other hand, watching a video triggers relatively high β-waves can stimulate complex thinking from the frontal temporal lobe.

4 Research Questions and Method

4.1 Research Questions

In this study, to learn about the differences for the human recognition between 2D video and VR, we will proceed with the following topic of VR as a more in-depth three-dimensional image. In practice, compared to 2D, the three-dimensional structure of VR often creates negative effects, such as fatigue, requiring further research.

- Research Question 1: How do α-waves and β-waves differ for 2D and VR videos?
- Research Question 2: Will different wavelengths in two channels appear for 2D and VR videos?

4.2 Research Method

Due to the complexity of the experiment design, in order to minimize trial and error, a pre-experiment was carried out in February 2016 and a re-designed experiment was conducted on October 9, 2016. We designed a pre-experiment to rule out the participants who felt especially uncomfortable with VR beforehand. For the experiment, a LG 40-inch CRT TV and Oculus Virtual HDM were installed and used. Participants consisted of a random section of 20 university students in Seoul, Korea.

The experiment conducted with three types of footage. Videos are categorized depending on their contents: (1) Sports (2) News (3) Advertisement/PR. Based on the nature of an experimental study as it cannot cover all genres, selection criteria of the three genres was centered on image characteristics for the purpose of this study. Moreover, in order to eliminate the effect of order, all videos were shown randomly to the participants.

5 Research Results

5.1 Data Analysis

The following results (Table 2) were analyzed based on the first research question. Using a paired t-test, the difference in alpha waves and beta waves between VR and 2D images can be compared respectively.

Table 2. Mean Difference between 2D and VR contents

Classification		2D	VR	t	Sig.
α-wave	Sports	17.41	11.39	3.01	.017
	News	17.46	12.001	2.64	.029
	Advertising	53.23	10.55	5.23	.003
β-wave	Sports	23.92	73.64	−8.39	.000
	News	24.40	67.06	−3.03	.018
	Advertising	33.04	58.85	−2.67	.027

The average waveform of alpha (α) waves and beta (β) waves were recorded for each image dimension and genre. Overall, it appears that α-waves were higher for flat images and β-waves were significantly higher for VR images. In the case of α-waves, all t-values are greater than zero and hold a significant probability lower than 0.05. Therefore, in terms of α-wave activity, 2D videos display a higher value than virtual reality videos. Thus, participants were more likely learning and "concentrated" when watching 2D videos. For the β-waves, all video types resulted negative t-values and the significant probability is greatly lower than 0.05. In terms of β-wave activity, VR videos hold greater value than 2D videos. For the participants, VR images overall created strong stimulus in the brain.

Next, we analyzed the data focusing on the second research question (Tables 3 and 4).

It can be seen that alpha waves appear higher for 2D videos in all three genres. Considering alpha wave EEG is associated with relaxation, corresponding to alpha waves the occipital lobe (channel 4 and channel 8) appears to be most active. Moreover, as advertisements are fast-paced and have rapid screen transitions requiring more concentration, alpha waves occurred higher. Sports and news videos had slower flows and fewer transitions.

Table 3. Comparison of alpha waves between 2D and VR (absolute value)

Classification			2D	VR	t	Sig.
Alpha (α) wave	Sports	Ch. 1	24.66	13.39	6.534	.000
		Ch. 2	26.12	13.63	6.127	.000
		Ch. 3	12.94	11.57	5.418	.000
		Ch. 4	13.19	11.26	4.583	.001
		Ch. 5	13.27	9.33	2.196	.056
		Ch. 6	12.89	10.82	4.719	.001
		Ch. 7	12.77	10.62	3.964	.003
		Ch. 8	15.40	12.15	3.482	.007
	News	Ch. 1	27.94	13.36	3.686	.005
		Ch. 2	28.74	13.58	3.906	.004
		Ch. 3	12.82	10.46	6.315	.000
		Ch. 4	13.39	10.37	5.953	.000
		Ch. 5	7.682	9.98	4.469	.002
		Ch. 6	11.91	10.57	3.827	.004
		Ch. 7	11.73	9.56	4.566	.001
		Ch. 8	23.87	10.13	2.094	.066
	Advertising	Ch. 1	45.65	12.69	1.879	.093
		Ch. 2	52.23	13.06	1.697	.124
		Ch. 3	19.37	10.20	2.509	.033
		Ch. 4	25.06	10.24	2.012	.075
		Ch. 5	53.67	8.65	1.121	.291
		Ch. 6	83.19	9.74	2.144	023
		Ch. 7	18.15	9.54	5.064	.061
		Ch. 8	48.90	10.12	3.144	.001

As mentioned before, VR videos stimulate the frontal lobe and complex thinking other than visual and auditory stimulation. Beta waves appeared most for sports genre and signaled intense brainwave stimulation. The waves appeared less for promotional and news video leading to lower stimulation.

For all three genres, it can be seen beta wave were all significantly higher for VR video viewing. The frontal lobe and prefrontal lobe channels 1, 2, 5, 6 had higher beta wave values: this was significantly high for channel 5 of the prefrontal lobe. In addition, it can be assumed due to strong visual effects and sounds of virtual reality, the associated function channels (channel 4 of the occipital lobe and channel 3 of the temporal) also show interesting results.

Table 4. Comparison of beta waves between 2D and VR (absolute value)

Classification			2D	VR	t	Sig.
Beta (β) wave	Sports	Ch. 1	35.17	81.78	−2.401	.040
		Ch. 2	33.16	85.05	−2.533	.032
		Ch. 3	17.62	59.65	−1.674	.128
		Ch. 4	14.18	49.75	−1.862	.096
		Ch. 5	39.37	121.73	−1.488	.171
		Ch. 6	25.66	79.13	−2.173	.058
		Ch. 7	11.46	41.40	−1.621	.140
		Ch. 8	14.74	54.57	−2.019	.074
	News	Ch. 1	42.57	77.02	−1.273	.235
		Ch. 2	39.78	83.35	−1.110	.296
		Ch. 3	20.20	40.47	−2.393	.040
		Ch. 4	16.10	30.82	−3.838	.004
		Ch. 5	28.95	134.80	−3.168	.011
		Ch. 6	22.10	92.75	−2.673	.026
		Ch. 7	15.65	25.32	−9.242	.000
		Ch. 8	25.77	27.87	−13.067	.000
	Advertising	Ch. 1	43.67	88.60	−2.921	.017
		Ch. 2	46.55	92.54	−2.098	.065
		Ch. 3	24.23	40.73	−4.228	.002
		Ch. 4	22.11	30.47	−5.408	.000
		Ch. 5	45.27	99.91	−2.934	.017
		Ch. 6	54.43	51.68	−5.671	.000
		Ch. 7	16.21	24.16	−6.311	.000
		Ch. 8	27.83	26.62	−8.723	.000

5.2 Brainwave Map

The following are example brainwave maps (Fig. 2) of the subjects when watching footages of 2D and VR. Throughout the brain, higher wave activity would be warm and active colors such as red or magenta. Less brainwave activity was shaded in blue or cool colors. Alpha waves were certainly active for 2D videos, especially sports and advertisements. Beta waves were significantly high for all genres in VR. Beta wave activity was also high for 2D news and even greater for 2D promotional videos.

Fast beta waves are powerful brainwaves and highly stimulating producing alertness, efficiency, test taking skills, multitasking, and high performance. However, too many “small” beta waves lead to discomforts such as anxiety, tension, stress and sleeping disorders. Further research on such beta brainwave effects is required in the future. As a limitation of this study, the amplitude (size: large or small) of brainwaves was not measured and could not be considered.

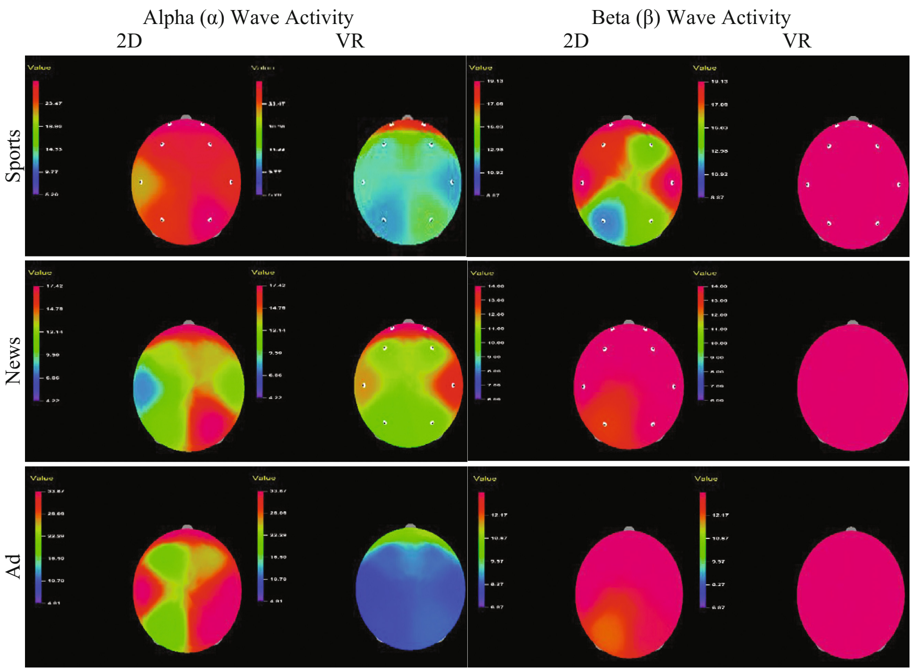


Fig. 2. Alpha (α) and beta (β) brainwave maps (categorized by video dimension and genre)

6 Conclusion and Implication

6.1 Summary and Conclusion

Following McLuhan's famous assumption "the medium is the message", VR is the message of future media. This study looks into the brainwave stimulus participants experienced when watching different types of VR videos. These stimulations and experiences shape the viewer's perception of virtual reality. Overall, VR videos had high beta brainwave activity for all genres: sports, news and advertisements. VR videos, having more dimensional and fast transitions, resulted in higher EEG activity. Beta waves appeared the highest for sports VR videos, perhaps due to the dynamic movements and rapid pace of sports and such video being in higher dimension. The excitement and liveliness is express vividly through VR than 2D. More senses and thinking is stimulated when watching VR; making videos feel more realistic to viewers, overall enhancing the experience of the video itself.

6.2 Implications of the Study

The implications of this study are to empirically confirm that the perception of VR videos accommodate different EEG activity and processes, compared to 2D. The results researched about human brainwaves in image or video recognition affect the viewer and, furthermore, the audience of videos. Such factors considered in this study can be applied to future experimental studies, e.g. physiological response studies, and video entertainment industry. Physical discomforts such as dizziness, headaches, fatigue and negative emotional activity such as anxiety and stress often occur when watching VR videos. The high levels of beta brainwave activity from our experiment confirm that. Further research on improving and overcoming such discomforts is needed for future use and development of VR videos.

Acknowledgements. This research is supported by SungKyun Research Fund 2016–2017.

References

1. Barfield, W., Weghorst, S.: The sense of presence within virtual environment: a conceptual framework. In: Proceedings of the Fifth International Conference on Human-Computer Interaction, vol. 5, pp. 699–704 (1993)
2. Emoto, M., Niida, T., Okano, F.: Repeated vergence adaptation causes the decline of visual functioning watching stereoscopic television. J. Disp. Technol. **1**, 328–340 (2005)
3. Cho, E.J., Kwan, M.L., Yang, H.C. Effects of stereoscopic movies: the position of stereoscopic objects (PSO) and the viewing condition. In: AEJMC 2011 Boston Conference Paper (2011)
4. Fukuta, T.: New electronic media and the human interface. Ergonomics **33**(6), 687–706 (1990)
5. Heeter, C.: Being there: the subjective experience of presence. Presence Teleoper. Virtual Environ. **1**, 262–271 (1992). MIT Press
6. Heeter, C.: Communication research on consumer VR. In: Biocca, F., Levy, M. (eds.) Communication in the Age of Virtual Reality, pp. 191–218. Lawrence Erlbaum Associates, Hillsdale (1995)
7. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S. Vengeance-accommodation conflicts individual performance and cause visual fatigue. J. Vis. **8**(3), 33 (2008). http://journalofvision.org/8/3/33/. doi:10.1167/8.3.33
8. Jerald, J.: The VR Book. ACM Morgan & Claypool, New York (2014)
9. Ijsselsteijn, W.A., de Ridder, H., Hamberg, R., Bouwhuis, D., Freeman, J.: Perceived depth and the feeling of presence in VRTV. Displays **18**, 207–214 (1998)
10. Ijsselsteijn, W.A., de Ridder, H.: Measuring temporal variations in presence. Presented at the 'presence in shared virtual environments' workshop, BT Laboratories, Ipswich, UK, 10–11 June 1998
11. Ijsselsteijn, W.A., Riva, G.: Being there: concepts, effects and measurement of user presence in synthetic environments. In: Riva, G., Davide, F., IJsselsteijn, W.A. (eds.) Being There: The Experience of Presence in Mediated Environments. Ios Press, Amsterdam (2003)
12. Lang, P.J.: Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski, J.B., Johnson, J.H., Willians, T.A. (eds.) Technology in Mental Health Care Delivery Systems, pp. 129–139. Ablex, Norwood (1980)
13. Livingstone, M., Hubel, D.: Segregation of form, color, movement, and depth: anatomy, physiology, and movement. Science **240**, 740–749 (1988)

14. Lombard, M., Reich, R.D., Grabe, M.E., Campanella, C.M., Ditton, T.B.: Big TVs, little TVs: the role of screen size in viewer responses to point-of-view movement. Paper presented to the mass communication division at the annual conference of the international communication association, Albuquerque, NM, May 1995
15. McLuhan, E., Zingrone, F. (eds.): Essential McLuhan. Basic Books, New York (1995)
16. McLuhan, M.: The Gutenberg Galaxy: The Making of Typographic Man. University of Toronto Press, Toronto (1962)
17. McLuhan, M.: Understanding Media: The Extension of Man. McGraw-Hill, New York (1964, 1985)
18. Meyrowitz, J.: No Sense of Place: The Impact of Electronic Media on Social Behavior. Oxford University Press, New York (1986)
19. Neuman, W.R, Crigler, A.C., Bove, V.M.: Television sound and viewer perceptions. In: Proceedings of the Joint IEEE/Audio Engineering (1991)
20. Nielsen, J.: Usablity Engineering. Academic Press, San Diego (1993)
21. Seuntiens, P., Vogels, I., van Keersop, A.: Visual experience of VR-TV with pixelated ambilight. In: PRESENCE (2007)
22. Sheridan, T.B.: Musings on telepresence and virtual presence. Presence Teleoper. Virtual Environ. **1**(1), 120–126 (1992)

Design of a Haptic Virtual System for Improving Fine Motor Skills in Children with Autism

Huan Zhao[1(✉)], Zhaobo Zheng[3], Amy Swanson[2], Amy Weitlauf[2], Zachary Warren[2], and Nilanjan Sarkar[1,3]

[1] Electrical Engineering and Computer Science Department, Vanderbilt University, Nashville, TN 37235, USA
Huan.Zhao@vanderbilt.edu

[2] Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University, Nashville, TN 37235, USA
{Amy.R.Swanson, Amy.S.Weitlauf, Zachary.E.Warren}@vanderbilt.edu

[3] Mechanical Engineering Department, Vanderbilt University, Nashville, TN 37235, USA
{Zhaobo.Zheng, Nilanjan.Sarkar}@vanderbilt.edu

Abstract. Fine motor skills, including gripping, holding and reaching, are useful and necessary in fundamental daily activities. However, children with autism spectrum disorders (ASD) show deficits and delays in fine motor skills compared to their peers. In this paper, we propose a virtual training system enabling the assessment and improvement of fine motor skills for children with ASD through haptic interaction in a low-cost and engaging virtual environment with real-time feedback (e.g., haptic, auditory and visual feedback). This system consists of a Haptic Gripper allowing users to manipulate (e.g., grip and move) virtual balls in the games and thus provides opportunities for users to improve finger and hand motor control. We present the system design, and a small usability study which verified the system functionality and indicated the potential of this system in improving fine motor skills of the users in move and grip manipulations.

Keywords: Fine motor skills · Haptic interaction · Virtual training system · Autism Spectrum Disorders (ASD)

1 Introduction

Autism spectrum disorders (ASD) characterized by deficits in communication and social interaction, represent a range of neurodevelopmental disorders [1, 2]. In addition to the core deficits, a growing number of research work has evidenced that children with ASD show motor deficits and delays regarding fine motor skills compared to typically developing (TD) peers [3]. Fine motor skills are commonly used in daily activities (e.g., using a key and drawing) that require coordination between fingers, hands, arms and visual perceptions [4]. Several studies have found that children with ASD show

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_20

abnormalities in precision grasping and reaching [5], eye-hand coordination [6], and motor planning and control [7]. It is believed that deficits in fine motor skills may negatively affect children with ASD in physical, scholastic and social activities [8, 9].

The growing applications of virtual reality (VR) technology in multiple types of ASD interventions [2, 10] suggest the potential of VR systems to engage children with ASD in a replicable, motivating and low-cost training platform with real-time feedback. Haptic devices with the capabilities of providing mechanical force feedback are beneficial in increasing immersion and improving quality of task performance. Several studies have demonstrated that haptic feedback can enhance training achievements [11–13]. For example, Broeren et al. [14] found that training in virtual reality tasks with haptics could improve motor rehabilitation after stroke. Kim et al. [15] designed a training system for improving children's handwriting skills with haptics and observed improvement in tracing precision. However, few haptic virtual systems are available for fine motor skill training of children with ASD [16, 17]. In this context, we focus on developing a virtual training system with haptic feedback for children with ASD to provide practice opportunities for the purpose of improving fine motor skills, in hopes of expanding the accessibility of motor interventions for children with ASD.

In this paper, we present the preliminary design of a haptic virtual fine motor skill training system and a small usability study. The system consists of a customized interactive tool, called Haptic Gripper, and a series of virtual games requiring the users to implement finger and hand manipulations (Fig. 1). The system is capable of automatically recording objective and quantitative data regarding the user's hand manipulative performance. The contributions of our current work are two-fold: (1) we provide a novel haptic virtual fine motor skill training system, with which the user can assess and practice fine motor skills via virtual games; and (2) we tested the system with a small usability study and found that participants found the system interesting and performed better after the training.

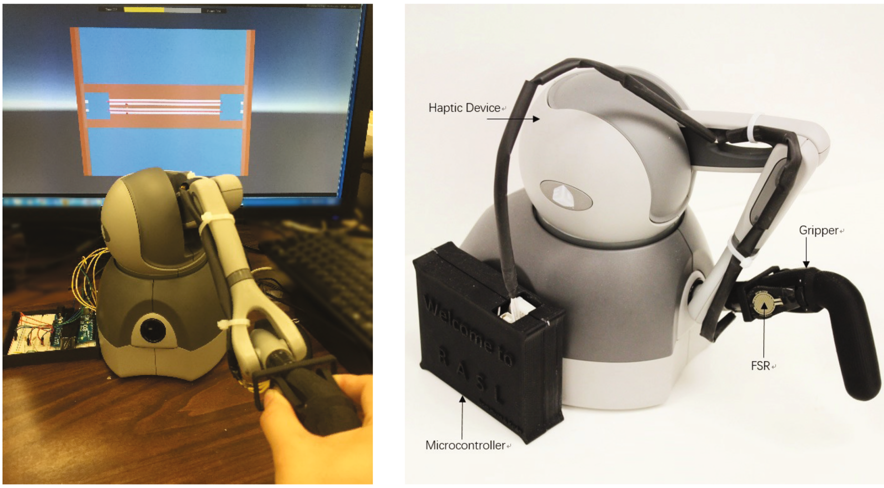


Fig. 1. The haptic virtual fine motor skill training system (*left*) and the Haptic Gripper used in this system (*right*).

2 System Design

The core of the proposed system is the virtual games that were designed in such a way that the user needs to apply appropriate finger pressure to adjust the inner-distance of two virtual balls and simultaneously move these balls through virtual paths toward targets. The user interacts with these games via the Haptic Gripper, which can detect the spatial position of the user's hand and finger pressure as well as apply haptic feedback on the user's hand. The haptic feedback combined with visual and audio feedback provides real-time guidance that helps the user to make appropriate manipulation adjustments in time.

2.1 Design and Use of the Haptic Gripper

The Haptic Gripper was constructed by augmenting a commercial haptic device with a customized gripper consisting of force sensing resistors (FSRs) (Fig. 1).

We chose the Geomagic Touch Haptic Device [18] as the tool to manipulate the virtual balls with the hand and give responses in the form of resistance. The device supports motion in 6 degrees of freedom (workspace: 6.4 W $\times$ 4.8 H $\times$ 2.8 D inch) and is able to provide an accurate measure of 3D spatial position (nominal position resolution: 0.055 mm). Additionally, it can apply force feedback, allowing user to feel virtual objects. In this system, we utilized the Haptic Plug-In [19] for Unity 3D [20] to develop the functionalities of haptic device.

To obtain finger pressure data, we made a 3D-printed gripper with FSRs to replace the stylus of Geomagic Touch Haptic Device. The size and functionality of the gripper were designed considering the hand features of children [21]. The gripper consists of two press plates and a hollow handle. To use it, the user holds the handle and presses his/her thumb and index finger on the press plates, where two FSRs are fixed (Fig. 2). We used the FSRs from Interlink Electronics [22] and connected the FSRs to an Arduino microcontroller [23], which gets the finger pressure data from FSRs, performs data filtering and then transfers the data to the games. The FSRs have 14.7 mm diameter active area and have the force sensitivity range from 0.2N to 20N, which are suitable for our application.

2.2 Design of the Virtual Games

We designed a series of virtual games that require the user to move two virtual balls in a group along several specific paths trying to reach the targets as quickly as possible while avoiding hitting the wall, with the purpose of strengthening the precision grip skills as well as the ability to generate steady motion.

Game Controls. In the games, the user is asked to manipulate the two balls, which are moving along the parallel trajectories. The reason we designed two controlled balls instead of one is that it is easier to reflect the grip manipulation through the relative position of two balls, and simultaneously it increases the requirement in hand-eye coordination. The user can control the balls through two forms of manipulations:

(1) moving the Haptic Gripper to adjust the 2D location of the balls; and (2) gripping the press plates of the Haptic Gripper to adjust the vertical distance of the two balls.

The movement manipulation is easy to perform. When the user holds the gripper and moves around, the vertical center of the two balls will follow the motion trace of the gripper. However, the motion of balls is constrained to lie along a 2D plane, while the gripper is allowed to move freely in the 3D workspace.

The grip manipulation is performed depending on the applied finger pressure. Because many fine motor activities (e.g., handwriting) require low-level forces, the gripper was designed to provide low-level grip control [24]. To simplify the grip manipulation, we divided the finger pressure into three ranges: (1) (0N, 3.57N); (2) (3.57N, 5.23N); (3) (5.23N, ∞), according to the positions of the press plates (Fig. 2). Three levels of finger pressure correspond to three specific vertical distances between the two balls. When the user grips the press plates with a small pressure that falls within the first range, the press plates remain in the original positions and the vertical distance between the balls will not change. As the user increases the finger pressure to reach the second pressure range, two press plates as well as the two balls will move closer to each other. And when a large finger pressure is applied to make the press plates tightly touch the gripper body, the vertical distance between the balls becomes further closer. In this work, as a measure of impact on performance, we used games requiring the medium finger pressure as the pre- and post-tests in the usability study to test the change of the user performance.

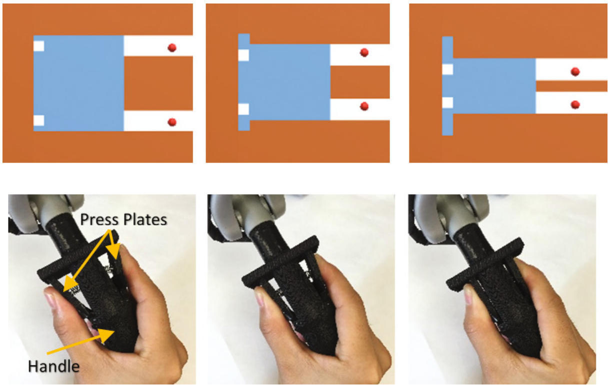


Fig. 2. The pictures in the first row show the corresponding relative positions between the balls as the user separately applies small (*left*), medium (*middle*) and large (*right*) finger pressure on the press plates of the Haptic Gripper using thumb and index finger.

Figure 3 gives an example of the virtual games. To complete a game, the user should move balls through white paths to first reach the right targets (white squares attached to the walls) and then come back to reach the left targets. First, as the vertical distance of the white paths varies in different games, the user should adjust the vertical distance between the balls by gripping the press plates to fit the vertical distance of two paths in order to make sure both balls can pass through the paths. In addition, moving balls as well as adjusting the vertical distance of two balls should be careful and stable avoiding hitting the wall if the user wants to achieve a high score. Once the balls hit the wall, the haptic Gripper will respond to provide the resistance feedback on the user's hand, which prevents the user from moving the balls in a direction that is not permitted due to constraints in the game. In addition, a warning text "−1" is shown and a "crash" audio is sounded to inform that the collision happened and the score decreases by 1. To help the user stay within the allowed paths, the paths in some games are marked with the visual cues (red stripes) where the balls are least likely to hit the wall.

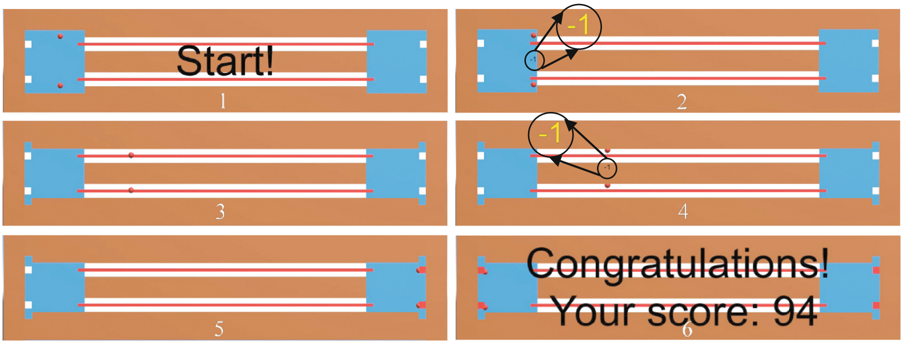


Fig. 3. A virtual game that requires medium grip pressure to manipulate the balls through the horizontal white paths. When the balls hit the wall, a warning text "−1" is shown to inform that the user's score decreases by 1 (see the picture 2 and 4). The game is completed after the balls touch the right and left targets (white squares that become red once being touched).

Game Variations. With the goal of encouraging the user to play the games in different hand postures, we developed 24 such games (Fig. 4) differentiated by the requirement levels on (1) finger pressure, (2) move orientation, and (3) eye-hand coordination. Accordingly, we adjusted the game configurations in terms of (1) the vertical distance between the white paths; (2) the orientation of the white paths; and (3) the existence of visual cues.

The change in vertical distance between the parallel paths requires an adjustment of vertical distance between the balls, which is controlled by the finger pressure and needs skills in grip force control. The orientation variable requires users to move the balls along different directions. And visual cues are provided in half of the games in the form of marking the optimal routes with the red stripes (the central lines within the paths) that help reduce the risk of wall-hitting. Without visual cues, it will become a little harder to move steadily in the allowed areas or determine which level of grip pressure to apply when the paths change. Figure 4 shows 12 game variations with visual cues created from altering the vertical distance and the orientation of the paths. Additionally, another 12 game variations without visual cues generated a total of 24 training games.

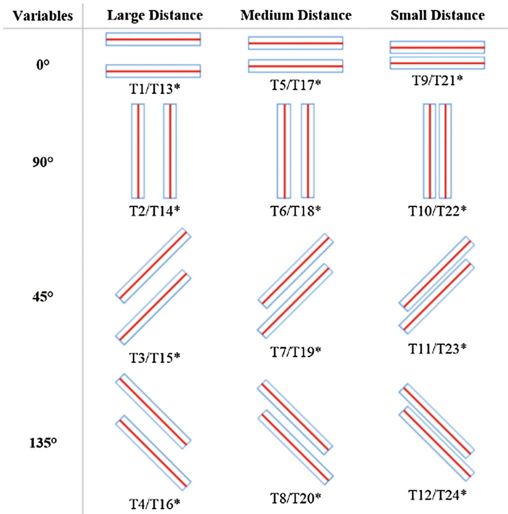


Fig. 4. Twelve game variations with visual cues (marked with red stripes). The game tags are shown below the game variations. The tags with "*" indicate the games with the similar shapes but without visual cues.

3 Usability Study

We conducted a usability study with 6 TD volunteers and 4 children/adolescents with ASD. The study was approved by the Institutional Review Board of Vanderbilt University. It is to be noted that this study was not designed to demonstrate improvement in motor skills. Instead it was a study to observe how the users felt interacting with the system and document their performances. A clinical study with protocols to improve motor skills will be designed in the future.

Participants. We recruited two groups of participants for this usability study. The first group included 6 TD volunteers (3 males and 3 females, age: Avg = 26.17, SD = 2.27), while the second group included 4 children/adolescents with ASD (2 males, ages: 17 and 14 years; 2 females, ages: 9 and 7 years). Each participant with ASD was given a $25 gift card as rewards.

Procedure. Each participant completed three sessions: a pre-test, an adaptive training session and a post-test. The pre- and post-test were short sessions consisting of 4 games (T17, T18, T19 and T20). These games required participants to grip and move balls through the paths with medium vertical distance but without the visual cues, making them more difficult to complete than the other virtual games. We used pre- and post-test to evaluate the impact of the virtual games on the participants by comparing their performances.

The adaptive training session contained 24 game variations and repeated some of these games to strengthen relevant skills of the participant. First, this session asked the participants to finish the 24 basic games and simultaneously recorded each game's score with the game tag. Based on a predefined score threshold (a threshold score of 90 was used in this usability study), finished games were divided into the "good performance" group (score $\geq$ 90) or the "poor performance" group (score < 90). Then, the "poor games" would be repeated for the participant to do more practices until his/her performance improves to a certain level (score $\geq$ 90) in these games. The participant thus got more opportunities to improve the weakness in targeted motor skills.

At the beginning of each experiment, the experimenters introduced the system and explained how to play the games to participants. Then participants would play a practice game (T1) to get familiarized with the system and games. In the end, each participant filled out a survey to give feedback.

4 Results

4.1 Participants' Feedback

There were seven questions included in the survey asking the participants' feedback on using the system. The survey results are shown in Table 1.

Table 1. Participants' feedback from the surveys.

Questions	TD Group	ASD Group
	Avg1 (SD)	Avg2 (SD)
1. How much did you like the games?	4.17 (0.69)	4.25 (0.69)
2. How easy was it to understand how to play games?	4.67 (0.47)	3.75 (0.69)
3. How easy was it to control the Haptic Gripper?	3.67 (0.47)	3.75 (1.19)
4. How much did the visual cues help you to perform stable manipulations?	4.17 (0.37)	4 (0.5)
5. Did you feel the haptic feedback? And will you try to adjust your manipulation when feeling the haptic feedback? a) Yes, Yes (TD: 4 | ASD: 3)[a] ; b) Yes, No (1 | 1) ; c) No, No (1 | 1).		
6. Compared to when you first started playing the games, how did you do by the end? a) Better (4 | 3); b) Same (2 | 1); c) Worse (0 | 0).		
7. Which games were more difficult for you? T3(1 | 0), T4(0 | 1) T5(4 | 1), T6(1 | 1), T7(2 | 2), T8(3 | 1), T9(1 | 2), T10(1 | 2), T11(2 | 2), T12(2 | 2), T15(1 | 0), T16(1 | 1), T17(4 | 1), T18(1 | 1), T19(3 | 2), T20(3 | 1), T21(1 | 2), T22(1 | 2), T23(2 | 2), T24(1 | 2)		

a. The answers for question 5-7 are shown in the form of "(the number of TD | the number of ASD)".

The first four questions were scored on a five-point scale (where 5 = "very much/easy", 1 = "not at all/very difficult"). In the following discussion, Avg1 and Avg2 stand for the average score obtained by the TD group and the ASD group, respectively. First, participants in both groups liked these games (Avg1 = 4.17, Avg2 = 4.25) and they felt that the games were easy to learn (Avg1 = 4.67, Avg2 = 3.75). Thus the games we designed seemed to be attractive and suitable for children with ASD. Second, they could learn how to manipulate the Haptic Gripper very well in a short time (Avg1 = 3.67, Avg2 = 3.75). However, we still found that due to the gripper size, it took a little longer for the younger participants to perform grip adjustments, especially when they tried to apply large finger pressure. In addition, the participants perceived the visual cues as helpful to guide and improve their manipulations (Avg1 = 4.17, Avg2 = 4).

The remaining three questions were choice questions. The question 5 asked about the haptic feedback. Most of participants (4 out of 6 in the TD group and 3 out of 4 in the ASD group) answered that they could feel it and it would affect their manipulations, which indicated that the haptic feedback was helpful to guide the participants' manipulations. As for the question 6, most participants (4 out of 6 in the TD group and 3 out of 4 in the ASD group) agreed that they played better by the end demonstrating their confirmations of the usefulness of these games in helping improve hand manipulations. In the next section, we will see that the objective performance analysis also demonstrated the performance improvement.

The last question required each participant to choose the most difficult games out of 24 basic virtual games based on self-assessment. Except the games T1, T2, T13, T14, all the games were referred to as difficult games by different participants. Some participants felt the games were challenging in specific move orientations, while some did not play well when requiring specific finger pressures or without visual cues. Their choices demonstrated individual differences of hand manipulation capabilities. However, we still can see that the games with the orientation patterns of 45°, 135° or with medium separating distance were more difficult for the participants. We found that the participants' answers were basically consistent with their performances.

4.2 Participants' Performances

We evaluated the participants' performance based on four metrics as below:

- Complete Time: it was the total time the participant used to complete a game. It suggested the difficulty of the game for the participant;
- Game Score: it was computed as (50 points for touching the right targets + 50 points for touching the left targets - times of ball collision with the wall). If the ball touched the wall for a long time, the score would decrease by 1 every one second. The participants were given enough time (120 s) to complete each game. So the main challenge for the participants was to avoid hitting the wall;
- Motion Trace Error: it was the root-mean-square error (RMSE) of the motion trace relative to the optimal route. It indicated the motion stability;
- Force Trace Error: it was the RMSE of finger pressure trace relative to the median value of the pressure data. It demonstrated the ability of the participant to maintain the certain finger pressure during the motion.

Because the difficulty of each game differed for the participants, we separately analyzed the performance results (Figs. 5 and 6) for each game in the pre- and post-test. First, we can see that the average complete time decreased and the average game score increased in the post-test for both groups. It indicated that both the TD participants and the participants with ASD improved their hand control of the virtual balls and reduced collisions on the wall after the training session. At the same time, they moved the balls faster and thus shortened the game complete time in the post-test. To compare the Group 1 and Group 2, we found that participants with ASD in Group 2 showed greater changes regarding the time (decreased by up to 66.21%) and score (increased by up to 24.76%). They achieved relatively lower scores and longer time than the participants of Group 1 in the beginning but finally reached the levels similar to or even better than Group 1. It is reasonable to see that adult TD participants played very well from the beginning, because they were more proficient in hand manipulations than children and these games were easier for them.

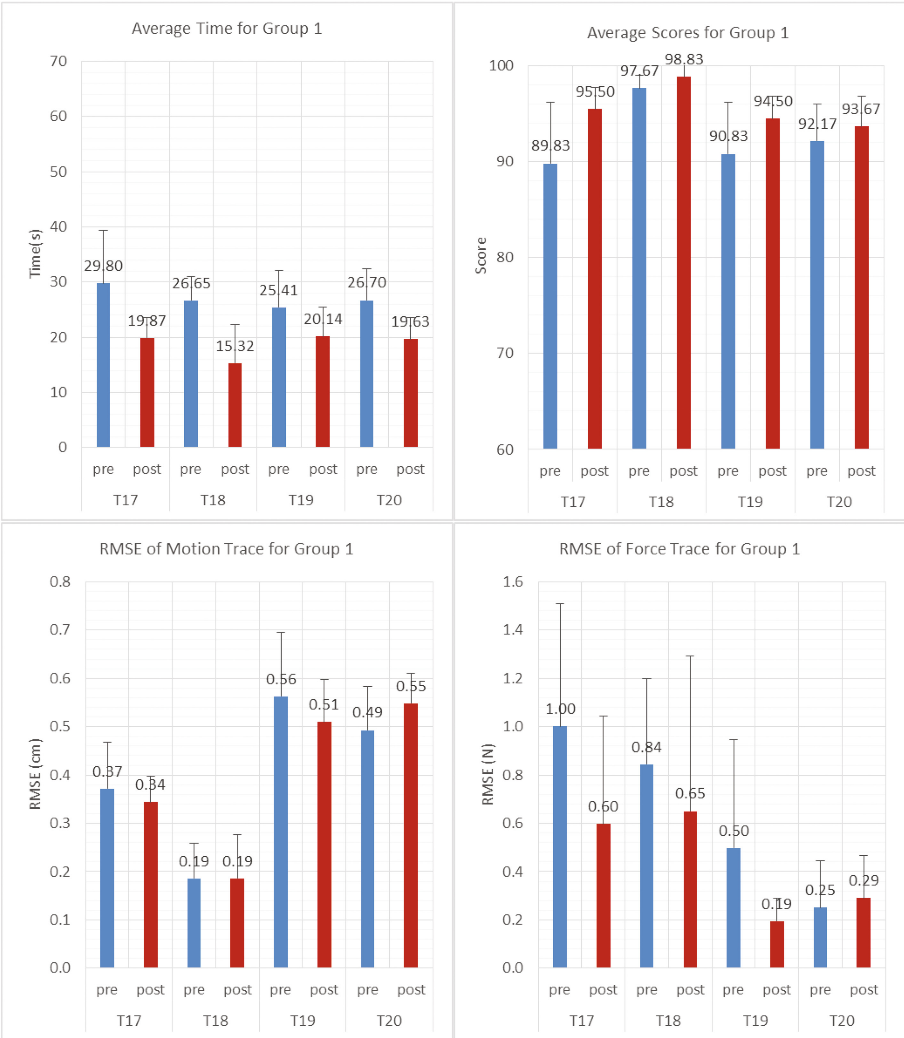


Fig. 5. The performance results of TD participants in Group 1.

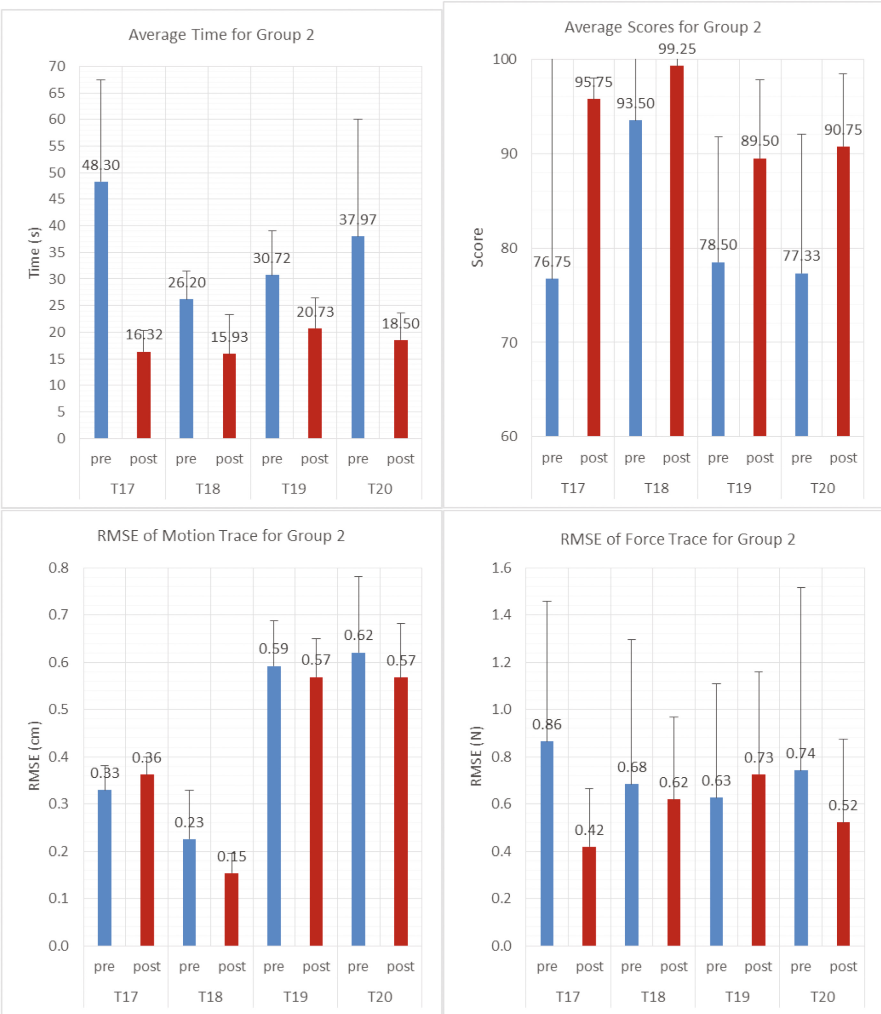


Fig. 6. The performance results of participants with ASD in Group 2.

We also found that participants in both groups played better in the game T18 (path orientation: 90°) than in the games T17 (0°), T19 (45°) and T20 (135°). The results were in agreement with the responses to the question 7 of the survey, where most participants reported that the games with the path orientation of 45° or 135° as the most difficult games. It revealed that playing these games could help participants understand and evaluate their hand manipulation skills. And the reason that all the participants played better in the game T18 was that the influence of the balls' gravity in hand control was reduced in the vertical direction. Thus, only a bit of effort was required for the participants to hold the balls.

In addition, though participants were not forced to move along the optimal routes, or to grip with a fixed pressure, decreased motion trace error and force trace error were found in some games, indicating that participants moved the virtual balls more smoothly and applied more stable finger pressure in the post-test.

5 Discussions and Limitations

The goal of our current work is to explore if a haptic virtual system that provides virtual games requiring fine motor skills that could help measure and eventually improve fine motor skills of children with ASD. The usability study provided preliminary evidence of our system's capability of providing fine motor skill training games in an engaging way, as well as collecting objective data for the performance assessment. The participants' feedback suggested that these training games were interesting and easy to play, and these games were helpful for the participants to understand and improve hand manipulation behaviors via visual, auditory and haptic feedback. The performance analysis also proved the positive impact of the system on the participants. The increased score, reduced completion time, and mostly reduced motion trace error and force trace error demonstrated the potential of this system for improving the user's fine motor skills, especially in gripping, holding, moving and reaching objects.

There were some limitations regarding the study as well as the design of the system. First, the system was only validated with a small group of participants with ASD. A follow-up user study with more participants with ASD will be conducted to investigate whether their fine motor skills significantly improve when they are trained by this system. We also want to include age-matched TD children in the future user study to explore developmental differences in motor skills between these two groups. Second, we will redesign the Haptic Gripper and make it more comfortable and adjustable for the users.

In addition, the virtual games in the current system were designed only for one user. We have been working on creating collaborative games with haptic feedback for two users in order to test and foster fine motor skills involved in the team work, as well as improve communication and collaborative skills among users. Based on the simple virtual game that was introduced in this paper, we also intend to develop games with complex paths requiring complex hand manipulations.

6 Conclusion

In this paper, we presented the design and implementation of a haptic virtual fine motor skill training system with the aim of engaging children with ASD in a virtual training environment to practice and eventually improve their fine motor skills. This system can provide adaptive training games with real-time feedback and record quantitative performance data. A small usability study was conducted with 6 adult TD participants and 4 children participants with ASD, who showed great interests in these games and improved performances in hand manipulations after playing these games. The study results suggested the usability and the potential of this system for fine motor skill training of children with ASD.

Acknowledgments. We would like to thank all the TD volunteers and the children/adolescents with ASD and their parents for their participation in the usability study. We are grateful for the support provided by NIH grant 1R01MH091102-01A1 for this work.

References

1. Baio, J.: Prevalence of Autism Spectrum Disorders: Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. Morbidity and Mortality Weekly Report. Surveillance Summaries, vol. 61, no. 3. Centers for Disease Control and Prevention (2012)
2. Zhao, H., et al.: A novel collaborative virtual reality game for children with ASD to foster social interaction. In: International Conference on Universal Access in Human-Computer Interaction. Springer (2016)
3. Provost, B., Lopez, B.R., Heimerl, S.: A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. J. Autism Dev. Disord. **37**(2), 321–328 (2007)
4. Cuffaro, M.A.: Fine motor skills. In: Encyclopedia of Child Behavior and Development, pp. 657–657. Springer (2011)
5. David, F.J., et al.: Coordination of precision grip in 2–6 years-old children with autism spectrum disorders compared to children developing typically and children with developmental disabilities (2012)
6. Crippa, A., et al.: Eye-hand coordination in children with high functioning autism and Asperger's disorder using a gap-overlap paradigm. J. Autism Dev. Disord. **43**(4), 841–850 (2013)
7. Forti, S., et al.: Motor planning and control in autism. A kinematic analysis of preschool children. Res. Autism Spectr. Disord. **5**(2), 834–842 (2011)
8. Jasmin, E., et al.: Sensori-motor and daily living skills of preschool children with autism spectrum disorders. J. Autism Dev. Disord. **39**(2), 231–241 (2009)
9. Piek, J.P., Baynam, G.B., Barrett, N.C.: The relationship between fine and gross motor ability, self-perceptions and self-worth in children and adolescents. Hum. Mov. Sci. **25**(1), 65–75 (2006)
10. Wade, J., et al.: Design of a virtual reality driving environment to assess performance of teenagers with ASD. In: International Conference on Universal Access in Human-Computer Interaction. Springer (2014)
11. Morris, D., et al.: Haptic feedback enhances force skill learning. In: Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC 2007). IEEE (2007)
12. O'Malley, M.K., et al.: Shared control in haptic systems for performance enhancement and training. J. Dyn. Syst. Meas. Control **128**(1), 75–85 (2006)
13. Srimathveeravalli, G., Thenkurussi, K.: Motor skill training assistance using haptic attributes. In: First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference. IEEE (2005)
14. Broeren, J., Rydmark, M., Sunnerhagen, K.S.: Virtual reality and haptics as a training device for movement rehabilitation after stroke: a single-case study. Arch. Phys. Med. Rehabilit. **85**(8), 1247–1250 (2004)

15. Kim, Y.-S., et al.: Haptics assisted training (HAT) system for children's handwriting. In: World Haptics Conference (WHC 2013). IEEE (2013)
16. Riederer, M., et al.: Development of tests to evaluate the sensory abilities of children with autism spectrum disorder using touch and force sensors. In: 2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth). IEEE (2014)
17. Morris, D., et al.: Haptic feedback enhances force skill learning. In: Second Joint EuroHaptics Conference, 2007 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. IEEE (2007)
18. The Geomagic Touch Haptic Device. http://www.geomagic.com/en/products/phantom-omni/overview
19. Poyade, M., Kargas, M., Portela, V.: Haptic Plug-In For Unity, 2014. Digital Design Studio (DDS). Glasgow School of Art, Glasgow (2014)
20. Unity3D. https://unity3d.com/
21. Hohendorff, B., et al.: Lengths, girths, and diameters of children's fingers from 3 to 10 years of age. Ann. Anat. Anat. Anz. **192**(3), 156–161 (2010)
22. Interlink Electronics. http://www.interlinkelectronics.com/FSR402.php
23. Arduino. http://store-usa.arduino.cc/products/a000066
24. Schwellnus, H., et al.: Writing forces associated with four pencil grasp patterns in grade 4 children. Am. J. Occup. Ther. **67**(2), 218–227 (2013)

Assistive Technologies and Natural User Interaction

Research on Driver's Visual and Psychological Characteristics Under Right-Turning Scenario for Car Head-Up Displays

Xi Wang[1(✉)], Yujia Wang[1], Xiaoqiu Zhang[2], and Chang Liu[2]

[1] School of Art and Media, Tongji University, Shanghai, China
yuer0487@outlook.com
[2] Pan Asia Technical Automotive Center, Shanghai, China

Abstract. Autonomous vehicles are increasingly becoming a reality, but for the driving vision information allocation is not clear enough. Driver distraction and inattention are the main causes of accidents. This paper presents experimental results of an eye tracker under real driving conditions, and analyses them from the design point of view. Clustering eye tracking data for each stage resulted in a pattern classification, as well as each type of driver's visual and psychological characteristics, from which common eye gaze practices could be summed up. The improved concept depends on the drivers' needs and desires and can serve as a basis for view management concepts of future HUD.

Keywords: Right-turning · Human-systems integration · Driver observation mode · Driver psychological · Head-up displays

1 Introduction

1.1 Background

Automotive safety is one of the main directions for future car development. The design and development of the intelligent vehicle is designed to assist or replace the driver to reduce the traffic accidents.

In the road traffic system, people, vehicle, environment is a mutual influence of the dynamic system, any unstable factors may lead to imbalance, and system is not stable, which leads to the traffic accidents caused by all sorts of contradictions. Moreover, many studies show that the driver factor is the first factor causing traffic accidents.

As the main factor of the human perception of the external environment, the research shows that driving occupies up to 90% of the visual channel depend- ing on the driving situation [1]. In the course of driving, the driver through the eye movement behavior to obtain the driving condition of the scene, in order to deal with visual information [2].

The deployment of visual attention as well as its response to changing conditions is coupled to our cognitive state. Eye movements allows us to track overt visual attention which is associated with a viewer's point of gaze [2].

In the course of driving, the driver through the eye movement behavior to obtain the driving condition of the scene, in order to deal with visual information.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_21

Driving is a task that requires a steady stream of visual input for the driver to adequately make decisions on vehicle positioning and evasive maneuvers. Even 2.0 s with a drivers' eyes away from the road can have devastating consequences as the drivers are not able to respond to the environments around them. While there are many resources invested in educating the public about safe driving and minimizing distractions (e.g. distracted.gov) [3].

There is little doubt that driver distraction and inattention are the main causes of accidents today. A recent study by Victor et al. [4] investigated the relation between visual glance behavior and crash for a large number of real rear-end collision events and came to the conclusion that crashes occur when the driver looks away from the forward roadway at the wrong moment (the driver has an "inopportune glance") [5].

Therefore, it appears that distracted driving caused by non-driving tasks will continue to be a concern in the future [3].

Recent research links time with eyes off of the road to increased chance of accidents, a problem that could be diminished when using HUDs [3].

The current understanding is that HUDs are preferable to in-built "head-down displays" (HDDs). Studies comparing HUD and HDD have concluded that HUD allows drivers to react faster to an abrupt event. In addition, it seems to support more consistent speed control and causes less mental stress. Nonetheless, another study suggest that the use of HUD could result in attention capture and increased reaction times to sudden events when compared to baseline performance. Also, the HUD benefits do not hold under the high workload of unexpected events, during which HUD users experience detriment to both the driving task and roadway event response [6].

HUDs keep the visual stimuli on a display at or above the driver's line of sight in a vehicle. This minimizes the time spent switching between looking at a roadway and at a center console or dashboard display, which are considered to be Head-Down Displays (HDDs). HUDs may allow drivers to use peripheral vision to maintain vehicle stability even when not focused on the road ahead. Because HUDs are not yet widely used, limited research has been performed to assess the validity of common standards in regards to HUDs [3].

The growing interest in eye tracking as a research tool in spatial cognition and related fields was the motivation for this Special Issue [5]. This paper puts forward a set of evaluation based on the vehicle vision eye movement test method. The method uses eye tracking system as a tool for collecting assessment information collection, evaluation of staff of region of interest in a particular context information, such as fixation time, fixation times, heat map, information trajectory maps, through the analysis of these qualitative information processing. We are most interested in extracting information about fixations and their durations, as these relatively steady eye movements indicate the regions of space subject to cognitive processing at the moment of fixation [2].

Redeploying visual attention to a new location is accomplished by saccades, the rapid movement of the eyes executed to reposition gaze. Gaze fixation data can indicate how viewers integrate information from text and diagrams [2].

In general, visual attention is governed by goal-driven intent, referred to as top-down attention. However, in the visual processing of a scene, there is interplay between top-down attentional shifts and bottom-up attention, drawn reflexively to spatial locations by external cues. Although computational models of top-down attention employ schemas, most are based on bottom-up models of vision, concerned with stimulus-driven, saliency-based attention [2].

The decision situations faced during both, locomotion and way finding, require the agent to collect information about the environment. Although humans can (and do) use several senses to acquire such knowledge [7], the visual sense is particularly interesting to study:

- Many way finding aids rely on visual information (e.g., maps, signage).
- Measuring visual attention is relatively easy, compared to measuring what the way finder perceives through other senses.
- The visual sense collects spatial information at a much greater distance and higher resolution than other senses.
- Humans actively direct their visual sense (in contrast to, e.g., the vestibular or auditory sense). With eye tracking, we therefore measure not only sensory input, but also the way finder's information acquisition procedure [2].

Combined with the driver's driving behavior record video, contrast analysis to the evaluation results.

Aiming at a deeper understanding of how these sub processes work, research in spatial cognition has used eye-tracking technology to study the visual search processes during way finding.

1.2 The Main Contribution of the Eye Tracking Study

Looking for the Breakthrough Point for the Development of Intelligent Vehicles in the Future. The driver's eye movement analysis can provide valuable information for the intelligent vehicle system, the intelligent vehicle system, capable of predicting or judging that in certain circumstances, a behavior of driver's intention and purpose, and timely according to the prediction results from the feedback or judgment, improve the interaction between cars and people, improve vehicle safety.

To Search for the Evaluation and Design Method of Vehicle Vision Based on Eye Tracker. As an experimental tool, eye tracker is widely used in vehicle testing, but its data and research results are used in the field of transportation, and the application in the field of evaluation and design of vehicle vision is less. In this paper, the eye movement experiment, data processing and application from the design point of view, hoping to provide help for the application of eye tracker in the field of automotive design.

2 Eye Movement Study

In the process of using the eye tracking experiment in different traffic conditions, the driver's personal habits, the eye tracking data is not properly chosen and other factors are likely to have an adverse impact on the experimental results, so we in the formal experiment before the pre-experiment, finally through the experiment time limit to ensure road traffic conditions are similar, the use of action camera video recording to facilitate the processing of data to understand the driver's behavior of inviting eye tracker professional engineers how to interpret the data acquisition and explain the meaning of the data. We finally identified the experimental content.

We have done nine daily scenario, in this paper we choose right scenario for analysis. Because of China is left driving, the driver turned his head wide when turn right, the scenario is relatively complex.

2.1 Evaluation Process

Evaluation Design. The experiment used the Tobii Glasses 2 to collect the fixation points.

Eighteen (9 female, 9 male) subjects took part in our evaluation. They had between 1 to 13 years (M = 7.45, SD = 4.23) of driving experience. Road conditions are city roads.

Analysis of Eye Tracker Data. The following is a description of the map area of interest: SPOT a is left front windshield; SPOT b is right rearview mirror; SPOT c is interior rearview mirror; SPOT d is Left outer rear view mirror; SPOT e is the left window; SPOT f is right window; SPOT g is console (Instrument panel area); SPOT h is right front windshield (Fig. 1).

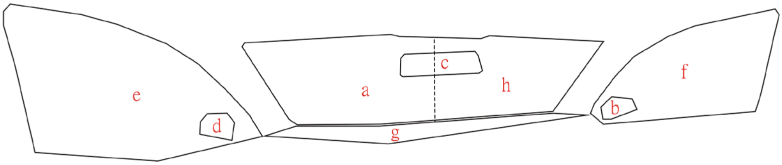


Fig. 1. AOI (*area of interest*).

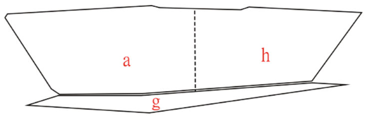
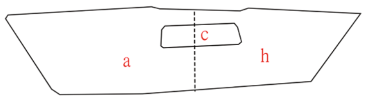
3 Analysis

3.1 The Visual Features and Psychological Characteristics of the Right Turning

4 Results and Discussion

The area derived from Tables 1, 2 and 3 according to the driver's observation mode is instructive.

Table 1. The decision-making stage is defined as the start of the received command and the tendency to execute.

Classification	Driver observation mode	Influence factor	Driver intention	Inspiration for design
1	Only in front of the field of vision (a/h/g), mostly for the state of the eye, the current side of the car/person, showing a short time smooth tracking state 	Car environment is relatively simple	The road is simple; the driver is more relaxed, look at the situation only to confirm the car environment	In this case, the vehicle self-monitoring will play a more important role, at the same time, due to the driver's eye movement behavior to watch; in order to let the driver noticed not blocking the line of sight, so remind information can be placed in the h zone, or to voice
2	In the three regions of the a/c/h, it is shown that the triangle is swept through the rear view mirror) 	Vehicles before and after the party line more people need to pay attention when driving vehicles, driving habits	Vehicles before and after the road conditions, due to the slow speed of pedestrians, so the driver through the rearview mirror to confirm the rear traffic conditions	At this point, the speed and the distance between the driver will be more attention to the information, the rearview mirror as the driver frequently scan area, you can display images and short text messages

(*continued*)

Table 1. (*continued*)

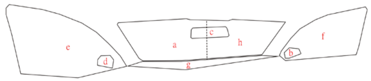
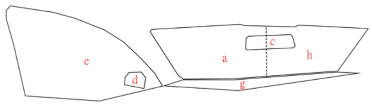
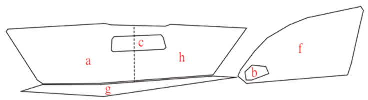
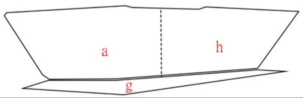
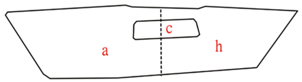
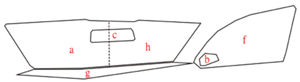
Classification	Driver observation mode	Influence factor	Driver intention	Inspiration for design
3	The driver line of sight in the left right between the three regions are substantially horizontal scanning, the driver's head rotation Angle is bigger also, scanning involves most of the area 	Speed is faster in all subjects, about 20 km/h	Cautious driver, quickly identify the car environment, and need to get visual information from as many areas as possible	When the vehicle speed is fast, the driver did not have enough time to get information, and because of the fast changing scene, the ability to identify the drivers of objects will be reduced. Due to the short time, can provide direct information on the threat of warning, do not provide extra information scattered the driver attention, try to simplify the information process of human identification
4	Watch the front and the left side view 	There are green belts in the right side of the bend, in front of a vehicle or pedestrian over	The driver is in front of the car and pedestrians	When there is an obstruction (such as a bush) around the vehicle, the driver and the vehicle are reminded
5	Watch the front view and the right side view 	Front without car	The driver is more cautious, from the right area to confirm the corner of the environment and the rear of the vehicle	According to the observation of the majority of the driver's right turn, the right rear vehicle distance and the right front situation through the image display in the h area

Table 2. The implementation phase of the body began to turn right, most of the subjects at this time turn the steering wheel.

Classification	Driver observation mode	Influence factor	Driver intention	Inspiration for design
1	Only limited to the front view (a/h/g), mostly gaze state, when the front of the car/person, showing a short time smooth tracking state 	There are no other vehicles and pedestrians on the road ahead	Simple road conditions, the driver is more relaxed, watching the case only to confirm the car around the week light environment	In this case, the vehicle self-monitoring will play a more important role, at the same time, due to reminders
2	In a/c/h three areas showing a triangular scan state 	The rear of the vehicle passing	Confirm the driver behind the traffic through the rearview mirror	It is possible to give an image of the rear traffic condition directly to the rearview mirror and inform the driver and the vehicle distance of the rear vehicle
3	Watch the front view and the right side view 	Fewer pedestrians, and there is no phenomenon from the right through, so guess the relationship with the driver's habits	Turn right when the driver is accustomed to seeing the right side of the window to confirm the upcoming lane, and through the right rearview mirror to confirm the rear vehicle. At this point the driver's observation mode is mainly to eliminate the anxiety of the heart, rather than get effective information to help turn right	We can give the driver enough information in the upper right corner of the area h in this case, to help the driver to enhance security

(*continued*)

Table 2. (*continued*)

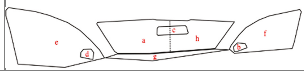
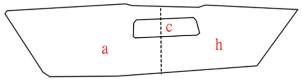
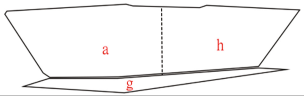
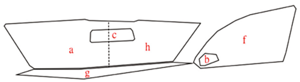
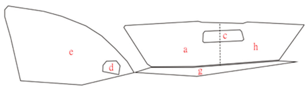
Classification	Driver observation mode	Influence factor	Driver intention	Inspiration for design
4	The driver line of sight in the left right between the three regions are substantially horizontal scanning, the driver's head rotation Angle is bigger also, scanning involves most of the area 	Driver habits: one is in the decision-making stage and adjust the stage to focus on the main, so in the implementation phase will be frequent scanning. One is in the vehicle during the process of a full scan as the main observation mode	Cautious driver, quick confirmation of car cycle environment, and need to get as much as possible from the area of visual information	Help the driver to filter the threat of information to remind the sound, you can use real-world technology directly to the information superimposed on the real object, so that the driver in the scanning process to read

Table 3. The adjustment stage refers to the basic completion of the right task; the driver did not adjust the state of the next phase of the road to prepare.

Classification	Driver observation mode	Influence factor	Driver intention	Inspiration for design
1	Only limited to the front view (a/h/c), but through the rearview mirror to observe the rear of the vehicle situation 	Vehicles on the road less pedestrians at the same time, through the first two stages of observation to understand the observation mode of the driver to focus on the main, small range of scanning	Into the straight state, within the rearview mirror to confirm the rear traffic conditions	Nothing
2	In a/g/h three regions showing a triangular scan state, mostly gaze state 	Due to the straight road, both sides of the pedestrian vehicles less, almost no threat to the driver of the factors	The driver enters the straight state	Nothing
3	Watch the front view and the right side view 	The right side is a large piece of lawn, so the external environment does not exist influencing factors	One type is the first two stages are not observed on the right side, so inertia adjustment stage to see the right; Another type is the right process from the right side are used to obtain visual information	For the more practical drivers, in order to make the private car more personalized, applicable to the user, you can use the user-defined model, in the case of options, the driver can choose according to their own driving habits of information location area
4	Watch the front and the left side view 	There is a car passing in the rear		In the rear of the vehicle beyond the safe distance, you can remind the left rearview mirror

We can conclude that, in the turn right scenario of this experiment, in the decision-making stage, the driver is more comprehensive observation, mainly because the traffic situation is slightly complicated, the driver needs to quickly confirm the car environment for the right after the right In this case I think the vehicle to provide information should try to avoid the text information, prevent the driver's mind thinking about it and inattention, and should try to give the warning information, such as security for their own vehicles are threatening information; in the In the execution phase, the driver observed mode to watch the front windshield, At this time the vehicle itself should help the driver to detect the vehicle around, and promptly reminded, this time the reminder information should be placed mainly in the h area, so that the driver watching the front while the remaining light can notice the information; adjustment stage, the driver observed mode close to the previous stage and straight task, the attention should be consistent with the previous stage.

In addition, in the case of multiple areas will see, HUD need multi-screen collaboration. Since the a/h area is the gaze area of all the drivers and the first gaze area, it is relatively important in multi-screen collaboration.

5 Limitations and Future Work

Due to the lack of time and ability, the above information is only based on the right-hand scenario of the driver observation mode, it is not enough to apply to all road conditions; Similarly, for this reason, we have no way to use only one scene to the driver Classification, so we classified the driver's observation pattern according to the three stages.

In the future study, we will complete the left, straight and change the analysis to compensate for the above mentioned deficiencies.

References

1. Haeuslschmid, R., Shou, Y., O'Donovan, J.: First steps towards a view management concept for large-sized head-up displays with continuous depth. In: 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM Digital Library, New York (2016)
2. Kiefer, P., Giannopoulos, I., Raubal, M., Duchowski, A.: Eye tracking for spatial research: cognition, computation, challenges. Spat. Cognit. Comput. J. **17**(1–2), 1–19 (2017)
3. Smith, M., Gabbard, J.L., Conley, C.: Head-up vs. head-down displays: examining traditional methods of display assessment while driving. In: 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM Digital Library, New York (2016)
4. Victor, T., BärgMan, J., Boda, C.N., et al.: Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk. Transportation Research Board, Washington, DC (2014)
5. Larsson, P., Niemand, M.: Using sound to reduce visual distraction from in-vehicle human-machine interfaces. J. Traffic Inj. Prev. **16**(sup1), S25–S30 (2015)

6. Villalobos-ZúÑIga, G., Kujala, T., Oulasvirta, A.: T9 + HUD: physical keypad and HUD can improve driving performance while typing and driving. In: 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM Digital Library, New York (2016)
7. Loomis, J.M., Klatzky, R.L., Giudice, N.A.: Representing 3D space in working memory: spatial images from vision, hearing, touch, and language. J. Multisen. Imag. 131–155 (2013)

Take-Over Requests Analysis in Conditional Automated Driving and Driver Visual Research Under Encountering Road Hazard of Highway

Fang You[1], Yujia Wang[1], Jianmin Wang[1(✉)], Xichan Zhu[2], and Preben Hansen[3]

[1] School of Arts and Media, Tongji University, Shanghai, China
365120435@qq.com, yujiawang49@gmail.com, wangjianmin@tongji.edu.cn
[2] School of Automotive Studies, Tongji University, Shanghai, China
zhuxichan@tongji.edu.cn
[3] Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden
preben@dsv.su.se

Abstract. In conditional automated driving, vehicles monitor the driving environment. Simultaneously, driver can also attend to a secondary task, but also need regain driving control when vehicle requests to intervene. The collaboration of vehicle and human driver support the driving experience in this situation. From automated driving to manual driving while in highway scenario, many researchers focus on secondary task engagement, take-over time and requests. This paper evaluates human performance while regaining driving control in conditional automated driving, research investigates take-over requests under highway hazard scenario through visual scanning analysis in lane changing situation. Different obstacles cause driver visual attention mode changes and adaptations. Results show that all participants can take-over in 6 s for voice chat tasks, while in electronic reading condition, not all participants complete can take-over even in 8 s.

Keywords: Take-over requests · Driver visual attention · Highway hazard scenario

1 Introduction

With the continuous development of surroundings detection and control systems, self-driving cars have developed from concept to a reality gradually. Driverless cars are expected to enter the daily life of ordinary people in 2025 [1]. At the same time, relative traffic safety problem has become more and more serious. In manual driving mode, driving safety is a serious challenge [2] in the worldwide and traffic crash is also a major world public health problem. Corresponding automotive security, human driver in-vehicle behavior and role definition, car intelligence degree has attracted the attention of researchers.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_22

SAE International's On-Road Automated Vehicle Standards Committee has defined six levels of driving automation [3]. According to these six levels, we can discuss the different levels of automation driving. In the level 0 (No Automation), the human driver need to undertake the full-time driving performance. Nowadays, almost no car resides at level 0. To improve the traffic safety, an increasing number of vehicles are currently equipped with advanced driver assistance systems (ADAS) to improve the autonomy of a vehicle (Level 1). At Level 2, we find Partial Automation status and as an example, we have the Tesla car. While having the degrees of car accidents shows that its driving automation system is not perfect and safe enough. Level 3, which involves conditional automation and CAD, the car has the possibility to monitor the driving environment (context), support guidance and control tasks that are conducted by the system within the car. At Level 4 (High Automation), if human drivers do not respond appropriately to a request to intervene, the driving mode-specific performance could be driven by the automated car. One example of cars on this level are Google autonomous vehicle. At the final level (Level 5), a higher level of driving automation is still in development.

Self-driving vehicle can be implemented on the highway with Full Speed Range Adaptive cruise control (FSRA) which means automatic constant speed control and adjustment of the vehicle-to-vehicle distance at full speed range and with Lane Keeping Assist (LKA) automatic control of the steering to run within the lane. In 2016, Tesla had several traffic accidents. In these cases, Tesla's autopilot mode activating, the limited capacity of sensing the environment and driving context resulted in driver casualties. Further, these kind of phenomenon inspired researchers to cooperate on issues related human driver and vehicle in partial and conditional automation.

2 Background

The summary of levels of driving automation for on-road vehicles, described level 3 as "*the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene*" [3]. It does not define the situation or time in detail. Therefore, the major problem of conditional automotive, is to define when manual driving takes control over automotive driving.

The selections and handling of secondary task engagement, along with the needed reaction times to comply with a Take-Over-Requests (TOR) is one of the main research questions. this also affects the design of TOR with respect to the human machine interface (HMI) [4]. In similar studies of take-over requests, the issue starts in a highway scenario and then researched actively in various other situations [5].

From safety priority aspect, the early warning system should alarm at an early point, but the early warning system will cause the driver to think that no alarm or false alarm is required, thus reducing the effectiveness of the early warning system [6]. This means that the problem of take-over request time's upper limit is affected by several complex factors. Furthermore, accompanied by advances in technology, the upper limit of time required is changing. Some scholars do research focus on take-over lower limit time. That is to say, human drivers at least need how much time to take over from automated driving. At level 3 conditional automated driving, there are some studies of the time required for manual driving to take over from automated driving. Through an experimental setting utilizing different driving conditions in an immersive driving simulator, we studied and tested take-over completed time. We studied the time used for the transition from automated driving to manual driving, before a simulated road hazard. These experiments included three parts. First, participants were given a practical exercise task (5–10 min) to get accustomed to the test setting and the simulated driving environment. Secondly, the participants performed the main experimental task involving the test car simulator in an automated driving mode. Finally, the participants was involved in a set of take-over required tasks in which the researchers implemented different time variables to be tested.

Damböck and Toshio Ito [5] set the automotive driving speed to 100 km/h, and the take-over request time was 4 s, 6 s, 8 s and 5 s, 7 s respectively [5, 7]. In the first study, 6 s conditions was used, and most of the drivers could take over the driving rights and complete all three scenarios under the driving task. In the 8 s conditions case, all the drivers completed the task, and the driving level and the driver is always the same driving process performance. In the other study, when using the time conditions of 5 s, 7 s, all six combinations (such as leave the steering wheel or not, eye on monitoring the driving environment or not) of the drivers' state are able to successfully take over the driving rights and to avoid rear-end accidents. In the analysis of this limited study, it seems that the driver's safe completion of the driving right takes 6 s to 8 s, and 8 s to be a good timeframe for an acceptable transition in a highway scenario.

van den Beukel et al. [8] simulated different jammed traffic situation by making the front car emergency braking. At the time conditions 1.5 s, 2.2 s and 2.8 s, the accidents percentage was 47.5%, 20.8% and 12.5% respectively Mok et al. [9] set 2, 5 or 8 s before collision obstacles Just a few drivers within the 2 s condition were able to safely navigate the road hazard situation, while the majority of drivers in the 5 or 8 s conditions were able to navigate the hazard safely.

Another important aspect to consider is the Time to Collision (TTC) that can reflect the relative speed and relative distance between the vehicle and the obstacle in front of the car, and is widely used in the manual driving analysis of vehicle longitudinal collision safety.

Driver distractions will involve many security risks when the car is in automation mode, and presently they rely on the system's capabilities of managing the distraction [10]. In take-over request studies, the use of secondary task engagement will be common activities in a future driving situation of an autonomous vehicle.

In a study by Borojeni [11], the participants were distracted in the vehicle while autonomous driving through video watching, perceived workload ratings on a screen in the center console was measured. In another study, the task load as a commonly-used ergonomics factor for measuring working quantity was adopted to the experiment design, like NASA Raw Task Load Index [12]. Within take-over required time and secondary task engagement, multiple test trails are designed to enrich and deepen highway human-vehicle cooperation research.

In addition to the problem of take-over request elements, from in-vehicle system giving drivers take-over requests to drivers pass the obstacles safely, the whole driving process need to be considered. We tried to figure out how drivers acquire environment information in lane changing and different driving performance under different secondary task engagement. Through these two aspects studies help to improve driving take-over request safely. This paper involves two interconnected studies: one is monitor and observe the drivers' visual scan while driving to investigate a drivers' needs when performing a lane-changing driving scenario. In the other connected study, an experiment will test the success rate in a take-over requests scenario during two different conditions: an electronic reading state and in a voice chat state.

3 Scenario Analysis

3.1 Scenario Analysis

Automation technology is still not mature enough, and today the vehicle cannot entirely perform the driving task itself without human input. Drivers always need to keep the environment under observation, and ready to take over if the driving situation requires that. Based on in-vehicle interaction system, we tried to improve the safety of driving through the collaboration of people and vehicles. When the autopilot is not able to complete the driving task, the driver need to take over the driving control. For our analysis, we selected and designed a highway scenario that included that the driver encountered obstacles on the road (Fig. 1).

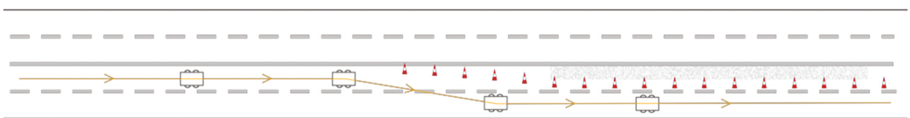


Fig. 1. Highway hazard scenario schematic diagram.

In the highway crash scene, the driver scenario is characterized as follows: when the driver observes the obstacles, the driver starts to slow down, sweep the target lane direction on the external mirror to confirm that the lane is safe and secure. The driver turns the steering wheel and activate the lights to change the course of the way, and finally, move the vehicle through the first traffic cone, and adjust the vehicle to complete the lane [13].

3.2 Lane Changing Visual Scanning Analysis

A previous and related research study shows that about 90% of the information acquired during the driving process comes from human visual system [14]. After understanding the encountering road hazard highway scenario, we started to analyze the driver's visual characteristics in the person-car-environment frame. Through recording and gathering the driver's eye movement data, we summarized the driver's visual scanning mode during the lane changing course. To goal with this was to extract information about the drivers' needs for in-vehicle HMI design.

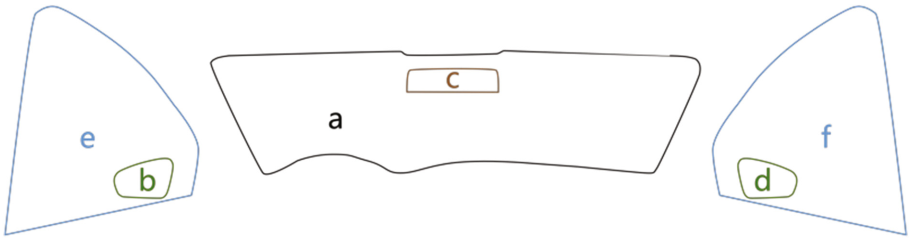


Fig. 2. AOI (interest of area). (a) Front view; (b) left outer rearview mirror; (c) rearview mirror; and (d) right outer rearview mirror; (e) left side window; (f) right side window.

We recruited 18 participants with the following requirement: a valid driving license and already having a private car. Gender distribution was as follows: males (N = 9) and females (n = 9). The participants reported that they owned their driving licenses from 1 to 13 years (M = 7.45, SD = 4.23). The 18 participants had a variety of background, such as teachers, security guards, office staff, and so on. Our experiment utilized an eye movement device called Tobii Pro Glasses 2, and installed the software Gopro HERA4[1] in the co-pilot in order to record the driver's head and hand movements. We recorded both the internal and external environment when the driver performed the task by using the camera inside the driver's eye movement instrument. At the end, the subjects were invited to score degree of AOI.

As for the data collection of the eye movements, we measured and analyzed the eye fixation points. Table 1 the eye movement area is divided into 6 AOI interest Area, such as Fig. 2.

Table 1. Right to left lane changing eye-tracking data.

AOI area	Total visit count	Percentage fixated (%)
a	521	100.00
b	226	100.00
c	96	38.89
d	57	38.89
e	46	27.78
f	41	33.33

The classification of the lane is divided by Worrall's stages of the lane change process, and the analysis of lane changing decision-making stages [15]. The preparation phase of the lane change is subjected to two elements, which are time and distance. Where time is the duration the vehicle needs to approach the target lane, and distance is the location of vehicle from the demarcation zone of the two lanes.

The execution phase of the lane changing activity starts when the driver decides to turn to the right, and normally, this is when most of the subjects start to turn the

[1] http://www.tobiipro.com/ and https://gopro.com/.

steering wheel. In adjustment stage, the lane changing task finished completely and the vehicle returns to the normal running state on the target lane.

In lane changing situations (see Fig. 2), the front view and left rear view mirror get the highest attention. The driver's observation, and participants vision scanning divided into three categories according if the front view and left outer rear was used or not: (a) observing the front view and left outer rear (three variations), (b) observing was more than front view and left outer rear; and (c) observation was less than front view and left outer rear, like only observation front view (Fig. 3).

Decision-Making Phase Vision Scanning Analysis.
In the decision-making phase, when the external environment is relatively simple, participants looked the front view area and left outer rear area to confirm the opportunity to changing lane. Vision scanning between front view and left outer rear.

When the external environment is more complex, and the obstacles closer to the vehicle, the driver may be in a "tense mode" and need to be alert in order to confirm the safe passing in multiple areas. Participants are not only concerned with the front view, left outer rear, but also the attention to other areas, like right outer rear and rearview mirror.

When the external environment is more complex and obstacles get to be far from the vehicle, the driver observe the road far ahead, watching road condition changing. In this situation the participants only fixed the front view.

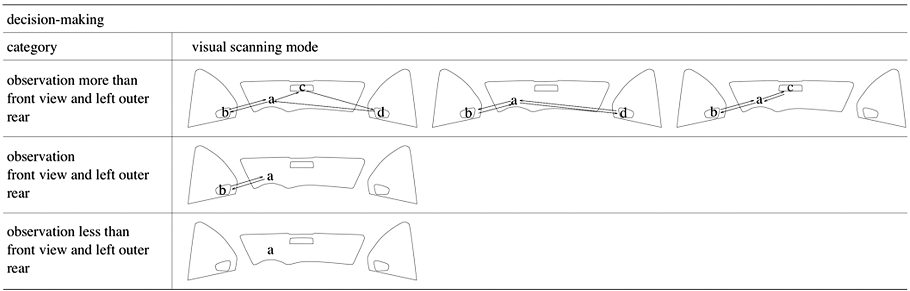


Fig. 3. Visual scanning mode summary of the decision-making phase. (a) Front view; (b) left outer rearview mirror; (c) rearview mirror; and (d) right outer rearview mirror.

Execution Phase Vision Scanning Analysis.
When original lane appears closer to the interference factor, it will cause participants to be vigilant and watch the right rearview mirror more often. When a vehicle, pedestrian or non-motor vehicle, appears in close distance and in front of the road of the experimental car, it results in persistent observation and then the participants only fixed the front view. Most of the participants in the execution phase, observed the left rearview mirror several repeated time to look for cars behind. Overall, in the lane changing

execution phase, the road is relatively calm (no or few quantities of vehicles). Furthermore, the environment outside the car did not contain any particular events that may attract the drivers' attention. This result in participants only scanning the front view and the left side of the rearview mirror two areas, to confirm pathway and driving safety (Fig. 4).

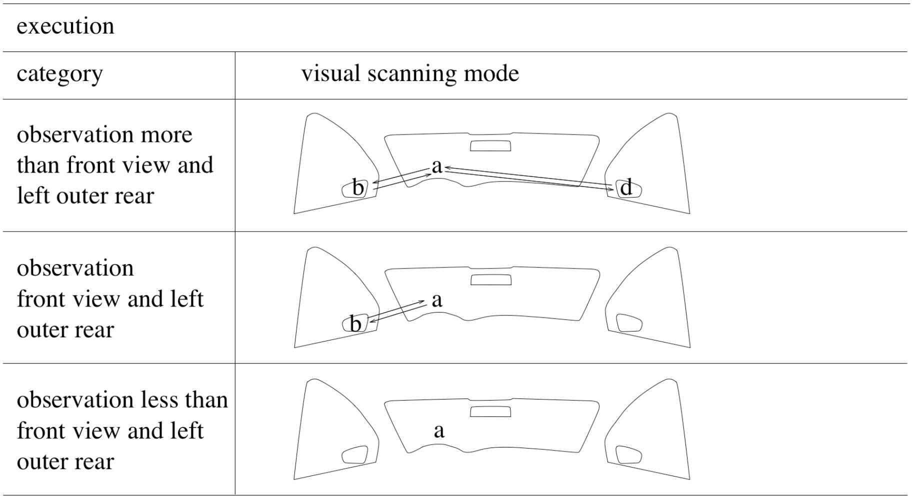


Fig. 4. Visual scanning mode summary of the execution phase. (a) Front view; (b) left outer rearview mirror; (c) rearview mirror; (d) right outer rearview mirror.

Adjustment Phase Vision Scanning Analysis.
This case involves more complicated traffic conditions and the interference factor around and close to the experimental vehicle caused participants to scan multiple areas to confirm driving safety. This groups of participants' driving speed significantly high or low than another two kinds of participants (Fig. 5).

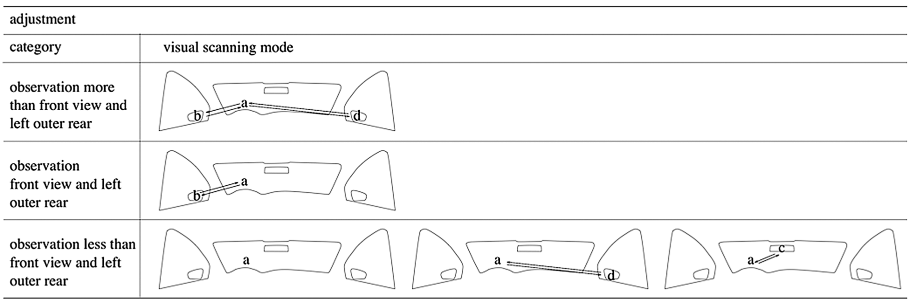


Fig. 5. Visual scanning mode summary of the adjustment phase. (a) Front view; (b) left outer rearview mirror; (c) rearview mirror; (d) right outer rearview mirror.

When, in the road ahead, interference factors like a vehicle, pedestrian or non-motor vehicle appeared close to the experimental car, the test participants only fixed on the front view. After changing line, a vehicle appearing on the original road caused participants to make a security observation to the environment security by scanning right rear mirror and inside rearview mirror.

4 Application and Evaluation

After the data analysis of the three different scenarios of phases in lane changing as described above, we found that distant obstacles will attract the participants' attention and generate interferes that will cause problem for the driver to complete the driving task. In the highway encountering hazard scenario, they will obverse and bypass the obstacles. When the driver cannot monitor the surrounding driving environment properly, the occurrence of traffic accidents may therefore increase.

When applying our finding on a take-over request problem, we suggest that in order to overcome the problem, the obstacle then need to be marked in advance. If drivers could understand lane condition faster, it will be helpful in a take-over safety situation.

This will draw the attention to the driver and result in a more unproblematic traffic condition. Head-up display has already been proved to improve information access and reading speeds compared to classic head-down displays [4]. Combining lane changing processes and highway hazard scenario analysis, it is necessary to mark the lane line which contain obstacles, in order to guide drivers to a quicker and more clear driving. We designed the heads up display in windscreen for the highway hazard scenario, including lane safety states, traffic cone azimuth, and sound for notice regain driving control (Fig. 6).

Fig. 6. The HMI prototype. The participant is shown in two situations: in reading condition (middle) and take-over requests to avoid obstacles in lane changing (right).

4.1 Evaluation Procedure

The simulated driving course contains three different sections and two alternative control transitions. In the first one, participants introduced to the virtual simulation driving environment and hardware facilities (5 min). Next, in the second experimental task, they switched to an automotive mode, which involved the car set to 80 km/h speed. The task duration was 10 min. In the third experimental task, the participants were in a situation where they and the vehicle within three time intervals (4, 6 or 8 s), encountered a takeover request, where the participants need to regain driving control.

Distraction events were based on common driving distraction in Chinese traffic situation. When the vehicle is in automotive mode, the participants was asked to perform different types of non-driving tasks, so the participants were in a distractive state. Participants' distraction task were introduced in two different conditions: (a) an electronic (mobile) reading condition in which the participants was free to browse in a SNS application, and (b) a voice chat condition, in which participants used a phone app to engage in a voice chat. Researchers studied six participants in a pilot study, mainly University teachers and students, 3 females, 3 males. The participants were between 19 and 28 years old with driving experience of 1–12 years.

4.2 Result

The result of the driver performances from the two distraction conditions of the electronic reading state and the voice chat is shown in Table 2 below. Regarding both the electronic reading condition and the voice chat condition rates, the driving performances are the same at take-over time in both the 6 s and 8 s conditions. However, in 4 s condition, participants in the voice chat state take-over requests success rate is better than in electronic reading state take-over requests success rate, respectively in 33.33% and 66.66%.

Table 2. Take-over requests success rate under electronic reading and voice chat condition.

Time required(s)	Electronic reading condition take-over requests success rate (%)	Voice chat condition rate take-over requests success rate (%)
4	33.33	66.66
6	83.33	100.00
8	83.33	100.00

Furthermore, in electronic reading condition, participants were in relatively strong secondary task engagement (driver take-over). When they needed to regain driving control, they wound put down the phone, and then put hands in wheel to take-over. In the voice chat condition, some participants were in different body posture: one hand on the steering wheel, the other hand with the mobile phone, which made it easier to regain manual driving. During the experiment, when participants received the take-over request, the behavior of the participants to manage the operation was ordered in the following categories: (a) the phone placed in the legs; (b) hands to take the steering wheel; (c) while others took the phone in their hand, and (d) while yet others were holding the phone and steering wheel.

5 Discussion and Conclusion

This paper investigated the visual driver behavior within the area of complex human-vehicle-environment system, using different environmental interference factors and shows that these results in different driving behavior is regarded with varied drivers' attention and decision-making processes.

In particular, when the interference factors, such as vehicle, pedestrian or non-motor vehicle, are far from the experiment car and with few obstacles on the road, the driver will manage the interference factor through the front view. When the interference factors (vehicle, pedestrian or non-motor vehicle) are near the experimental car as well as with a large amount of obstacles of various types, the driver will be in saccade between multiple areas.

In our take-over request experiment, despite the sampling for the study is low, it indicates different performances behavior within the secondary task engagement. During electronic reading condition, not all participants managed finish regaining driving control, not even within the 8 s interval. For the electronic reading condition, all participants completed the state transition from manual driving to automated driving in 6 s. Finally, different body gestures encountered and discovered which affected the take-over request success rate. In future research, more subjects is needed and additional secondary task engagements needs to be tested in order to discover and explain more effective take-over request procedures.

References

1. Self-driving cars moving into the industry's driver's seat. http://news.ihsmarkit.com/press-release/automotive/self-driving-cars-moving-industrys-drivers-seat
2. World Health Organization: Global Status Report on Road Safety 2015. WHO Press, Geneva (2015)
3. SAE levels of driving automation. http://www.sae.org/servlets/pressRoom? OBJECT_TYPE = PressReleases&PAGE = showRelease&RELEASE_ID = 3544
4. Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., Bauer, W.: Take-over requests for automated driving. In: 6th International Conference on Applied Human Factors and Ergonomics. Elsevier B.V., Las Vegas (2015)
5. Ito, T., Takata, A., Oosawa, K.: Time required for take-over from automated to manual driving. In: SAE 2016 World Congress and Exhibition. SAE International, New York (2016)
6. Singh Tomar, R., Verma, S.: Safety of lane change maneuver through a priori prediction of trajectory using neural networks. J. Netw. Protoc. Algorithms **4**(1), 4–21 (2012)
7. Damböck, D., Bengler, K., Farid, M., Tönert, L.: Übernahmezeiten beim hochautomatisierten Fahren. Tagung Fahrerassistenz, München (2012)
8. van den Beukel, A.P., van der Voort, M.C.: The influence of time-criticality on situation awareness when retrieving human control after automated driving. In: 16th International IEEE Annual Conference on Intelligent Transportation Systems, Hague (2013)
9. Mok, B., Johns, M., Lee, K.J., Miller, D., Sirkin, D., Ive, P., Ju, W.: Emergency, automation off: unstructured transition timing for distracted drivers of automated vehicles. In: IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, New York (2015)

10. Walch, M., Sieber, T., Hock, P., Baumann, M., Weber, M.: Towards cooperative driving: involving the driver in an autonomous vehicle's decision making. In: 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, New York (2016)
11. Borojeni1, S.S., Chuang, L., Heuten, W., Boll, S.: Assisting drivers with ambient take-over requests in highly automated driving. In: 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, New York (2016)
12. Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., Bauer, W.: Take-over requests for automated driving. In: 6th International Conference on Applied Human Factors and Ergonomics, Elsevier B.V., Amsterdam (Las Vegas)
13. Wang, X., Yan, L.I.: Characteristics analysis of lane changing behavior based on the naturalistic driving data. J. Traffic Inf. Saf. **1**, 17–22 (2016)
14. Summala, H., Lamble, D., Laakso, M.: Driving experience and perception of the lead car's breaking when looking at in-car targets. J. Accid. Anal. Prev. **30**(4), 401–407 (1998)
15. Worrall, R.D., Bullen, A.G.R.: An empirical analysis of lane changing on multilane highways. In: Highway Research Record 303, pp. 30–43. National Research Council, Washington, DC (1970)

Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces

Gustavo López(✉), Luis Quesada, and Luis A. Guerrero

University of Costa Rica, San José, Costa Rica
{gustavo.lopez_h, luis.quesada}@ucr.ac.cr,
luis.guerrero@ecci.ucr.ac.cr

Abstract. Natural User Interfaces (NUI) are supposed to be used by humans in a very logic way. However, the run to deploy Speech-based NUIs by the industry has had a large impact on the naturality of such interfaces. This paper presents a usability test of the most prestigious and internationally used Speech-based NUI (i.e., Alexa, Siri, Cortana and Google's). A comparison of the services that each one provides was also performed considering: access to music services, agenda, news, weather, To-Do lists and maps or directions, among others. The test was design by two Human Computer Interaction experts and executed by eight persons. Results show that even though there are many services available, there is a lot to do to improve the usability of these systems. Specially focused on separating the traditional use of computers (based on applications that require parameters to function) and to get closer to real NUIs.

Keywords: Human factors · Intelligent personal assistant · Smart device

1 Introduction

Alexa, Siri, Cortana and Google Assistant are the most common and extensively used Intelligent Personal Assistants available in the market. From the academic perspective, they are Speech-based Natural User Interfaces (NUI).

In the Human-Computer Interaction (HCI) domain, speech-based NUIs are systems that users operate through intuitive actions related to natural human behavior via voice instructions. There are many names for speech-based NUIs. However, this paper focuses on voice-activated intelligent personal assistants deployed in smartphones or smart speakers.

Nowadays, most smart-device manufacturers have their own voice assistant application. However, the designs vary significantly and their functionalities are very different [1].

The hypothesis of this paper is that the run of manufacturers to release new versions of their voice-activated personal assistants has had a large impact on their products, especially in their simplicity and usability.

An example of this assertion is that when voice-activated personal assistants were first released their behavior was limited and the commands were very rigid and

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_23

structured [1]. Afterwards, several efforts were conducted to create more natural interfaces [2–5]. However, some of the current devices went back to rigid and structured commands.

This paper presents a functional and usability test of some of the most prestigious voice-activated personal assistants in the market (i.e., Amazon Alexa, Apple Siri, Microsoft Cortana and Google Assistant).

Each personal assistant answered the same request and a group of eight people rated each response on two categories: natural feeling and correctness. Results show that Google Assistant is the most natural personal assistant. However, it is also the less correct. On the other hand, Siri is the most correct personal assistant and the least natural.

The rest of this paper is arranged as follows: Sect. 2 illustrates the methods used in this work. In Sect. 3 the evaluation process is explained. Section 4 presents the results, which are further discussed in Sect. 5. Finally, Sect. 6 is the conclusion.

2 Evaluated Systems

In this paper, we describe an evaluation of four intelligent personal assistants. This section introduces the four evaluated systems: Amazon Alexa, Apple Siri, Microsoft Cortana and Google Assistant. These are the most popular assistants.

2.1 Amazon Alexa

Alexa is an intelligent personal assistant developed by Amazon. Alexa is linked with Amazon's smart speaker "Echo". According to Amazon, Echo is a hands-free, voice-controlled device that plays music, controls smart home devices, provides information, reads the news, and sets alarms, among other functionalities [6].

Moreover, Alexa is capable of voice interaction, music playback, making to do lists, setting alarms, streaming podcasts, playing audiobooks, and providing weather, traffic, and other real time information. Alexa can also control several smart devices using itself as a home automation hub [7].

Some of the promotional commands for Alexa include: "Alexa, find me a Chinese restaurant.", "Alexa, re-order paper towels". "Alexa, what's on my calendar today?", "Alexa, set a timer for 20 min", "Alexa, play Adele from Prime music", and "Alexa, what's my commute?" [6].

Additionally, Alexa services can be invoked from compatible devices developed by Amazon (i.e. Kindle Fire, Amazon Fire Phone and Fire TV) [7].

2.2 Google Assistant

Google Assistant is an intelligent personal assistant developed by Google. It was designed to allow conversational usage [8]. Google Assistant's predecessor is Google Now. Google Assistant is linked with Google Home, a voice-activated speaker [9].

Google Now uses a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of services. Moreover, it delivers information to users predicting their requirements.

Some of the promotional commands for Google Assistant include: "Ok Google, Remind me to pick up a birthday card", "Ok Google, Book me a table for 6 at Quartino for 8:30", "Ok Google, Who invented sushi?" [8].

2.3 Microsoft Cortana

Cortana is an intelligent personal assistant created by Microsoft for Windows powered devices (both PC and mobile). Cortana can set reminders, recognize natural voice without the requirement for keyboard input, and answer [10].

Some of the promotional commands for Cortana include: "What's the weather like?", "Call Sarah", "What is five miles in kilometers", "What's the definition of 'table'?" [10].

2.4 Apple Siri

Siri is a personal assistant for Apple devices. Originally released in 2010, it is the oldest of the most used intelligent personal assistants. The most common uses for Siri are web browsing and dictation [11]. Currently, Siri is currently under a redesign process to add new functionalities and adapt it to new devices.

Some of the promotional commands for Siri include: "Show my photos from Utah last August", "What movies are playing today?", "Find videos I took at Iva's birthday party", "Text Pete 'See you soon smiley exclamation point'", "Read my latest email", "Find a table for four tonight in Chicago" [11].

2.5 Summary

This section summarizes the devices and intelligent personal assistants evaluated in this paper. Table 1 shows the two evaluated smart speakers.

Table 1. Device summary

Device	Manufacturer	Release date	Current price	Connectivity
Echo	Amazon	Nov 2014	$179.99	Data, Bluetooth, Wi-Fi
Home	Google	Nov 2016	$129	Wi-Fi

Table 2. Personal assistant summary

Personal assistant	Manufacturer	Release date	Supported languages
Siri	Apple	2011	20
Cortana	Microsoft	2014	8
Alexa	Amazon	2014	2
Google Assistant	Google	2016	4

Google Home functions with Google Assistant as artificial intelligence, Echo works with Alexa. However, this paper also considers two intelligent personal assistants that do not have a specific device (i.e., Cortana and Siri). Table 2 summarizes the four evaluated intelligent personal assistants.

3 Methods

This study included the participation of eight subjects (average age = 26, StaDev ± 5, males = 6, females = 2) familiar with technology and the use of Intelligent Personal Assistants.

During the execution of the evaluation, all participants gather in the same place. They listened to the responses of each personal assistant. Table 3 describes the devices used for each personal assistant and Fig. 1 shows the devices.

One of the main researchers followed a script of requests and repeated each request to each device separately in random order. Neither Text to Speech nor computer-generated speech were used during this research.

Table 3. Devices used to access each personal assistant

Personal assistant	Device
Siri	IPhone 7 plus
Cortana	Microsoft surface book
Alexa	Echo dot
Google Assistant	Google home

Fig. 1. Devices linked to the intelligent personal assistants.

There are many classifications of assistant functionalities. To define the evaluation procedure we will use the following classification [12]: shopping and buying assistant, care assistant, travel and entertainment assistant and administrative assistant.

The script used in this research had four categories (i.e., shopping and buying assistant, travel and entertainment assistant, administrative assistant, and miscellaneous). The care assistant category only focused on functionality it was evaluated separately.

One of the researchers presented the devices with a request using the same voice tone and pace. All the participants listened to the answers and gave a response using a 5-point Likert scale for naturality of the response and correctness.

In the shopping assistant category, four features were considered: managing shopping lists, ordering products from the internet, finding restaurants or stores and finding store schedules.

For the travel and entertainment category, five requests were included: sports update, movie theater information, transportation, distance and arrival time providing two locations, and identifying a song.

In the administrative assistant category, six features were considered: managing multiple timers and alarms, to-do lists, reminders for the same date and a specific date, meeting scheduling and composing emails.

Finally, for the miscellaneous category 21 features were evaluated: newscasts, traffic updates, weather forecast, device management, games, random functions (flip a coin, random numbers, roll a die), conversions, measurements, currency, tip calculator, solving math problems, and other functionalities such as (telling jokes, beatboxing, answering random questions or telling facts).

To evaluate framework compatibility, a review was executed and the devices were tested with the smart home devices available for the research team.

4 Results

This section presents the results of evaluation described above. Also, compatibilities of the intelligent personal assistants are presented.

4.1 IFTTT Framework Compatibility

If This Then That (IFTTT) is a public platform of services that has gained much industry momentum and adoption. IFTTT provides a logical paradigm for controlling services. Users log onto the IFTTT platform and configure their available devices to be activated by a large set of triggers.

In this research, a web review was conducted to assess the compatibility of each intelligent personal assistant with IFTTT framework and the only personal assistant that does not have full compatibility with the framework is Siri. However, Apple is working on an update that will supposedly add this feature.

4.2 Smart Home Framework Compatibility

This research also reviewed the compatibility of the personal assistants with smart home devices (the available devices during the research). In this case, Alexa, Google

Assistant and Siri were able to easily configure and access smart home devices. Finally, Cortana could not be configured. Reviewing Microsoft's website, this functionality is going to be added soon.

4.3 Evaluation Results

This section presents the main results of this research. The evaluation outcomes will be presented divided by features; afterwards, the results of each category will be described. Finally, we will present the general results of this research. Table 4 shows the best and the worst personal assistant in both correctness and naturality for each one of the features evaluated.

Table 4. Best and worst personal assistants by feature. Abbreviations: S = Siri, C = Cortana, A = Alexa, GA = Google Assistant, N/A = Not applicable

Feature	Best		Worst	
	Correctness	Naturality	Correctness	Naturality
Shopping and buying assistant				
Introductions	S	GA	C	A
Managing shopping lists	C	A	GA	GA
Shopping online	A	A	S, C, GA	S, C, GA
Finding restaurants	C, GA	C	A	S
Store schedules	C	C	S	GA
Travel and entertainment assistant				
Sport updates	A, S	S, C	GA	GA
Movie info	C	S, GA	A	C
Transportation info	S	A	GA	C
Transportation time	GA	GA	A	A, S
Identifying a song	C	C	GA	GA
Administrative assistant				
Alarms and timers	A, S, GA	A, S	C	C
Managing to-do lists	A	A	GA	GA
Reminders	S, C	S	GA	GA
Reminders in a date	C	C	GA	A
Schedule a meeting	C	C	GA	GA
Compose an email	S	S	A	A
Miscellaneous				
Local news	C, GA	GA	S	S
Traffic updates	C	C	GA	GA
Weather forecast	A, S, C, GA	S	N/A	C
Device management	A	A	GA	GA
Games	GA	GA	S	S

(*continued*)

Table 4. (*continued*)

Feature	Best		Worst	
	Correctness	Naturality	Correctness	Naturality
Random (coin)	GA	GA	C	C
Random (dice)	A, S, C, GA	GA	N/A	S
Random questions	GA	GA	S	S
Random facts	S	GA	C	A
Conversions	GA	C, GA	A, S, C	S
Measurements	C, GA	S, GA	A	A
Currency	A, S, GA	GA	C	C
Holiday information	GA	GA	C	C
Calculate tips	GA	GA	A	A
Math problems	S, C, GA	GA	A	A
Time zones	GA	GA	C	C
Jokes	GA	C	S	S
Beatbox	GA	GA	A	A
Translation	GA	GA	A	A
Word definitions	A	A	S, C, GA	S, C, GA

In the shopping and buying category, the best personal assistant was Cortana. However, the only one capable of shopping online was Alexa. Alexa and Cortana gave the most natural responses in this category. The least correct personal assistants in this category were Siri and Google assistant; also, Google Assistant gave the least natural responses.

In general, for the shopping and buying category, the Echo dot, using Alexa was the most liked device. This was expected as it is designed for shopping. The request with the best response in all devices was to find a restaurant. Ordering things was only available on Alexa. However, the request with the most natural response was the store schedules. The request with the worst response was shopping online as only Alexa was capable of doing it.

In the travel and entertainment category, Siri and Cortana share the first place in correctness and Siri was the one with the most natural responses. Again, Google assistant was the least correct in this category. Cortana and Google Assistant gave the least natural responses. Also, in this category, the request with the best (correct) and most natural response in general was sports update.

In the travel and entertainment category, Google Assistant gave the best response to the location and arrival time query. Alexa gave the worst response when prompted to get information about a movie.

The administrative assistant category has a tie in the most correct personal assistant between Siri and Cortana. The most natural personal assistant was again Siri. Google Assistant was the least correct and natural personal assistant in this category. The best-responded query was timers and alarms in both correctness and naturality. The best response in general was the handling of timers. However, Cortana gave excellent

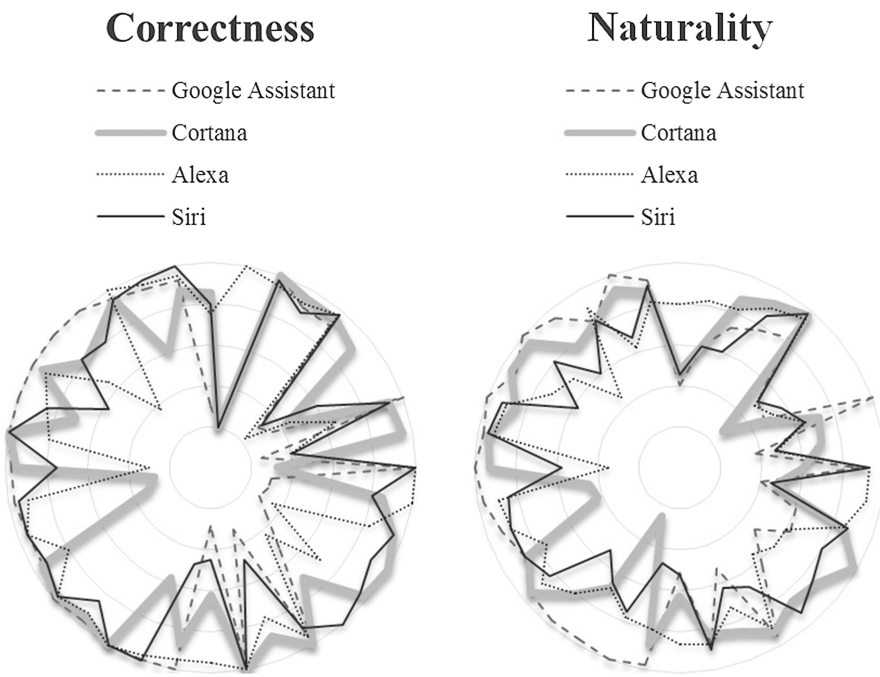


Fig. 2. Correctness and naturality by personal assistant. Interior circle means “poor” correctness/naturality; exterior circle means “excellent” correctness/naturality. Lines represent all tested features.

results sending emails and setting reminders. Google Assistant had the worst response when prompted to set a reminder.

Finally, Google Assistant overwhelmingly won the miscellaneous category both in correctness and in naturality. The difference in this category is of large; the second place (Siri) has less than half the points that Google Assistant has for correctness. As for the naturality, Google Assistant won this feature by over 400%. Cortana gave the worst results in correctness for this category, closely followed by Alexa and Siri. Alexa and Siri also were the least natural.

In this category, the best-responded request was the conversions. Google Assistant had the best-rated response in 11 of the 21 categories (perfect score). Cortana had the worst response when prompted to know which day of the week a holiday was. Figure 2 shows the results.

The results presented in this section were unified giving one point to each personal assistant when they had the best or worst response.

In general, performing the statistical analysis no significant differences were found between the systems. In other words, the evaluation shows no preference for any system, neither in naturality nor in correctness. This result was extracted of an analysis of all the results together.

Siri was the most correct device; however, Google assistant was the one with the most natural responses. Google assistant suffered because it does not support some of the features evaluated in this research. Siri had the worst results in naturality.

5 Discussion

Science fiction has shown the concept of intelligent assistants through movies, TV shows and books. New technologies and algorithms have allowed the development of personal assistants for commercial use for some years. Available assistants provide a vast variety of functionalities. However, this variety depends on the personal assistant implementation and its purpose.

There is an opportunity for improvement in all tested devices. No trend was observed. All devices showed at least one functionality with wrong or unnatural answers. Nevertheless, Google Assistant was the best in miscellaneous category and the worst in the other categories.

A remarkable feature of Google Assistant was the naturalness of answering some questions. The tone and pace of the female voice used by the Google device expressed surprise, suspense and joy. These features were no always offered by Siri, Cortana and Alexa. Even though, some answers were catalogue as natural using these devices.

Siri and Cortana enhance theirs answers with visual information. Also, they show what phrase or words are recognizing. Google Assistant and Alexa can show additional information too, but using their respectively apps. This additional information was helpful in some features using maps, photos, or graphics.

Finally, new studies must be conducted. The intelligent personal assistants' potential could be tested in unexplored areas like counseling, marketing, learning, and sales. Also, combination of these devices and technologies with robots, data centers, and machine learning techniques provides new opportunities.

Acknowledgements. This work was partially supported by *Centro de Investigaciones en Tecnologías de la Información y Comunicación* (CITIC), *Escuela de Ciencias de la Computación e Informática* (ECCI) both at *Universidad de Costa Rica*. Grant No. 834-B6-178.

References

1. Ghosh, S., Pherwani, J.: Designing of a natural voice assistants for mobile through user centered design approach. In: Kurosu, M. (ed.) Human–Computer Interaction: Design and Evaluation, pp. 320–331. Springer International Publishing, Cham (2015)
2. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Commun. **40**, 5–32 (2003)
3. Dryer, D.C.: Getting personal with computers: how to design personalities for agents. Appl. Artif. Intell. **13**, 273–295 (1999)
4. Becker, C., Kopp, S., Wachsmuth, I.: Why emotions should be integrated into conversational agents. In: Nishida, T. (ed.) Conversational Informatics: An Engineering Approach, pp. 49–68. Wiley, New York (2007)

5. Ball, G., Breese, J.: Embodied conversational agents (2000)
6. Amazon Inc.: Amazon Echo. www.amazon.com/echo%0A
7. Amazon Inc.: Alexa skills kit. https://developer.amazon.com/public/solutions/alexa/alexa-skills-kit
8. Google: Google Assistant. https://assistant.google.com/
9. Google: Google Home. https://madeby.google.com/home/
10. Microsoft Coorporation: Cortana. https://www.microsoft.com/en-us/mobile/experiences/cortana/
11. Apple: Siri. http://www.apple.com/ios/siri/
12. Sathi, A.: Cognitive devices as human assistants. In: Cognitive (Internet of) Things: Collaboration to Optimize Action, pp. 29–39. Palgrave Macmillan US, New York (2016)

Human Factors Affecting the Development of Smart Device-Based Notifications

Marcelo Guzman(✉), Gustavo López, and Luis A. Guerrero

Universidad de Costa Rica, San José, Costa Rica
{adrian.guzman,gustavo.lopez_h,
luis.guerreroblanco}@ucr.ac.cr

Abstract. Ambient Intelligence (AmI) promotes a world in which almost every device has computational capabilities. Such devices will allow a natural user interaction with the digital world. In addition, in several contexts developers have used smart devices to gather information of the user and provide notifications. The system consists of three main modules: one for data collection, one for analysis and decision-making, and one information deploying. More importantly, we conducted a questionnaire and the used the results as basic rules for the decision-making module on the system. The questionnaire had 60 respondents.

Keywords: Notifications · Smart devices

1 Introduction

Notifications are an important feature that allows information systems to inform people. However, attending notifications often has a disruptive effect on the task that people are performing [1].

Some trends promote the use of smart devices to deliver notifications. For instance, the living-room TV could deliver notifications when the user is sitting on the couch, the refrigerator could deliver notifications when a user is approaching to get food, the bedroom lights could change colors to deliver a message to the users in bed or smart speakers around the house could deliver notifications using natural language.

Smart devices are electronic artifacts that are connected to a network and can operate interactively and autonomously [2, 3]. Currently, smart devices are becoming ubiquitous. One of the core features of smart devices is they are able to deliver notifications in different formats, depending on their capabilities [4].

The problem here is that this avalanche of notifications can have a negative impact to our mind and mental health, it can make us addicted to devices, or we could ignore important notifications [5]. One way to organize this problem is to know from the user what kind of notifications are important and in which device they want to receive them.

In this paper, we propose a framework that allows decision making on which device should deliver notifications, depending on the devices available in a person's context, the notification degree of urgency and where and what the user is doing at the moment of receiving a notification.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_24

Moreover, we conducted a questionnaire to determine what users think would be a good way to deliver notifications. The results were tabulated and introduced into an expert system (detailed in a previous work [6]).

The structure of the paper is as follows: Sect. 2 presents an overview of the proposed system; Sect. 3 describes the methods of the study, focusing on the questionnaire. Section 4 describes the results of the questionnaire and a discussion. Finally, Sect. 5 concludes this paper.

2 System Overview

The system considered in this paper was described in other work [6]. However, to consolidate and understand the results presented in this paper, it is necessary to understand and know about the system. Figure 1 shows the basic architecture of the system.

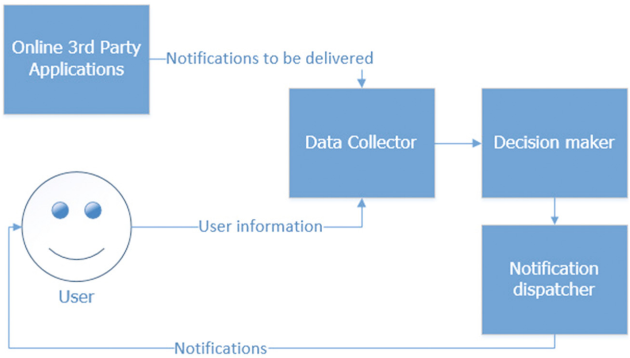


Fig. 1. System architecture

Online third party applications are those applications that require delivering notifications to the users. The data collector is the recipient of the information that the system is able to gather about both the user and the notification requests. The decision maker module uses the information contained in the data collector to determine the proper way to deliver the notification. Finally, the notification dispatcher delivers the notification.

In this paper, we describe a questionnaire designed to gather information that would allow the creation of a first set of rules for the decision-making module.

3 Method of the Study

Most of the previous work done in this field focused on the effects of interruptions and framework proposals to deliver notifications using smart devices. While this is certainly important, the human factor in this matter is also important. Therefore, we conducted a survey to add the human perspective in this decision making process. This section describes the design and procedure of the study, the instruments, and the participants.

3.1 Design

In this study, we used the results of a previous work in which we asked HCI experts to define which devices a notification system should choose to deliver notifications. Following this, we asked 60 people to answer a large questionnaire with a set of situations and a set of smart devices.

The scenarios reflect everyday situations of the potential users of the system both in context and situations. Each scenario consists of two parts: the event and the device. The event describes a fictitious situation, defined based on experiences and comments of the participants. The interviewee decides based on the description the level of urgency to choose a device for receiving the notification. The device selection is the question that addresses given a situation, which device will suit the notification requirements and would be best to deliver a notification.

We chose the devices considering different characteristics. We included wearables, home and office devices. Moreover, we considered personal and shared devices. Some of the devices are traditional devices with new functionalities (i.e., smart TV, refrigerator), other devices are a current buying trend (e.g., smart speaker, smart watch). Finally, we chose some device for their use or location (e.g., car, office and home computer). Figure 2 shows the included devices.

It is worth mentioning that some users do not feel identified with some of these devices because they do not currently use them, this makes us assume that the preference of choosing a device to receive the notifications has to do with the availability of a device in the market, cost, and time to learn how to use it.

Fig. 2. Devices considered in the study.

3.2 Instruments

We used Google Docs form capabilities to set the questionnaire and distribute it around. The survey presented an introduction stating: "This questionnaire's goal is to feed an experimental expert system to centralize notifications in a user's context, depending on the notification type, urgency and user location".

Afterwards, we provided instructions and contact information. The first set of questions was demographic (i.e., gender and age). The second addressed the users' knowledge and capabilities with smart devices. The third one presented scenarios and asked for an answer of which smart device should deliver the notification. Finally, we included three questions regarding the amount of notifications received per day, the likelihood of using smart devices to receive notifications and an open question for comments.

3.3 Participants

We recruited participants using social media and mailing lists by describing the study and stating that we are looking for participants with knowledge of smart devices. Sixty people responded the questionnaire. All of them were from Costa Rica. Figure 3 shows the distribution of participants by age and by gender.

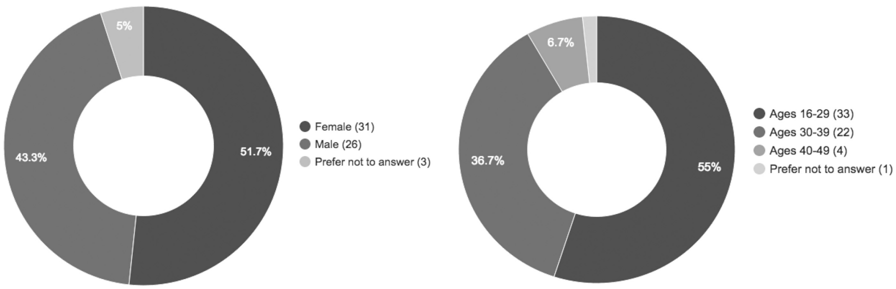


Fig. 3. Participants' distribution by gender (left) and age (right)

4 Results and Discussion

Based on our previous work [6], a new questionnaire was made with more scenarios, in order to retrieve more data from end users. For review and discussion purposes, we divided the results based on the location of the user when the notification occurs.

In this case, the results are divided if four areas: office, home, "on the street" and "while driving". The difference of the last two areas is that, in one we are observing the preferences when the focus of the user is an important task that requires all the possible attention (like driving), and the other ("on street") gathers information while the user is performing less important tasks.

After executing the questionnaire, based on Fig. 4, we realize that all interviewees feel comfortable using smart devices, where only six people feel an average level of confidence, the rest of respondents feel that they take advantage of their devices.

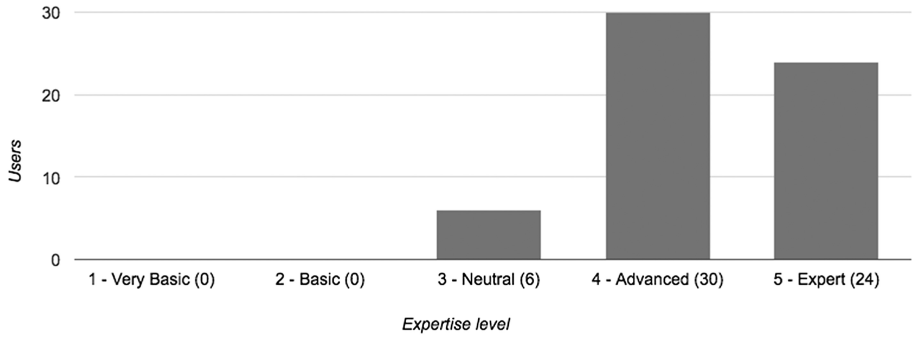


Fig. 4. Smart device technology knowledge

We also asked the participants about their willingness to use a system like the one presented in this paper (after a brief introduction). None of the participants stated that they will never use it, 15% stated that they might use it, and 23% stated that they do not have an opinion. The positive queue of the distribution gathered 62% of the respondents between "I will use it" (40%) and "I will always use it" (22%).

Another question asked for the user criteria on the amount of notifications that they receive per day. In this case, the options ranged from "very few" to "too many". Again, none of the respondents stated that they receive very few notifications and 5.4% stated that they receive a few. The largest amount of respondents stated that they receive a normal amount of notifications (41.1%). Moreover, 37.5% respondents stated that they receive many notifications and 16.1% receive too many notifications. Even though we did not explicitly ask each user the amount of notifications they receive per day, this result is an indicator of the user satisfaction on the amount of notifications received. The next sections describe each scenario and the main findings.

4.1 Location: Office

Table 1 shows the results of nine test scenarios with the office as location. The columns on Table 1 represent the device that would deliver the notification, the top three devices chosen by the respondents. We highlighted the device with a highest score in each situation.

The first place is occupied by the office computer, where we assume the user spends many hours of the day, is the best option to receive these notifications, followed by the smartphone, for being a personal device and which our intervieewees already accustomed to use them and keep their notifications in it. The third place shows an

Table 1. Results using "office" as location

Location: office	Office computer	Smartphone	Smartwatch	Other devices
Payroll	**29**	28	2	1
Car accident	16	**31**	3	10
Stock change	**34**	23	2	1
Intruder alarm	**30**	24	5	1
Airline ticket alert	**28**	26	6	0
Meeting alert	**36**	14	9	1
Shopping delivery	15	**39**	3	3
Speeding ticket	**28**	25	1	6
Fuel price alert	**25**	21	4	10

interesting trend, the smartwatch gains popularity on end users while its models improve their features and their cost of purchase is becoming smaller.

We can interpret the results obtained, for most users the work computer is the best place to receive notifications. In addition, we can assume that users do not perceive as negative to receive some personal notifications like a speeding ticket notification (28 users selected the office computer vs. 25 that selected the smartphone), or a change in the stock market (34 office computer users vs. 23 smartphone).

In the case of the price change for an airline ticket, the difference is small. This could be related to the high amount of mobile apps on the market focused on travel discounts for flights, hotels, and tours, are best positioned that their website counterparts, for watching offers and discounts.

In two cases, the smartphone was considered the best option. First, a car accident notification on the user's route to go home, this could be to allow the user to change their commute route having this information at hand when leaving the office. Second, when the user makes a purchase, the difference here is of 39 vs. 15. This could be due to the type of products the user usually buy (e.g., personal use, luxury, niche products) a personal device is preferred instead of receiving this information on the work computer and exposed to nearby colleagues.

In only two cases, some users preferred to have their notifications on the screen car display instead of the 2 most popular devices. When there is a car accident and a fuel price alert. We can assume these users feel that having the information in the car can get their attention while driving.

4.2 Location: In the Car, Driving

In the case in which the user is in the car, driving, the main preference for receiving notifications remains to be the smartphone. In this category, 5 scenarios were analysed. Table 2 shows the results.

Table 2. Results using "in car" as location

Location: in car	Smartphone	Car screen	Smartwatch	Other devices
Car accident	**31**	27	2	0
Stock change	**34**	15	4	7
Intruder alarm	**39**	18	0	3
Groceries list	**42**	10	0	8
Bank notice	**28**	18	7	7

The second device selected was the car screen. However, the difference between the smartphone and the car screen is significant. The only notification that is close for both devices is the one related to driving (i.e., car accident).

We can assume those preferences are based on the availability for a user to own a smartphone compared to having a device in his car with internet connection, like a touch screen car stereo. In the future, when car screens become more available to new and used cars, this preference might change.

The smartwatch is again a third viable option to receive these notifications, perhaps this device is perceived as a complement of the smartphone. With other devices, it is possible to conclude, for example, the change in the stock market and the notification from the bank, does not represent a priority when the user is driving, and this type of notification can be postponed to be reviewed later, may be in a device that offers more interactivity and options for the user.

4.3 Location: At Home

When the respondents are in the house, the preference for a device to receive notifications seems to be related to the activity being performed (watching TV, cooking or resting). In this context, and based on Table 3, the smartphone is still the device preferred by the interviewees. However, other devices such as smart TV, home computer, and smart speaker are options also considered.

We can assume the choice of a smart speaker for some people, is because they can be early adopters of new technologies, can take advantage of this device.

Table 3. Results using "at home" as location

Location: at home	Smartphone	Smart TV	Home computer	Smart speaker	Other devices
Car accident	**25**	5	1	15	14
Stock change	**31**	0	14	4	11
Groceries list	20	0	6	8	**26**
Bank notice	**32**	13	11	0	4
Speeding ticket	**31**	11	12	1	5
Favorite show	11	**40**	1	4	4
Airline ticket alert	**23**	**23**	5	4	5
Payroll	**38**	7	4	3	8

4.4 Location: On the Way

When the interviewees are on the street, the smartphone and the smartwatch are the popular choices, based on Table 4. We see in this scenario, the smartphone has more people willing to use it to receive all their notifications than any other device. The closest choice, the smartwatch has only a few users willing to use it in comparison to the smartphone. We assume the choice of those two devices, due to the availability for the user, cost and ease to use when walking, watching a movie, shopping and doing any other activity on the street.

Table 4. Results using "on the street" as location

Location: on the street	Smartphone	Smartwatch	Other devices
Emergency call	**55**	5	0
Credit card fraud alert	**52**	6	2
Groceries list	**50**	5	5
Bank notice	**40**	3	17
Unauthorized access	**48**	11	1
Water supply outage	**51**	7	2

The only exception to the results presented in Table 4 is when the notification is a bank notice, where the smartphone has 40 users who prefer it, and 13 respondents preferred to receive their information at home, on their computers, instead of the smartwatch, the closest option in general for this scenario. This type of notifications has no relevance when the user is in activities that do not involve having a computer or a similar device nearby to review the bank website, to get details about that bank notification.

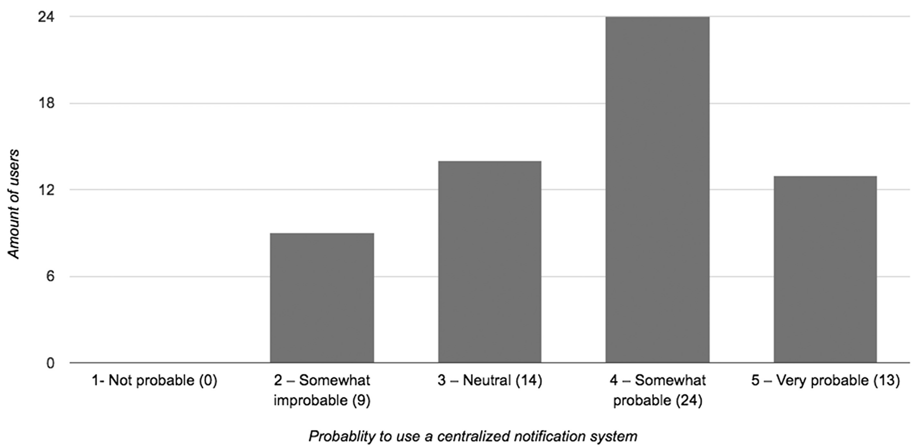


Fig. 5. Results to the question "will you be willing to use the system?"

5 Conclusion

In this paper, we investigated notifications in smart devices, focused on human factors affecting the decision of which device suits each notification. We conducted a survey with 60 participants and 6 different devices (speaker, office and home computer, car screen, refrigerator, smart watch, smart TV, tablet, smart phone). Additionally, we gather information on the use of smart devices and amount of notifications received by the participants.

The system that is described in this paper includes three main modules. However, the focus of this work is the decision-making module. The questionnaire was executed to gather insights from users on how they would like to receive notifications. The most common response was the smartphone. In other words, users prefer their notifications on their phones. We believe this result is caused by the extended use of this device.

The adoption of new smart devices could lead to a change in this perspective from the users, and works like the one presented in this paper are a step forward in that direction.

It is interesting to point that the participants of this study stated that they would be willing to use the proposed system. Figure 5 shows the distribution of the responses to the question: Will you be willing to use the system?

In general, the results of this paper show that users like smartphone notifications. However, it also shows that they will be willing to get their notifications through other mechanisms, including smart devices.

Acknowledgements. This work was partially supported by *Centro de Investigaciones en Tecnologías de la Información y Comunicación* (CITIC), *Escuela de Ciencias de la Computación e Informática* (ECCI) both at *Universidad de Costa Rica*. Grant No. 834-B6-178.

References

1. Kubitza, T., Voit, A., Weber, D., Schmidt, A.: An IoT infrastructure for ubiquitous notifications in intelligent living environments. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, pp. 1536–1541. ACM, New York (2016)
2. Weiser, M.: The computer for the 21st century. Mob. Comput. Commun. Rev. **3**, 66–75 (1991)
3. Poslad, S.: Ubiquitous Computing: Smart Devices, Environments and Interactions. Wiley Publishing Inc., London (2009)
4. Weber, D., Voit, A., Kratzer, P., Henze, N.: In-situ investigation of notifications in multi-device environments. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1259–1264. ACM, New York (2016)
5. Okoshi, T., Nozaki, H., Nakazawa, J., Tokuda, H., Ramos, J., Dey, A.K.: Towards attention-aware adaptive notification on smart phones. Perv. Mob. Comput. **26**, 17–34 (2016)
6. López, G., Guzman, M., Marín, G., Guerrero, L.A.: Towards smart notifications—an adaptive approach using smart devices. In: International Conference on Ubiquitous Computing and Ambient Intelligence (in press). Springer International Publishing, Berlin (2016)

Developing a Proxy Service to Bring Naturality to Amazon's Personal Assistant "Alexa"

Luis Carvajal(✉), Luis Quesada, Gustavo López, and Jose A. Brenes

University of Costa Rica, San Pedro, San José, Costa Rica
{luis.carvajalrojas, luis.quesada, gustavo.lopez_h, joseantonio.brenes}@ucr.ac.cr

Abstract. Amazon's Alexa is an intelligent personal assistant, developed to be used jointly with a Bluetooth Speaker and microphone hardware called Amazon Echo. Even though Alexa is supposed to be a natural user interface, its use is not very natural. As a user is intending to use Amazon's Alexa system, they must follow a very tighten and structured way to provide the commands for the system to achieve its goal. In this paper, we propose a proxy service called "Plis". This service was developed as a Skill to be used with the Amazon Echo. Therefore, a user can say: Alexa, "plis" and our functionality will start its job. Our skill determines from the natural way in which the user speaks what they are asking for. With that information, we create a query that would be further sent to other skills already providing the functionality required.

Keywords: Intelligent personal assistant · Amazon Echo · Alexa

1 Introduction

Natural user interfaces are software systems developed to feel as natural as possible to the user [1]. However, it is not enough to use non-traditional interfaces to call a system natural (i.e., using speech recognition does not create a natural user interface). In this paper, we address the interaction naturality of a speech-based personal assistant developed by Amazon, called "Alexa" [2, 3].

Alexa uses speech recognition as interaction mechanisms. However, the way in which its capabilities are implemented are computerized (i.e., non-natural). The core of Alexa (developed by Amazon) has a natural interaction. However, the API provided by the company to develop functionalities (skills as they call them) allows third- party programmers to add functionalities and these functions are not natural.

An example of the lack of naturality in the interaction with third-party skills can be determined just by observing how they are lunched. "Alexa, open Starbucks" is the way to activate Starbucks Reorder skill. However, this skill is only able to reorder products that you have already order. This interaction is not natural. The user should be able to say "Alexa, reorder from Starbucks".

An even worst example is the way in which skills are positioned as people. For instance, Starbucks reorder also offers a functionality activated by this sentence: "Alexa, tell Starbucks to start my order". Why would a user say tell Starbucks rather than telling Alexa (its personal assistant) to start my order.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_25

In this paper, we address this issue by providing a proxy service that helps users to interact with Alexa in a more natural matter. We provide a framework that classifies skills and allows the user to access them from one centralized skill called "plis". Therefore, the user can state "Alexa, 'plis' start my Starbucks order" and our skill will look from the available skills one that can achieve this request and redirect it.

2 Related Work

This section presents works related to the one proposed in this paper. The basic foundations for this work are based on several years of research. For instance, since 2002, Michael F. McTear [4] described the main components of spoken dialogue technologies. He included talking capabilities, natural language processing, external communication, speech generation and voice synthesis. By this year, authors were already considering and proposing the way in which intelligent personal assistants should interact with the users.

With this base, several works have been performed addressing the use of intelligent personal assistants. For instance, Wenbin Li, Ning Zhong, Yiyu Yao and Jiming Liu [5] described a system that works as a personal assistant. Their system provides several capabilities specially focused on email management. However, this type of systems does not focus on the user interaction.

Another type of systems emphasizes more on the adaptive nature of personal assistants. For instance, Pauline M. Berry [6], developed a project that evaluates a system that helps with calendar and schedule management. This paper focuses on the learning capabilities of the personal assistant and its adaptation to the user needs.

Focusing on the measurement off personal assistants, Umair Saad, Usama Afzal, Ahmad El-Issawi and Mohamad Eid [7] evaluated the user experience reported with the use of two different personal assistants. In that paper, authors describe how the hearing queues help user and improve the general interaction.

It is important to consider the application context of a personal assistant. For instance, Bassam El Saghir and Noel Crespi [8] address the need of personal assistants in context-aware systems. In that work, authors focus specially in the communication characteristics, and the basic requirements for effective communication (i.e., autonomy, proactivity, reactivity and adaptability). Another example was deployed in research and development by Emerson Paraiso and Jean-Paul A. Barthès [9] who described the application of intelligent personal assistants in a computer supported cooperative work environment. Specifically, they tested the assistant usefulness in semantic interpretation research.

3 Amazon's Alexa

This section describes Amazon's Personal Assistant functionality. Moreover, there is a study abridging the potential skill that can be used by the developed skill ("Plis"). Finally, this section presents a description of the development process and tools.

3.1 Functionality

Alexa is an intelligent personal assistant developed by Amazon. Alexa's adoption growth was boosted by the use of it in the Amazon Echo Bluetooth speaker. There are two types of functionalities for Alexa: Alexa commands and third-party skills [3].

Amazon developers are working on new commands available for Alexa every day. However, Alexa also provides an API that allows third-party developers to create and publish skills that can be used through Alexa. Skills are accessible through the basic commands of Alexa. A user can say "Alexa, open [name of skill]" and this will launch a skill.

In order to use the Amazon Echo a user has to follow a rigid structure. First, the keyword (Alexa or Echo) should be pronounced. Second, the name of the Skill (name of the applications for Alexa) should be mentioned; otherwise Alexa's core system will be used. If a third-party skill is being used, the name of the functionality and its parameters should be provided.

As an example, a user can be heard saying 'Alexa (Keyword), use TellMeTheScore (App name), results for (Functionality) Real Madrid vs. Barcelona game (parameter)'. This is not the way natural user interfaces (NUIs) are supposed to work.

Currently, there are a lot of limitations on what skills developed by third parties can do. Therefore, to evaluate our skill, we simulate the communication with other skills.

The main problem solved by this proposal is the way in which current skills work. Alexa's core system is very well defined and implemented (i.e., usable and natural). However, most third-party skills are developed using a set of provided functionalities and are left to be used in a very structured way. This proposal helps to make more usable those third-party Skills.

Some of the abilities of Alexa embedded in the core of the system (i.e., without using third-party skills) include: core commands, media controls, time and dates, to-do lists management, news and weather basic functionalities, entertainment, definitions, dictionary or math abilities. Finally, other embedded capabilities comprise smart home management and purchasing online.

Alexa's basic commands provide volume management, mute or unmute, stop or resume audio, access to help, documentation and account management.

The media controls give access to music services, playing songs or audiobooks. It also allows playing by author, playlists, getting information of the audio currently playing, general audio management. Timer and alerts let on setting, canceling, snoozing and checking the time. Similarly, to-do lists allow users to create, delete and manage lists.

In entertainment, Alexa allows the user to ask about movies, theater, bands and other famous people. Math and grammatical help provide users to recognize errors, perform grammatical or mathematical tasks.

Alexa can integrate with some of smart home platforms, such as SmartThings, Wink, Insteon, Lutron, Belkin WeMo, Philips Hue and more [3]. Some of these capabilities are built in the core of the system, other require additional skills to be installed.

The next section describes how the skills published in the Alexa website were categorized in order to be used in this project.

A metadata database was created to list the functionalities of the third-party Skills and as the user speaks, a comparison with the metadata database is done to try and select a proper Skill and its functionality. We compare the way in which the Skills are traditionally used, and the way in which users can achieve their goal using our service.

3.2 Skill Categorization

Some of the categories specified by Amazon for Alexa skills that this research considers are: business and finance, shopping, news, music, sports, social, transportation, weather and miscellaneous. Table 1 shows the skills considered in the business and finance category. These skills include business news, taxes information, economic indicators and financial related data (currency, exchanges or stocks).

Table 1. Skills considered for the business and finance category.

Skill name	Description
Federal status	Check for federal offices available
TFM tool (unofficial)	Lookup transactions codes based on debit and credit card
TXU energy	Get account balance, make a payment and get information
BizTalk	Latest business news
Indicators	Keep up with the state of the economy
Currency today	Ask for currency rate and convert between currencies
Stock info	Provides the quote of any stock listed on US stock exchange
Sales tax calculator	Quickly calculate taxes

The shopping category skills are shown in Table 2. Skills in the shopping category include shopping recommendations and store schedules. The news category is not shown as a table since most of the skills are feeds provided by newscasters and their functionality is just to give a flash briefing.

Table 2. Skills considered for the shopping category.

Skill name	Description
Make my look	Help you to match your outfit
Black Friday calendar	Ask for sale hours for major retailers
Gift genie	Will suggest a gift instantly
Kit	Ask for recommended products for your hobbies
My deal	Show the top deals and on sale events
Still on	Find store hours an business Information in your area
Shopper	Tell where is the cheapest meal
Starbucks reorder	Reorder your usual from a nearby starbucks

Table 3 shows the music category. In the music category, the skills include music services, radio stations and services for musicians. Moreover, some skills to help amateur musicians learn were considered.

Table 3. Skills considered for the music category.

Skill name	Description
Spotify	Play music from spotify
Radio Richard diamond	Listen to the golden age of radio
Ear trainer	Helps to improve your interval recognition
Starfish audio	Play your music stored on cloud one drive
Guitar tuner	Tune your guitar
Musixmatch	Helps you to learn song lyrics
Fluffy radio	Request your favorite song to FluffyRadio.com

One of the categories with the largest amount of skills available is the sports category. Table 4 shows the skills considered in this category. The sports category includes skills developed for specific teams, of sport type. Moreover, skills that let on users to find other ways to follow sport results were considered.

Table 4. Skills considered for the sport category.

Skill name	Description
Racing results	Horse racing results from UK and Ireland
Ravens fan	Ravens stats & fun
Seahawks fan	Seahawks stats & games
AZ cardinals	AZ cardinals last news
Chicago baseball (not official MLB)	Standings and the last game results for Chicago
Unofficial arsenal fixture	Gives information about arsenal FC
Arizona sports 98.7 FM	Listen to arizona sports 98.7 FM
Follow the game	Find out what radio station is playing.
Soccer bot	Last matches and info about soccer.
Falcons fan	Falcons stats & games
Gymnastics facts	Info about gymnastics events
Patriots fan live	All about patriots
Catch sports	Personal assistant for live sports
MLB game audio	MLB.com live radio
Formula one race information	Positions of formula one

The social category is show in Table 5. The social category includes communication and notification mechanisms and access to social media systems. The transportation category is shown in Table 6. The transportation category includes skills that allow users to access information about travel times, ground transportation services in different cities and flights.

Table 5. Skills considered for the social category.

Skill name	Description
Vaa you	Send hands free voice activated notifications
Ask my buddy	Alert your contacts
Message wall	Write anonymous notes for each other
Olabot	Create a virtual version of you
Starfish me	Record your own responses for phrases
AT&T send message	Send messages with your phone
Mail reader	Check your Gmail
Shootipedia	Browse, find, rate, share and submit photography locations
Choice assistant	Guides for making decision
Twitter reader	Can read your home timeline, mentions, retweets and likes
Agog reader	Listen to the last 10 tweets of the most popular accounts

Table 6. Skills considered for the social category.

Skill name	Description
Travel time	Calculate how long it will take you to get to where you have to go
City bike	Find out how many available bikes there are at your saved city bike station
NYC subway	Checks the status of NYC's subway lines
Station agent	Train arrival times
Flight deals	Ask for flight deals from your city
Dave's airport	Obtains airport weather data and reads aloud
Lyft	Ask for a ride and get picked up in minutes
Uber	Ask for a uber ride and get picked up in minutes
Expedia	Check for your upcoming trips and your flight status
Kayak	Track flights and check prices on flights, hotels and rental cars

The last two categories are weather and miscellaneous. These categories are shown in Tables 7 and 8, respectively. The weather category includes services to get information about tides, winds and temperature. Moreover, services that recommend clothes for the weather were considered. The miscellaneous category includes services that are not specific to any of the other categories.

Table 7. Skills considered for the weather category.

Skill name	Description
Tide guide	This skill will give the tide report for most locations with a NOAA station
WeatherBug	Provides current weather conditions, including weather alerts and lightning information
What should I wear today?	Tell you what you should wear on a given day based on the current temperature
Temperature inside	Ask temperature inside point of sale
Wind report	Ask for the wind conditions in your city
Fish wise	Analyzes weather patterns and determine the best times and places to fish
Weather	Amazon weather skill

Table 8. Skills considered for the miscellaneous category.

Skill name	Description
This day in history	Find out about the historical events that happened on this day and every other day of the year
Magic 8-Ball	Get a random yes/no answer for difficult decisions
Sleeptracker	Ask Alexa about your sleep and control your Sleeptracker device

3.3 Developing Process

Amazon provides a sundry services and tools for the development in its devices, such as Amazon Developer console and Amazon Web Services (AWS) which includes: AWS Lambda, Amazon DynamoDB, Amazon VPC, Amazon API Gateway, Amazon IoT Platform, among others [10]. To develop an Alexa skill requires the Amazon Developer Portal and AWS Lambda specifically. The first step to program a skill is to write or upload the code in Lambda and then to configure the skill in the Developer Portal.

A skill that is being developed has its code in Lambda; see Lambda as "… a serverless compute service that runs your code in response to events and automatically manages the underlying compute resources for you…" [10]. The code can be written in Java Script, Python and S3 (Amazon programming language). When a user makes a request, it is dispatched to Lambda service for execute the corresponding intents and instructions, then Alexa retrieve the response from Lambda and emits it to the user.

The Amazon Developer Portal functions are:

- Skill Information: Set the skill type, language, skill name, application id and invocation name.
- Interaction Model: Write the Intent Schema and Sample Utterances.
- Test: Use the Service Simulator to write some utterances and see or hear the Lambda request and response.
- Publishing Information: Set the category, sub-category, testing instructions, skill description, example phrases, key words and images.
- Privacy & Compliance: Set the privacy request, privacy policy and terms of use.

4 "Plis" Architecture and Deployment

In this paper a skill for the Amazon Alexa personal assistant was developed and deployed. The development process of a skill has two general steps, write or upload the code using Amazon AWS Lambda service and add the skill information and interaction model with the developer portal.

In order to allow a natural interaction with our skill and avoid using a code name, we decided to call our skill "Plis" based on "please". Therefore, a user could ask "please" and the present a request that our skill will redirect to another skill with the capabilities to achieve what the user is asking for.

Figure 1 shows the general architecture of the "Plis" skill. The system uses a database to store the meta-data of the skills available. A hash structure is used to store the information.

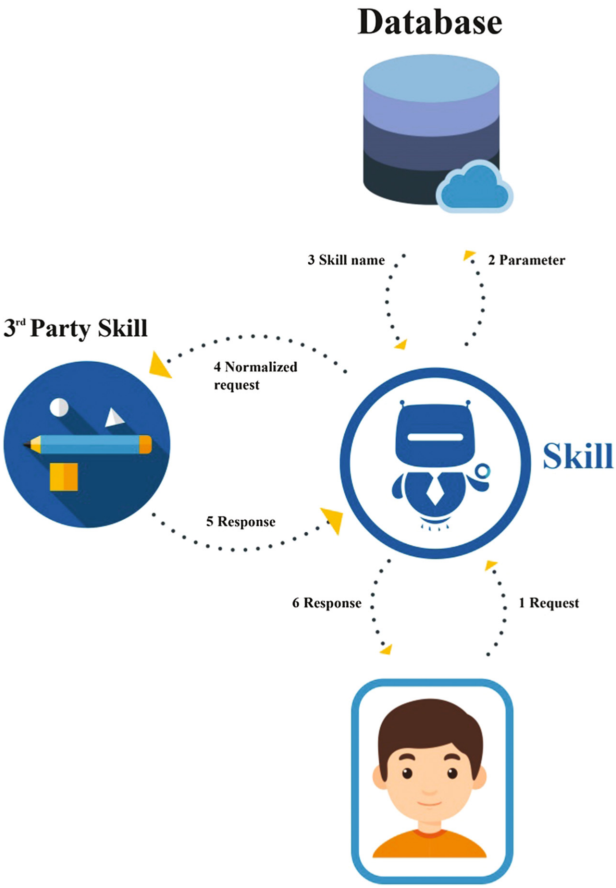


Fig. 1. "Plis" basic architecture

According with the architecture showed in Fig 1. The system works as follows:

1. A user asks "plis" something.
2. "Plis" skill runs the request on the database to see what the user is requesting and the parameters are stored separately.
3. After searching in the database, a skill name will be provided and that skill will be summoned with the previously stored parameters.
4. The request to the third-party skill is what we call normalized request.
5. The third-party skill response is then gathered and delivered to the user.

For a complete understanding of how the system works, let´s see an example. An user ask "*Alexa, plis what was the score of Patriots versus Eagles last game*", when Alexa hears the trigger word "plis" it will invoke our skill, being in the "Plis" Proxy it get the parameters, in this case they will be: score, Patriots, Eagles and last. These parameters, except the score, are stablished in the Interaction Model and received by the Lambda code. Each parameter has its own process, score parameter, for example, is the question as well, Patriots and Eagles are the team names stored in a hash related to their corresponding sport with a numeric categorization, and last is the time parameter. After processing all the parameters, the skill use them to find the corresponding third-party skill (if there is one) in the metadata database, then pass it those parameters and executes the request in the third-party as "*Alexa, ask Follow the game what was the score of Patriots versus Eagles last game*", then Plis Proxy will retrieve the response and emit it on the Echo.

Table 9. Differences between call a third-party skill against and using call "Plis" skill as a proxy service.

Calling a 3rd party skill	Calling a 3rd party skill using "Plis" Proxy
Alexa, ask Soccer bot about the next Manchester United match	Alexa, plis when is the next match of Manchester United?
Alexa, ask Soccer bot about the last Barcelona versus real madrid game	Alexa, plis tell me about the last real madrid vs. Barcelona game
Alexa, open travel time	Alexa, plis tell me the travel time to New York
Alexa, travel time to New York	
Alexa, ask Uber to call me an Uber SUV to work	Alexa, plis call me an Uber SUV to work
Alexa, ask sleep tracer to star recording sleep	Alexa, plis record my sleep
Alexa, open falcons fan live	Alexa, plis tell me the last score of falcons
Alexa, ask falcons fan last score	
Alexa, ask MyBuddy to alert James	Alexa, plis alert James
Alexa, ask AT&T to text Mike	Alexa, plis text Mike
Alexa, Ask Follow the top 3 in the table.	Alexa, plis tell me the top 3 football teams
Alexa, open choose assistant	Alexa, plis help me to choose an appetizer
Alexa, ask choose assistant for appetizer assistant	
Alexa, launch this day in history	Alexa, plis give an event of September 3rd
Alexa, tell me about another event of September 3rd	
Alexa, ask WebMD what are the side effects of Nexium	Alexa, plis give me the side effects of Nexium
	Alexa, plis tell me the side effects of Nexium
Alexa, ask Expedia to reserve a car	Alexa, plis reserve a car with Expedia
Alexa, ask my valet how much charge does my car have	Alexa, plis how much charge does my Tesla have
Alexa, ask my extra brain to remind me to wash the car on saturday	Alexa, plis remind me to wash my car on Saturday

In some cases, when the user's request searched in the database and no skills are found or sports don't match, "Plis" skill will immediately give a response letting the user to know that his request cannot be fulfilled.

There are several requests that the user can ask to "Plis" skill. Table 9 shows the difference between the traditional way in which third-party skills would have to be launched and how the same request can be performed through "Plis" skill in a more natural way.

The skill detailed in this section was developed and tested, simulating access to other skills. A major drawback of the current configuration of Amazon Alexa is that skills are segregated. Calling a skill form another one can only be performed with permissions and access to detailed information that only the provider manages.

Therefore, in order to achieve massive adoption of this system, collaboration between third-party skill providers would be necessary.

5 Conclusion

This work presented a proxy service developed to bring naturality to the use of Amazon's Alexa personal assistant. For the most part, Alexa's capabilities are usable and natural. However, since Alexa offers an API for third-parties to develop skills, these skills are not as natural as the basic services of the personal assistant.

By using third-party skills, a user can find himself asking something like this: "Alexa, ask TellMeTheScore to give the results for Real Madrid vs. Barcelona game". Listening to this request it is certainly not natural. A user would prefer ask an assistant: "Alexa, what was the score of yesterday's match?" or even "Alexa, how was Real Madrid's game".

Our proposal serves as a proxy service with the codename "Plis". The proxy service gathers a request in natural language and processes it in order to determine a third-party skill that could provide the corresponding response. This helps the user to ask in a more natural way, instead of forcing the user to know the name of the skills and what they are capable of.

To develop the proxy, we build a metadata database that stores skills and keywords for their capabilities. In this way, when capturing a request from the user, a search in the database could be done and obtain a skill to solve the request. Then, the request is automatically called by "Plis" and the response is processed and delivered to the user.

The main goal of this project is not to centralize third-party applications, but to give a more natural access to the services they provide. In order to scale and deploy this skill, changes in the way skills work for Amazon Alexa would be required. Moreover, collaboration between third-party skill providers would be necessary.

Acknowledgments. This work was partially supported by *Centro de Investigaciones en Tecnologías de la Información y Comunicación* (CITIC), *Escuela de Ciencias de la Computación e Informática* (ECCI) both at *Universidad de Costa Rica*.

References

1. Ghosh, S., Pherwani, J.: Designing of a natural voice assistants for mobile through user centered design approach. In: Kurosu, M. (ed.) Human-Computer Interaction: Design and Evaluation, pp. 320–331. Springer, Cham (2015)
2. Amazon Inc.: Amazon Echo. www.amazon.com/echo%0A
3. Amazon Inc.: Alexa Skills Kit. https://developer.amazon.com/public/solutions/alexa/alexa-skills-kit
4. McTear, M.F.: Spoken dialogue technology: enabling the conversational user interface. ACM Comput. Surv. **34**, 90–169 (2002)
5. Li, W., Zhong, N., Yao, Y., Liu, J.: An operable email based intelligent personal assistant. World Wide Web. **12**, 125–147 (2009)
6. Berry, P.M., Donneau-Golencer, T., Duong, K., Gervasio, M., Peintner, B., Yorke-Smith, N.: Evaluating intelligent knowledge systems: experiences with a user-adaptive assistant agent. In: Knowledge and Information Systems, pp. 1–31 (2016)
7. Saad, U., Afzal, U., El-Issawi, A., Eid, M.: A model to measure QoE for virtual personal assistant. Multimed. Tools Appl. **76**(10), 12517–12537 (2016)
8. El Saghir, B., Crespi, N.: A generic layer model for context-aware communication adaptation. In: Wireless Communications and Networking Conference, pp. 3027–3032 (2008)
9. Paraiso, E.C., Barthes, J.-P.A.: An intelligent speech interface for personal assistants in RD projects. In: International Conference on Computer Supported Cooperative Work in Design, pp. 804–809. IEEE (2005)
10. Amazon Inc.: Amazon Lambda Description. https://aws.amazon.com/es/lambda/

A Customizable Calculator Application with 3D-Printed Cover for the Visually Impaired in China

Yifan He, Bo Zhu, and Pinata Winoto(✉)

Media Lab, Department of Computer Science, Wenzhou-Kean University, 88 Daxue Road, Wenzhou, China
{heyif, zhub, pwinoto}@kean.edu

Abstract. To date, a wide variety of applications are developed and available in iTunes App Store, Google Play and Windows App Market. However, the visually impaired are still unable to enjoy most of them because of the non-optimized design of these applications for them. Though some mobile operating systems provide auditory assistive functions, slow and tedious interaction style lead to a poor user experience. Therefore, it is necessary to design specific applications for the visually impaired, providing not only basic functions, but a good user experience as well. To achieve the goal, in this research, we focused on a frequently implemented task on the mobile phones. Specifically, a Windows Phone calculator application for visually impaired has been designed. Similar to a typical calculator, it allows users to perform some basic calculations such as addition and multiplication. However, unlike typical calculators, it has a special layout and audio clues designed for the visually impaired. To complement the application, we have also designed a 3D-printed cover with holes on it, serving as the tactile interface for the application. The application is highly customizable so that users can reassign each button according to their preferences. The entire solution, including the Windows Phone calculator application and the 3D-printed cover, can provide visually impaired users the basic function of calculations as well as a rich user experience. With this solution, a considerable improvement with respect to the operational speed was found during a simulation test. A pilot experiment has been designed to test the working efficiency using blindfolded users.

Keywords: Visually impaired · Assistive technology · Tactile interface · Mobile interaction · Customizable · China

1 Introduction

Recent years mobile phones are becoming a necessity in the everyday life of many people. In China, the largest mobile phone market in the world, there are over 1000 million mobile phone users by the end of 2016 [1]. One of the features which have attracted such a large population is the touch screen. Compared to the traditional keyboard input, a touch screen can provide more interaction styles and a better user experience. However, the visually impaired usually have difficulties to interact with a touch-screen-based mobile phone. In China, there are over 55 thousand of the visually

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_26

impaired per million populations until 2010 [2] which demands more attention in how the traditionally designed application could server their daily needs.

Such convenience can be further extended to include the visually impaired with certain optimization. To achieve the goal, auditory assistive function such as Google TalkBack, has been frequently applied [3]. But the slow and tedious interaction style brings a poor user experience. Therefore, researchers have attempted to apply tactile assistive technology to provide not only the basic function but a rich interaction as well. In 2016, a research group proposed a system to help visually impaired customize the 3D printed tactile maps. With the 3D printed maps, users can obtain a spatial understanding of their environment [4] which is crucial for them to navigate the environment with ease. Such tactile interface layer technology has been used to solve problems like navigation and browsing webpage for visually impaired [5–7]. However, when it comes to another frequently implemented task on the mobile phone, mathematical calculation, few research has proposed a proper solution with the tactile assistive technology. In present study, a Windows Phone calculator application and the corresponding 3D-printed covers were designed for the visually impaired to perform mathematics calculation on a touch screen mobile phone. Similar to a typical calculator, it allows users to perform such basic calculations as addition and multiplication. However, unlike typical calculators, it has a special layout and audio clues designed for the visually impaired. The application is highly customizable so that users can reassign each button with a new value according to their preferences. A pilot experiment has been implemented to test the working efficiency with blindfold users.

The rest of paper is organized as follows. Section 2 presents and discusses the prior studies on assistive technologies for the visually impaired; Sect. 3 presents our design solution. The experiment design is presented in Sect. 4, followed by experiment results. A brief analysis and discussion is presented in Sect. 5. We conclude this paper by pointing out the limitation of current research and the future path we will follow in Sect. 6.

2 Related Work

Though many researchers probed mobile interaction for the visually impaired, most of them focused on gesture-based operational style [8–10]. For instance, the *Slide Rule* [8] examined the efficiency of using sets of slide gestures to accomplish corresponding tasks. The result indicates that error rate significantly increases when applying such interaction style in a touch screen device. *Input Finger Detection* [9] explored a gesture-based way of text input, researchers compared the input efficiency with iPhone's built-in VoiceOver and positive result was received in the testing. However, when it comes to a task like performing calculation, which requires both high accuracy and speed, gesture may not be the optimal choice.

In 2015, a study used the covered phone screen to guide users when using touch screen phones. The experiment result proved the feasibility of this solution [11], such screen cover makes it easy for the blind to get access to touch screen operations. But lack of tactile cues on the cover lead to a lower operation accuracy. However, such a feasible solution still motivates our research here.

To our best knowledge, existing calculation applications for the visually impaired on the mainstream App markets are mostly voice-enabled. These applications require users to provide a voice command, and the result will be returned with an audio message. TalkingCal is a typical voice-based calculator for the visually impaired on Apple's App Store [12]. Though the auditory interaction can provide the basic function of calculations for the visually impaired, it also has limitations. The noise in the environment could affect the voice recognition. Also, the slow and tedious interaction style cannot provide a good user experience. Therefore, we believe that it is necessary to apply tactile assistive technology, specifically, a screen cover to provide a better user experience.

Based on existed calculator applications on mainstream application market and prior studies, we combined the mobile phone applications and the corresponding tactile screen covers. Such a solution can partly solve the difficulties of using touchscreen devices for the blindness and the visually impaired.

3 Our Design Solution

This section presents the design details of our solution. We first proposed an initial design of the application interface (Fig. 1) as well as the cover and implemented a simulation test. Based on the test result, we applied certain adjustments (Figs. 2 and 3). The adjusted designs had then been used in our pilot experiment.

3.1 Initial Design

Similar to the common design for the visually impaired, the application provides auditory cues. Besides, the initially designed application contains 18 buttons: 10 numeric buttons (0 to 9), a fraction point button, 4 operator buttons ("+", "–", "*", "/"), an "=" button, a "Clear" button and a "Verification" button. The layout of the buttons is shown in Fig. 1. These buttons were designed into rectangles and the size of the button

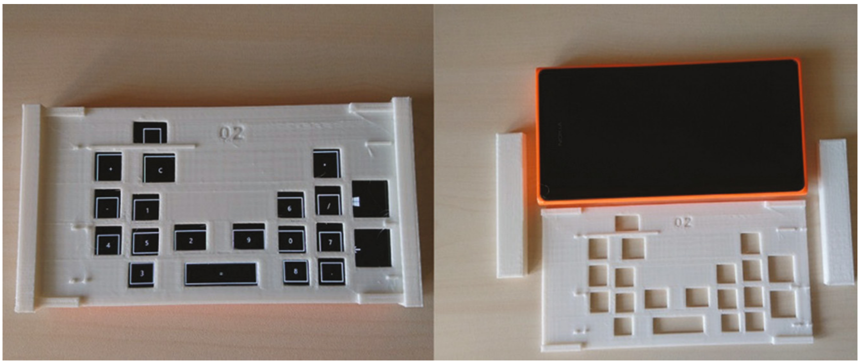


Fig. 1. Photos of initial design including application interface and 3D-printed cover during the initial design stage.

is larger if the distance between the button and edge of the mobile phone is larger. On the 3D-printed cover, there are 18 corresponding holes and other 2 holes for “Home” button and “Back” button. On the surface of the cover, there are several embossments which can remind the users the position of their fingers. Also, two 3D-printed fixtures are designed to lock the cover firmly into the phone. Such delicate design, we believe, is crucial to enhance the acceptability of our proposed solution.

When working with the initially designed solution, users will follow these steps: (1) open the application and lock the cover into their mobile phones; (2) tap the “Verification” button to ensure the cover has been locked correctly; (3) start to calculate.

Simulation Test on the Initial Design. Following previous testing protocols as in [11], we performed a simulation test with blindfolded users. During the test, we found that some users cannot work with the tactile cover precisely for the lack of tactile information as well as the layout of buttons. Also, some users mis-tapped on the “verification” button, “Home” button and “Back” button during the test.

3.2 Adjusted Designs

Based on the simulation test result, we adopted the following adjustments:

1. Rearrange the button layout into a more orderly form.
2. Add embossments near the holes to provide more tactile feedback.
3. Remove the holes of the buttons which are not necessary in the calculation (“Home” button, “Back” button and “Verification” button).

Also, we add a customization function which allows users to rearrange the numeric buttons and fraction point button based on their own preference. With the help of voice clues, the user can reset the button in the order of “.”, “0”, “1”, “2”, “3”, “4”, “5”, “6”, “7”, “8” and “9” respectively.

Design I. The entire interface is divided into two parts: on the left, there are 10 numeric buttons, a fraction point button and 2 buttons used for customization; on the right, there are 6 buttons including 4 operator buttons, one “Clear” button and a “=” button. Certain adjustments were adopted on the corresponding cover except for the layout. The eleven holes on the left side were made into circles. Also, the “Clear” button and “=” button were adjusted into rounded-rectangle. Besides, the embossments near the holes can serve as the tactile hint for the visually impaired. Users can sense the direction to the holes by putting their finger on the embossments (shown in Fig. 2, right).

Fig. 2. The first adjusted design (Design I) with embossment to provide more tactile cues.

Design II. Similar to the first adjusted design, the second design also contains 2 parts. The main differences between two cycles are the six buttons and the corresponding holes on the right side: they were designed in shape of circle and arranged in an ellipse.

Fig. 3. The second adjusted design II.

4 Pilot Experiment and Results

Apparatus. A Nokia Lumia 730 phone running Windows 8 is used to load our application. Two covers were tested to pair with our application.

Participants. A total of 9 participants (7 males and 2 females) aged from 20 to 22 participated in the pilot study. Among them, 6 used the cover of Design I (Fig. 2) and the rest three of them used the cover of Design II (Fig. 3).

Study Procedure. A two-repeated measurement experiment was designed to test the working efficiency of two adjusted solutions with blindfolded participants. During the experiment, they were required to perform 4 different sets of calculations (shown in Table 1) in the phone (Fig. 4).

Table 1. The calculation sets used in the experiment.

Set I	Set II	Set III	Set IV
$3 + 6$	$15 + 23$	$127 + 346$	$3.14 + 50.2$
$4 - 2$	$64 - 37$	$512 - 256$	$67.9 - 0.37$
7×9	89×70	780×100	8.2×2
$8 \div 2$	$99 \div 33$	$314 \div 157$	$17.9 \div 5.6$
$3 + 6$	$15 + 23$	$127 + 346$	$3.14 + 50.2$

Before doing the calculation, the blindfolded participants were required to be familiar with the covers with the help of experimenters. In this period, participants can choose to customize the layout of numeric number buttons or use the default layout. Then, they were required to do calculations read by experimenters from 4 calculation sets, using one of the covers. Task duration and error rates of each calculation were recorded. After the two-repeated measurement, they were interviewed with the following questions:

1. Did you feel nervous when you were blindfolded during the experiment?
2. Do you think it is inconvenience when using the cover to perform calculations?

Fig. 4. A participant is seen calculating with the adjusted cover.

The task duration and mistake rate will be reported in the next section. Interestingly, no participant chose to customize the button layout.

Experiment Result of Design I. The average task completion time of calculating Set I are 21.15 s (SD = 1.40) for the first measurement and 12.04 s (SD = 7.55) for the second measurement; average time costs on Set II are 25.04 s (SD = 6.45) for the first measurement and 18.01 s (SD = 3.22) for the second measurement. When calculating Set III, the average time costs are 31.65 s (SD = 2.80) for the first measurement and 26.10 s (SD = 15.59) for the second measurement. It costs 49.60 s (SD = 29.60) for the first measurement and 30.07 s (SD = 4.51) for the second measurement in average. The average mistake counts of doing all calculations with the first design are 15.83 for the first measurement and 12.83 for the second measurement. The improvement indicates that with increasing manipulation, users will be more efficient with our design. Figure 5 presented the results.

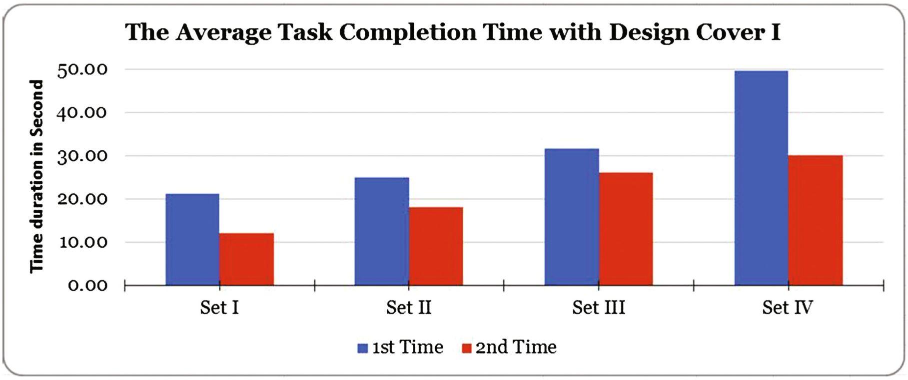


Fig. 5. The average task completion time using Cover I.

Experiment Result of Design II. 3 participants were tested with Design II. The results are shown in Fig. 6. The average time costs for them to calculate Set I are 16.04 s (SD = 12.88) for the first measurement and 11.86 s (SD = 7.41) for the second measurement. The average time costs for them to calculate Set II are 21.38 s (SD = 14.80) for the first measurement and 10.80 s (SD = 3.44) for the second measurement. The average time costs for them to calculate Set III are 25.96 s (SD = 16.66) for the first measurement and 15.08 s (SD = 6.02) for the second measurement. The average time costs for them to calculate Set IV are 32.57 s (SD = 31.94) for the first measurement and 22.37 s (SD = 13.39) for the second measurement. The average mistake counts are 13.33 for the first measurement and 8 for the second measurement.

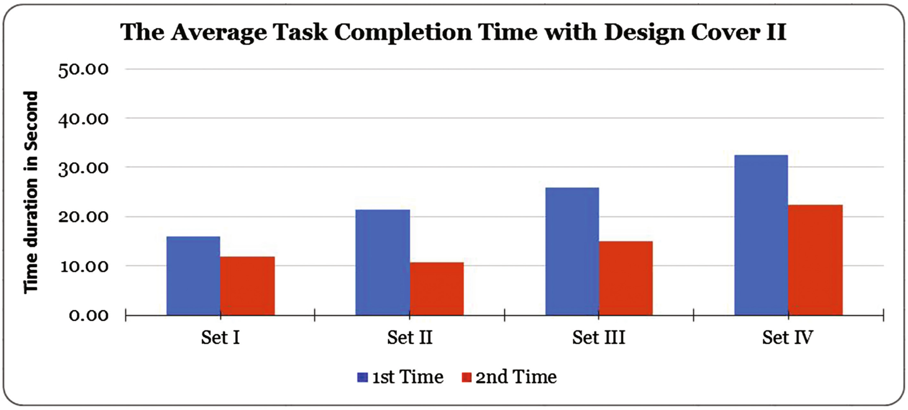


Fig. 6. The average task completion time using Cover II.

5 Brief Analysis and Discussion

Figures 5 and 6 showed the average task completion time for the participants when performing calculations with two different covers respectively. The results clearly indicated that participants spent less time in the second measurement than in the first measurement with both covers, which implies that they got more familiar to the covers after certain practices. Though the time cost becomes bigger when the calculations become more complex in both covers, the time cost using the second cover is 7.87 s less than using the first cover in the first measurement and 6.69 s less in the second measurement on an average. Also, participants made 2.50 times less mistakes when using the second cover in the first measurement and 4.83 times less in the second measurement. Therefore, we believe that the second design can provide a better user experience in terms of error rate.

Though the data shows the feasibility of our solution, during the post-test interview sessions, participants still pointed out several limitations of our adjusted solutions as well as future experiment design factors. All of them reported the mis-tap on the buttons nearby during calculation. One participant claimed that the thickness of the cover is not enough which may cause the mis-tap. She also raised the concerns that some holes are too far away from her fingers. Another participant reported that it was difficult to differentiate the holes of numeric buttons, because of the same shape. Some participants also made comments on the experiment. Two of them stressed that they felt nervous in a blindfolded status which might affect their performance. Another participant commented that she was anxious when doing the last calculation due to the long experiment time.

6 Conclusion and Future Study

In the present study, we provided two design solutions for the visually impaired to use Calculator application on the mobile phone, including the Windows Phone applications as well as corresponding 3D-printed cover. The solutions provide audio cues as well as a customization capability; both solutions are complementary in offering compensatory strategies for vision restoration to better exploit this population's excellent sensory substitution capabilities [11]. We first implemented a simulation test on our initial design and made adjustments on two final designs based on the test result. A pilot study with blindfolded users implied the feasibility of the two covers. Users spent less time and made less mistakes when doing the second measurement for both adjusted designs, which implies our designs can be mastered after a period of training. However, some participants pointed out the limitations of our designs.

The future study will focus on performing the comparative experiments to further tap into the visually impaired users' sensory substitution strategies (including auditory sensory substitution) and investigate how such strategies could inform our design accordingly.

Acknowledgements. The financial support to the last author through Student Partnering with Faculty (SpF) Program is greatly acknowledged. Thanks also go to those volunteers who participated in our study.

References

1. Statista. Number of mobile phone users in China from 2013 to 2019 (in millions). https://www.statista.com/statistics/233291/forecast-of-mobile-phone-users-in-china/
2. WHO. http://www.who.int/blindness/publications/globaldata
3. Google Inc. - 15 December 2016 - Everyone: Google TalkBack - Android Apps on Google Play. https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback
4. Taylor, B., Dey, A., Siewiorek, D., Smailagic, A.: Customizable 3D printed tactile maps as interactive overlays. In: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS 2016 (2016)
5. Tang, H., Beebe, D.: An oral tactile interface for blind navigation. IEEE Trans. Neural Syst. Rehabil. Eng. **14**, 116–123 (2006)
6. Rotard, M., Knödler, S., Ertl, T.: A tactile web browser for the visually disabled. In: Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia - HYPERTEXT 2005 (2005)
7. Götzelmann, T.: LucentMaps. In: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS 2016 (2016)
8. Shaun, K.K.: Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In: ASSETS 2008 (2008)
9. Duarte, K., Cecilio, J., Furtado, P.: Overview of assistive technologies for the blind: navigation and shopping. In: 13th International Conference on Control Automation Robotics and Vision (ICARCV) (2014)

10. Shiri, A.: Input finger detection for nonvisual touch screen text entry in Perkinput. In: Graphics Interface Conference (2012)
11. Pinata, W., Tiffany, Y.T.: Sensory substitution to enable the visually impaired to play an affordable wearable mobile game. UbiComp/ISWC 2015 (2015)
12. TalkingCal. https://itunes.apple.com/us/app/talkingcal/id545866051?mt=8

Impact of Cognitive Learning Disorders on Accessing Online Resources

Alexander Cadzow(✉)

C3L, Sawbrideworth, UK
alex@cadzow.com

Abstract. The present document assesses the impact of learning cognitive disorders on accessing online resources. This is a thought exercise study and examines how technology, represented by online resources, is made available without discriminating against those with learning cognitive disorders. The learning impact has been assessed by the Human Factors based methods of Soft Systems Methodology (SSM), Usability Engineering (UE) and Ontology Sketch Modelling (OSM). The major finding of the study is that whilst there are tools that make access to many resources more equitable, there is still limited understanding of how those with cognitive disorders assimilate information. In addressing this aspect, the present document has examined the wider role of semantic web technology that by introducing codification of context and pragmatics can improve the presentation of knowledge such that those with cognitive learning disorders can achieve non-discriminatory access to online resources.

Keywords: UX and usability · User experience

1 Introduction

The present document assesses the impact of learning cognitive disorders on accessing online resources. This is a thought exercise study and examines how technology, represented by online resources, is made available without discriminating against those with learning cognitive disorders. The learning impact has been assessed by the Human Factors based methods Soft Systems Methodology (SSM), Usability Engineering (UE) and Ontology Sketch Modelling (OSM). The major finding of the study is that whilst there are tools that make access to many resources more equitable, there is still limited understanding of how those with cognitive disorders assimilate information. In addressing this aspect, the present document has examined the wider role of semantic web technology that by introducing codification of context and pragmatics can improve the presentation of knowledge such that those with cognitive learning disorders can achieve non-discriminatory access to online resources.

2 Background Information

This project concerns accessibility of online resources which refers to the degree to which an interactive product, i.e. the online resource, is accessible by as large a proportion of the accessing population as possible. There is a need to strike a balance in

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_27

the design of the online resource and the means to access it between usability and accessibility for all people regardless of whether they a disability or not. This often requires a need for a site or system to allow for assistive technologies [1, p. 18] The aides focus on reading, speaking and listening to enhance accessibility. Often these are third party tools which are applied to sites and systems instead of being built in. Since there needs to be a balance between usability and accessibility often these tools are not built into sites or systems so there are sometimes issues of compatibility [1, pp. 82–83]. This project considers accessibility and usability when assessing the impact of learning cognitive disorder on accessing online resources.

Furthermore, as stated in the EU's Digital Single Market initiative [2]: "Web accessibility is not just a question of technical standards and of web architecture and design. It is not a concern for web developers only, but also a question of political will and of moral obligation now enshrined in the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD). Article 9 of the Convention requires that appropriate measures are taken to ensure access for persons with disabilities, on an equal basis with others, to inter alia information and communication technologies, including the Internet" [3].

It can be asserted therefore that universal web access is not a "nice to have" but should be defined as an essential element of non-discriminatory behaviour. The concern addressed in the present document is how to address accessibility for cognitive as opposed to physical disability.

3 Definitions

- *Cognitive Disorder:* a category of mental health disorders that primarily affect learning, memory, perception, and problem solving, and include amnesia, dementia, and delirium
- *Dyslexia:* a learning difficulty that primarily affects the skills involved in accurate and fluent word reading and spelling
- *Dysgraphia:* neurological condition that impairs writing and memory processing [4]
- *Semantic Web:* a set of formats and protocols for the interchange of data (where on the Web there is only an interchange of documents)
- *Usability:* effectiveness, efficiency, and satisfaction with which specified users achieve specified goals under certain environments' [5]
- *Accessibility*: extent to which products, systems, services, environments and facilities can be used by people from a population with the widest range of characteristics and capabilities, to achieve a specified goal in a specified context of use [6].

4 Approaches

4.1 Soft Systems Methodology

Systems-based methodology for tackling real-world problems in which known-to-be-desirable ends cannot be taken as given. Soft systems methodology is based upon a phenomenological stance (applying the philosophical study of the structures of experience and consciousness to problem-solving) [7, p. 318]. When used to examine the impact cognitive learning disorders have when accessing online resources, it aims to show how users interact with the system being analysed and the rationales for decisions to use aids tools and how they are used themselves with the system.

Rich Picture is the expression of a problem situation compiled by an investigator, often by examining elements of structure, elements of the process, and the situation climate [7, p. 317]. How people with learning cognitive disorders use mySearch and how it works (Fig. 1).

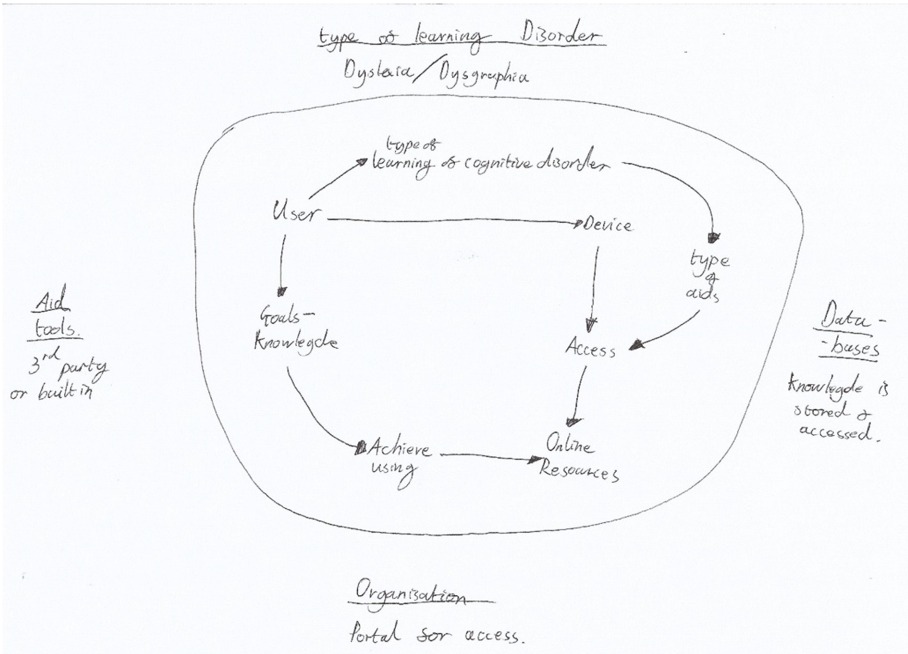


Fig. 1. Rich picture.

Root Definitions – Customer Actor Transformation Worldview Owners Environment. A concise, tightly constructed description of a human activity which states what the system is; what it does is then elaborated in a conceptual model which is built based on the definition [7, p. 317]. The CATWOE method is used here to develop the root definitions and their conceptual models.

Conceptual Models. They allow for a systemic account of a human activity system. These models presented here contain the minimum necessary activities for the system to be the one named in the root definition [7, p. 313]. There are two root definitions paired with two conceptual models presented here in this paper looking at how a user with a cognitive learning disorder would use the site or system. The first root definition and conceptual model show the use of aid tools with a site and the second shows User-system interaction (Tables 1 and 2; Figs. 2 and 3).

Aid Tools.

Table 1. CATWOE method.

Domain	Description	Justification
Customer	User with learning disorder	The User being described has the cognitive learning disorder
Actor	Aid tools	The devices or software used by the user
Transformation	Applying aids to a system	How improved understanding would be gained from the site
Worldview	Apply aids to an article to read processing of an article	How the user is using the site or system
Owners	User of the aids	These are the people who are going to use the aids with the site or system
Environment	Are the aids compatible with the system?	The user needs to know that their chosen aid tools with work with the site or system

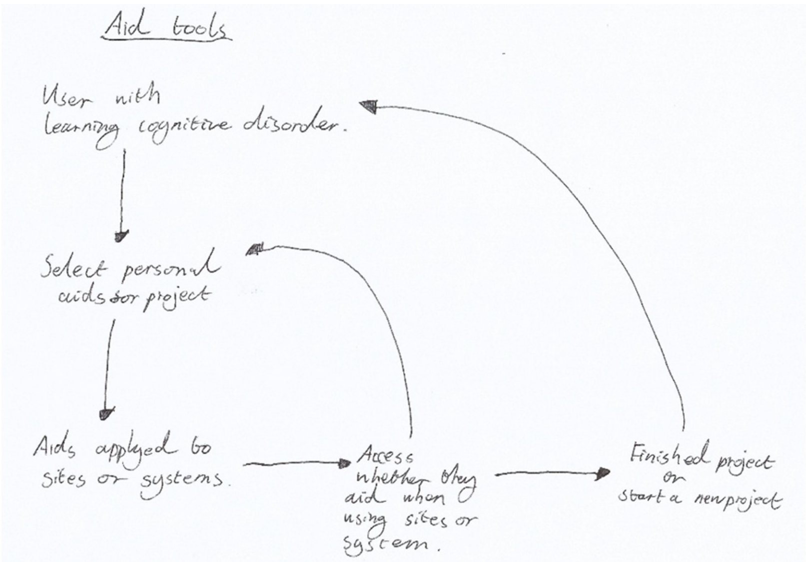


Fig. 2. Conceptual model of aid tools.

User-System Interaction.

Table 2. CATWOE method.

Domain	Description	Justification
Customer	User accessing the site or system	How the task will start
Actor	System or site architecture	What the User will interact with
Transformation	Using aids to with the results	To gain a better understanding of the results
Worldview	People with cognitive learning disorders using aids with the site or systems	Interaction between the User and the System
Owners	System or site administration	Responsible for ensuring aids work with the site or system
Environment	Compatibility of aid tools	The application of the 3rd party aid tools to user interfaces of the site or system may not work as desired

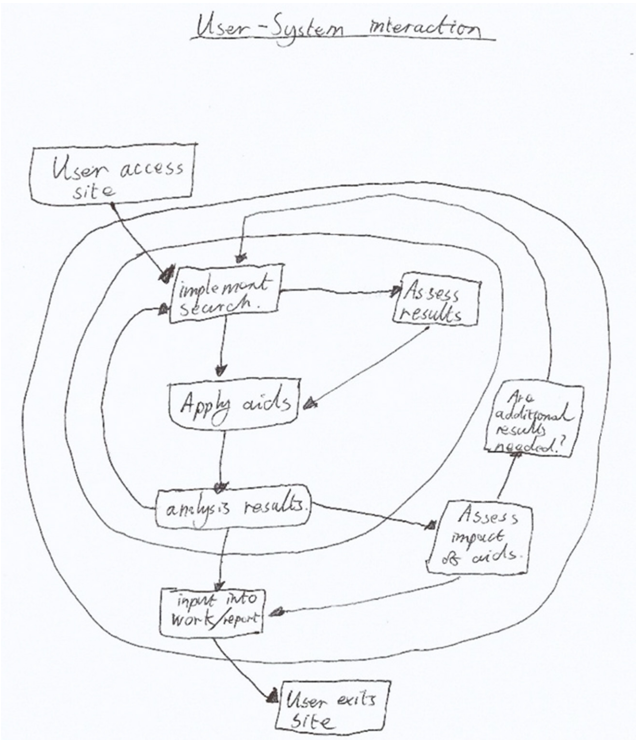


Fig. 3. Conceptual model of user-system interaction.

Comparison with the Real World. The user-system interaction shown here are generalised overviews about how the users use the aids to better their processing of the articles their searches would find. Though there are problems with these generalised overviews is that people who have cognitive learning disorders or learning disabilities are that the people have different levels of them when diagnosed and each person develops their own unique coping mechanism.

The lessons learned from SSM here is though while there are difficulties with generalised models it shows the process would be the same for all people with learning difficulties as they use online resources. This means the decision process can be studied and analysed to show how they can be improved when users interact with the system or site.

The findings and work process of the first approach of SMM will be used to lead and to influence the second approach. This is to allow a consistent analysis of accessing the impact of cognitive learning disorders when accessing online resources.

Usability Engineering (Inspection/Evaluation). Usability engineering is an area of which inspects and evaluates human-computer interactions and human-computer interfaces (HCI). Here Personas paired with relevant cognitive walkthroughs will be used to inspect and evaluate the HCI.

4.2 Persona

Each Persona is a final year university students who is a twenty-one years' old who each have a cognitive learning disorder. They are undertaking a final year dissertation. This scenario was chosen since it would require the use of research and articles which would entail the use of university resources to find these sources.

Dysgraphia – Aids for Writing and Memory Processing.

Persona Details. The final year student is twenty-one years' old, female, who is undertaking an anthropology course. She requires the aids in taking notes and observations. Finally, these aids would be used the write up her dissertation.

- Voice recognition software. To allow for natural voice searching on sites and systems.
- Mind map software. To pull and organise data from the sites and systems.
- Audio notes. A digital recorder or a smart phone or a digital program.
- Reference and citation software. Endnote. To ensure the correct use of referencing but also to allow clear organisation. Pulled from their source.

Dyslexia – Aids for Spelling, Note Taking and Proofreading.

Persona Details. The final year student is twenty-one years` old, male, who is undertaking a Forensic Science course. He uses the aides to double check his aids to double check his work and ensure it is grammatically correct. Also by using the screen reader aid in processing audibly and visually sources he uses and allows him to better proofread his own work.

- Dictionary, physical or digital.
- Note taker. Use of a programme such as OneNote, Evernote, Windows Journal and Google Keep. Enables thought and ideas to be easily tracked. Data can be pulled directly from sites or systems.
- Screen reader. To to be used one the accessed sites and systems. Also, can be used on articles and sources.
- Reference and citation software. Endnote. To ensure the correct use of referencing but also to allow clear organisation. Pulled from their source

Cognitive Walk-Through. Acognitive walkthrough shows how a person with one of these learning disabilities would hypothetical use the system or site being accessed. The Persona will have a task list with the action sequence that details the specific task flow from beginning to end.

Dysgraphia Task List.

1. Access myBU online search resource mySearch.
   1.1. Enable 3rd party voice recognition software to facilitate natural voice searching.
2. Enter the Search term for a topic.
3. Go through the search results.
   3.1. Note any chosen articles or journals found into categories in the mind mapping software.
   3.2. Create audio notes for each chosen articles or journal.
   3.3. Use Endnote to create a list of references using the BU-Harvard referencing system.
4. Either enter a new search term or exit mySearch.

Dyslexia Task List.

1. Access myBU online search resource mySearch.
   1.1. Ensure the physical or digital dictionary is nearby or open to allow for unfamiliar words or terms to be looked-up.
2. Enter the search term for the chosen topic.
3. Go through the search results.
   3.1. Use the screen reader to go through the chosen journals or articles.
   3.2. Enter sections of note from these journals or articles into Notetaker software.
   3.3. Use Endnote to create a list of references using the BU-Harvard referencing system.
4. Either enter a new search term or exit mySearch.

The results of applying these tasks would be that each persona would have a selection of notes and their relevant references. These can then be inserted into the main body of their dissertation with minimal fuss.

The key lessons learned from working the process of Usability Engineering is the lack of built-in support in mySearch for Users with cognitive learning disorders to aid them when they are undertaking searches for information. The use of personas to inspect mySearch showed that is not a big step-up to enable aid tools for Users with cognitive learning disorders. Since the tools are not specific for people with learning disabilities though they are the users most likely to take advantage of them. From this evaluation, the 3rd party aid tools generally work well when using the mySearch and any future developing could suggest versions of these tools being built into mySearch.

The findings from the usability engineering will influence the direction and pattern when developing of the OSM. The process of task lists will help in direction of the pathways when developing the OSM.

4.3 Ontology Sketch Modelling (OSM)

OSM when paired with usability and accessibility testing, this will bring the concepts and links discussed in the previous approaches into a single structure and flow. The WAVE tool will provide a baseline from which to work with when undertaking the OSM [8]. By using the MySearch page [9] with the WAVE tool as a baseline can be determined in order to see problems with accessibility which is one of the concerns of this impact. See next page for the image showing the layout of the web page in question.

Aside mySearch is built on technology and services from EBSCO host and its Research Databases. It provides subscriptions and access to different publishers of journals, magazines and books. So, an organisation such as Bournemouth University can gain access through a single portal instead should have individual subscriptions from those publishers. A summary of what the WAVE tool detected the following: three Errors, three hundred and thirty-five Alerts, seventy-four Features, twenty-six Structural Elements, one-hundred and fifty HTML5 and ARIA and fourteen Contrast Errors. As can be seen from the figure these are in the end only minor issues which would easily be fixed and generally will not have a negative effect on a user with cognitive learning disabilities. It also illustrates that even though mySearch is powerful search engine tool it is still accessible to the widest possible range of users. It also helps to serve to illustrate when developing the OSM and the links between the semantic web how the options and data being pulled together into a single user interface (Fig. 4).

For illustrating the OSM the software XMind has been used to show a concept of how the links and connections of users with learning disabilities, access the knowledge and how the use of aid tools to improve accessibility and usability come together.

The OSM is centred on the key topic of accessibility and usability since this the natural conclusion of the SSM and the UE. This then comes back to the issue of guidelines and standards for sites and systems to be designed for Users with cognitive learning disorders. Though in the end, it is up to the organisations that provide these sites and systems to follow the Web standards and guidelines which can be tested up tools such as WAVE. Since Users will use their own personal selection of aid tools there is a need for seamless integration because largely people except programs to work straight away as plug and play are considered a key cornerstone of universal accessibility and usability. Though there is a need to remember that the use of aid tools is not specific to cognitive learning disabilities since anyone can use them but this does not

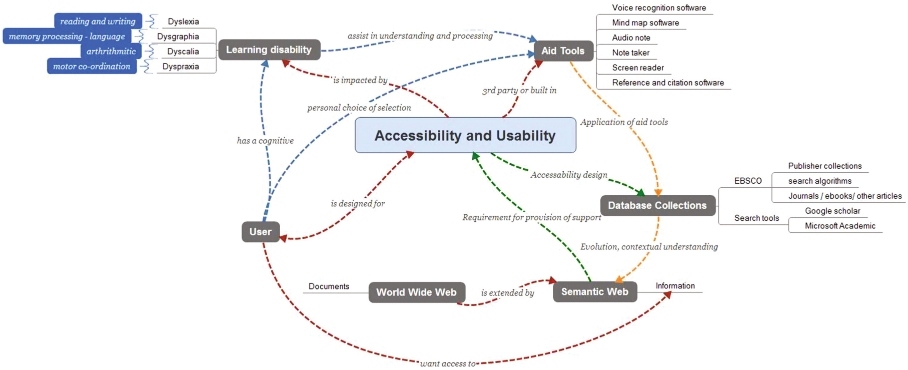


Fig. 4. OSM

undermine their importance in them being able to work without issues with any site or system. Since this assessment concerns the accessing of mySearch which use the services of EBSCO which collate published collections of articles, eBooks and articles the OSM focuses on the issues aid tools working with mySearch. The design of mySearch as shown by the WAVE tool has an acceptable usability and accessibility rating. So, the impact of cognitive learning disorders on accessing online resources through mySearch is minimal which leads then into what improvements could be applied which in the OSM is provided by the Semantic Web.

This in the future there could be developments which will hopefully see mySearch extended by the Semantic Web since it should be able to provide better contextual support when looking up information for topics which could mean better-personalised searches and more focused results. Also in theory, if the kind of aid tools discussed in this assignment in the future could be built into or more integrated into sites or systems such as mySearch as it should allow people with or without learning disabilities the option of greater accessibility by providing Users more options to process and access the desired information.

5 Discussion

The discussion will examine a wider role of semantic web technology on how humans understand and find or catalogue knowledge. The semantic web concerns a Web of data while the traditional internet is the Web of documents. There is a need, regardless of the complexity of the back-end system, for all people no matter their ability or any possible impairment (physical/mental) to interact with the user interface and the underlying content to be accessible to the widest extent and for all content to be accessible to the same degree (no content should be inaccessible through impairment).

The Ontological Sketch Model (OSM) presented in this paper is structured as an informal representation of the essential underlying structure of a system forming a basis for usability assessment. The primary aim is to develop an approach that is usable and that yields useful results. Studies have found that OSM allows a way of thinking about the usability of a system, including capturing important cognitive dimensions [10]. This relates to the relationship shown in the OSM presented earlier aims to the actions of a user and those accessing sites or systems of a domain. The OSM shows in the future the need for semantic web support which could potentially evolve the services such as EBSCO, Google Scholar and Microsoft Academic into better knowledge engineering tools. A published paper has shown that cognitive support tools can be used to leverage human abilities to increase human understanding and cognition of challenging problems [11]. This line of thinking could be applied to our OSM since the application of cognitive support tools also have the potential to improve universal accessibility and usability.

The nearly exponential growth in the availability of published digital information creates issues of how it can be usefully accessed. Conventional data-centric models such as those using in Relational Database modelling require complex normalisation and indexing to be accessed quickly. The underlying problem is that the information is not structured and cannot be readily indexed to allow for search. Adding keywords extends the data and if content creators do not take the time to provide keywords or phrases there is no guarantee that existing search engines will catalogue web pages. Furthermore, where search engines such as those from Google work it is based on page based keyword analysis and link indexing, furthermore indexing results can be skewed by buying index words and adding things like Analytics meta-data. Search results using existing models are not reliable or realistic given the commercial imperatives for site indexing. Solutions have been proposed in similar situations using concepts of a formal ontology paired with semantic process modelling. They concluded that formalisation of the semantics in conjunction with the use of inference engines allows the improvement of query functionalities [12]. These possible solutions could allow for better search methods. This connects to a case study which looked at digital libraries in the knowledge era. The purpose was to show the expression and the exploitation of humanity's collective knowledge. They found that Semantic Web technologies and knowledge management technologies expand the frontiers of knowledge representation and sharing [13]. Though these solutions are about the underlying architecture that facilitates the ability to search for topics. Though ideally people should see no difference in the user interface while experience more accurate and precise results. It relates back to the issue of the impact of learning cognitive disorder on accessing online resources which any future improvements would also need to consider of accessibility and usability which should remain consistent regardless how improved the underlying search architecture is.

6 Conclusion

From this thought exercise, there are a few key lessons learned from examining the impact of learning cognitive disorder on accessing online resources. The impact can be minimal since guidelines and web standards if followed ensure sites and systems have adequate accessibility and usability design or built into the interfaces from the beginning. Also, it could be argued for greater built-in support to allow for less reliance on 3rd aid tools since while that would give benefit to users with cognitive learning disorders it would also benefit all users since it is natural to take advantage of tools which can benefit the overall work process. Finally, a future study should cover a more in-depth assessment which could involve a blind test of tasks with a mixture of people with and without learning disabilities to ensure a more in-depth examination of the impact of learning cognitive disorders on accessing online resources.

References

1. Preece, J., Sharp, H., Rogers, Y.: Interaction Design – Beyond Human-Computer Interaction, 4th edn. Wiley, Glasgow (2015)
2. European Commission, Proposal for a Directive of the European Parliament and of the Council on the accessibility of public sector bodies' websites. https://ec.europa.eu/digital-single-market/en/news/proposal-directive-european-parliament-and-council-accessibility-public-sector-bodies-websites. Accessed 05 Mar 2016
3. United Nations, Convention on the Rights of Persons with Disabilities. http://www.un.org/disabilities/convention/conventionfull.shtml. Accessed 05 Mar 2016
4. Dysgraphia.org.uk/. http://dysgraphia.org.uk
5. W3, Usability – ISO 9241 Definition. https://www.w3.org/2002/Talks/0104-usabilityprocess/slide3-0.html. Accessed 10 Apr 2016
6. Accessibility Week – Design for all. https://www.cencenelec.eu/news/brief_news/Pages/TN-2014-018.aspx
7. Checkland, P.: Systems Thinking, Systems Practice – Includes a 30-Year Retrospective. Wiley, Chichester (2005)
8. WAVE Web Accessbility Evaluation Tool. http://wave.webaim.org/
9. EBSCO Publishing. http://eds.a.ebscohost.com/eds/search/advanced?sid=e1bdc0ef-1c70-4c60-96a3-dc54dfd65687%40sessionmgr4004&vid=0&hid=4211&preview=false
10. Blandford, A., Green, T.R.G.: OSM: an ontology-based approach to usability evaluation. In: Workshop on Representations. Queen Mary College, London (1997)
11. Ernst, N.A., Storey, M.-A., Allen, P.: Cognitive support for ontology modelling. Int. J. Hum. Comput. Stud. **62**, 553–577 (2005)
12. Thomas, O., Fellmann, M.: Semantic process modelling – design and implementation of an ontology-based representation of business processes. Bus. Inf. Syst. Eng. **1**, 438 (2009)
13. Lytas, M., Sicilia, M.-A., Davies, J., Kashyap, V.: Digital libraries in the knowledge era – knowledge management and Semantic Web technologies. Libr. Manag. **26**(4/5), 170–175 (2005)

A Study on the Interface of Dashboard Camera Designed for the Elderly

Shu Yao Chang[1], Shuo-Fang Liu[1(✉)], Ming-Chyuan Lin[2], and Ming-Hong Wang[1]

[1] Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan
liusf@mail.ncku.edu.tw
[2] Department of Creative Product Design and Management, Far East University, Tainan, Taiwan

Abstract. One of the consequences of the global aging is an increase in the driving by the elderly. Due to their slower perception, processing and cognition, elder drivers face greater risk of traffic accidents. This study is an analysis of the color collocation and text of existing dashboard cameras. After analyzing the results of the questionnaire and interviews concerning the visibility and load of the subjects through the cognition questionnaire, this study established an experimental interface to simulate the situation where an elder driver is reading the dashboard camera in driving. Moreover, we conducted a simulation with changeable color collocation which enabled the subjects to select the best color collocation for the interface of the dashboard camera. In this way, the driver can save energy and concentration in reading the dashboard in driving and pay more attention to road safety, thus reducing the risk of traffic accidents.

Keywords: Elder drivers · Driving behavior · Road user safety · Driving record

1 Introduction

The issue of an aging society has attracted global attention. Because of the natural physiological decline of the elderly, we have to pay attention to the safety of the living conditions for them. According to the statistics about "the number of people with a driving license" released by the Ministry of Transportation, the proportion of elder people (aged over 60) with a driving license rose from 9.08% in 2007 to 14.98% in 2013. It is obvious that the elder people with a driving license will account for a larger proportion in the future society [1]. According to the results of previous studies, driving can enhance elder people's sense of independence and improve their life [2]. The technologies for the elderly can be divided into four types, namely, travel (including transportation and action), communication (including information processing and understanding), household (space design), and household health care. To ensure safer driving for elder drivers, we need to find out a solution to protect their driving right and safety. In the future, many smart mobile devices will be designed on the basis of the touch and vision of elder users [3], and be operated through a touch screen [4].

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_28

However, smart products requires more attention from elder drivers in driving, which indirectly increases the safety risk. With focus on the design of the interface of dashboard camera, this study aims to design an interface suitable for elder drivers, so as to enhance their driving safety.

The global aging has resulted in an increasing number of drivers across the world. Because of the less concentration and slower reaction caused by organ decline, elder drivers face greater risk in driving [5]. And their limited acquisition of knowledge as well as slow information absorption and understanding often lead to traffic accidents [6]. Lightening the visual and cognitive load of elder drivers in driving can effectively reduce the risk and contribute to safe driving. In case of a dispute over traffic accidents, the dashboard camera is often taken as an effective evidence to solve the problem. During the driving, drivers would closely follow the warnings display on the dashboard camera; as a result, they have to pay extra attention to the warnings. Through an optimal adjustment of the interface of dashboard camera, we can save some energy in reading the interface and pay more attention to road safety, which will reduce the risk in driving [7].

2 Methods

2.1 Planning of Experiment Samples

In terms of interface design, young people tend to add more interesting features to the device, while the elderly place more emphasis on "perceived usefulness" [8]. For the elderly, a poorly-designed interface features much distraction [9], repetitious coverage or disorder [10, 11]. Therefore, most of the subjects of this study were the middle-aged and the elderly.

According to the Japanese Old Age Association, a person aged 75 or above is regarded as an elderly one, but the subjects of this study aged from 50 to 70, which is beneficial for the extension and continuity of this study. For a higher accuracy of users' perceived usefulness, the middle-aged and elderly people who drove at least once a week [12] and had used the products relevant to dashboard camera were chosen as the subjects.

2.2 Data Collection

In the preliminary experiment, the focus group was adopted to collect the data about the use of the widely-adopted dashboard cameras through operation and experience. Then, the data about the user experience were gathered, and the 1,300 sets of data about the color, font, size and leading of all the images on the dashboard cameras were analyzed. A 4-inch (6.15 cm in width and 8.2 cm in length) screen was used as the interface of the experiment.

According to the statistical analysis of the background and surface colors of the dashboard camera [13], we found that gray (K80) and gray (K40) were the two colors which were most frequently used as the background color of an interface. Hence, they were applied as the color collocation prototypes of the background color of the interface. The text color was selected according to the statistics of the previous studies. The top 3 text colors included white (58%), black (K100) (29.2%) and blue (6.9%), and the top 4 background colors were white (19.1%), dark cyan (13.9%), gray (K20) (12.6%) and dark gray (K40) (3.8%). The above text and background colors were combined with an orthogonal array to form the demands for color collocation and text color, so as to facilitate the design of the questionnaire.

2.3 Design of Questionnaire

As for the color collocation prototypes in the questionnaire, 24 color collocation combination were used for the use evaluation experiment. The evaluation experiment scale comprised five dimensions: learnability, effectiveness, afford, visibility and favorability.

Meanwhile, the Likert 5-point scale divided each evaluation attribute into five levels, and the interface with the most suitable colors and color collocation for the subjects was recorded to ensure that the chosen interface met the user's demand. The scores were rounded up to the second number after the decimal point of the total averages, and the scores of all attributes were obtained for reference (Fig. 1).

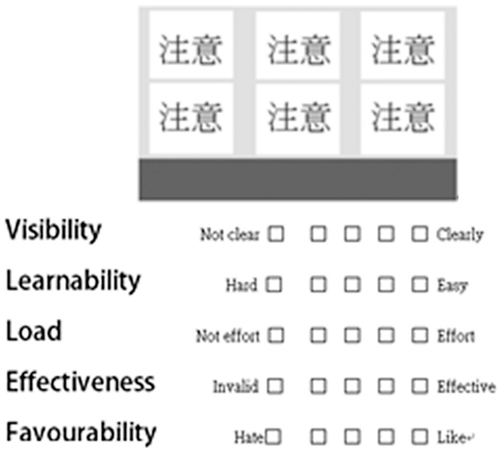


Fig. 1. The questionnaire included the image color collocations of a simulated inter-face of dashboard camera, and it also involved the five measurement standards of a visibility questionnaire: visibility, learnability, afford, effectiveness and favorability. Also, the Likert 5-point scale was employed to quantify the selection of the subjects.

2.4 Making of the Experiment Program

According to the results obtained in the previous stages, the subjects were in favor of the interface featuring black text with a white background or blue text with a white background as well as framed with white or blue. According to [14], the top 3 colors favored by the elderly were red, blue and green. In the elderly were the least able to recognize green or blue patterns due to a declining eyesight [15]. Hence, red was chosen as a reference color for the frame.

As for the making of the experiment interface, the JAVA GUI development interface by NETBEANS was adopted to construct the image of a simulated dashboard camera interface. The subjects could click the color menu to adjust the text and background colors of the tested image as well as the color of the frame to achieve a satisfactory interface. In the experiment, the color collocation of the interface could be repeatedly adjusted according to the need of the subjects (Fig. 2).

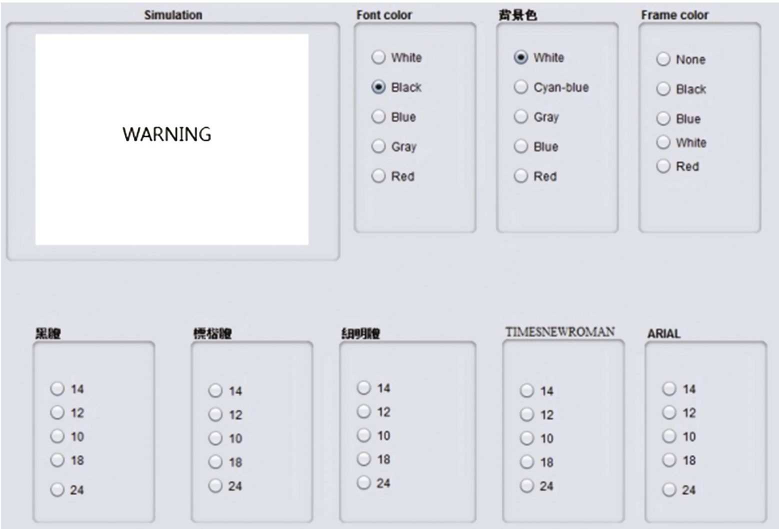


Fig. 2. The program included the display frame of the simulated dashboard camera interface on the upper left of the image as well as the options of text and background colors which could be adjusted. Moreover, there were options for the color of the interface frame.

2.5 Implementation of the Experiment Program

The program implementation comprised two stages. In the first stage, the subjects adjusted the user interface for an optimal one. In the second stage, the subjects made a full-screen projection of the tested image on the laptop to simulate the actual driving, and the image was presented at an angle of 45° in the corner of the subjects' vision. If the subjects didn't like the color or wanted to change the color in the test, they could change the setting to get a satisfactory one and then recorded the reset color. Then, the experiment came to an end.

3 Result

According to the test results obtained in the first stage, the total average of the experiment interface featuring blue text and white background was the highest, 4.95 points. The top 5 were "blue text and white background" (4.95), "black text and white back-ground" (4.82), "black text and gray background" (4.53), "gray text and blue back-ground" (4.36) and "white text and blue background" (4.19). The same ranking could be found in visibility and learnability. But in terms of afford and effectiveness, the ranking was as follows: "blue text and white background" (5.00 points), "black text and white background" (4.87 points), "black text and white background" (4.67 points), "gray text and blue background" (4.33), and "white text and blue background" (4.20), and "blue text and dark gray background" (4.20). In respect of favorability, the ranking was as follows: "blue text and white background" (4.80), "black text and white background" (4.73), "gray text and blue background" (4.27), and "white text and blue background" (4.01).

After the questionnaire, the subjects were asked about their favorite color collocation, and the results were as follows: black text (0.53), blue text (0.33), white text (0.14), white background (0.93) and cyan background (0.07). According to the results of the experiment interface based on these data, the text colors adopted by the subjects were black (0.53), white (0.2) and blue (0.27); the selected background colors were white (0.8) and cyan (0.2). As for the design of the interface of the dashboard camera for the elderly, the experiment results showed that the best design featuring black text and white background (Figs. 3, 4, 5, 6, 7, 8 and 9).

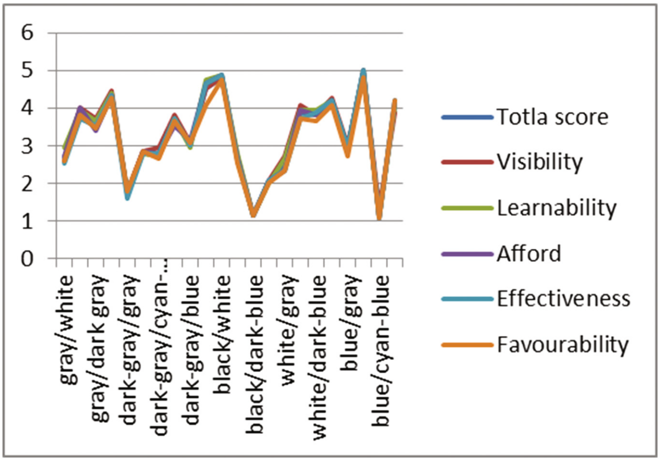


Fig. 3. The X axis was the combination of font color, and the ranking was gray text and white background. The Y axis was the quantified collocation. On average, "black text and white background" and "blue text and white background" were the highest, while "black text and cyan background" and "blue text and cyan background" were the lowest.

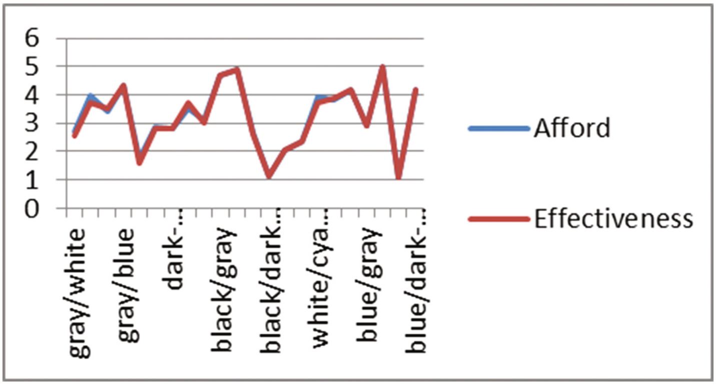


Fig. 4. T As for the overall comparison of afford and effectiveness, "white text and blue background" and "gray text and blue background" shared the same afford and effectiveness.

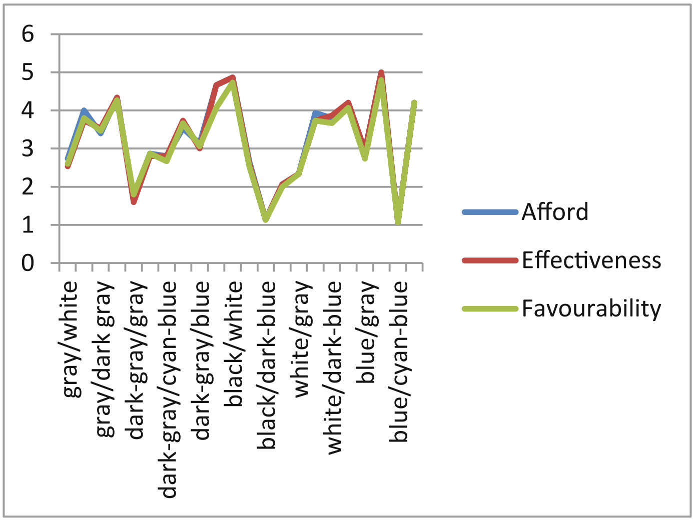


Fig. 5. In the comparison among favorability, load and effectiveness, the subjects preferred "blue text and gray background" to "black text and gray background" and "white text and blue background" which had a higher rank.

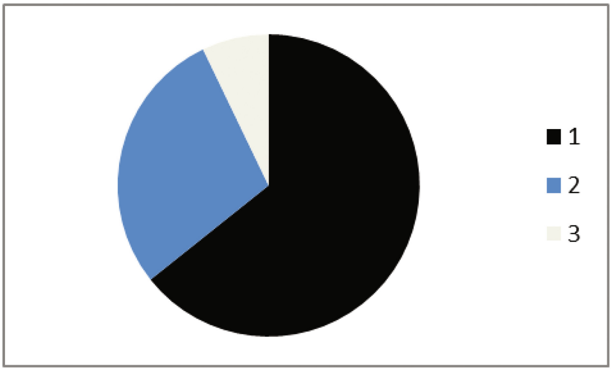


Fig. 6. In terms of text color, no experiment has been done, and the subjects were asked to select their favorite color. In the pie, "1" represents "black"; "2", "blue"; "3", "white". "Black" accounted for 60%.

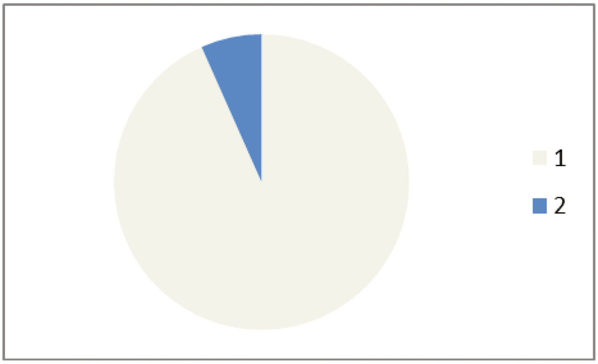


Fig. 7. As for background color, most of the subjects chose "white" as the background color, accounting for 93%.

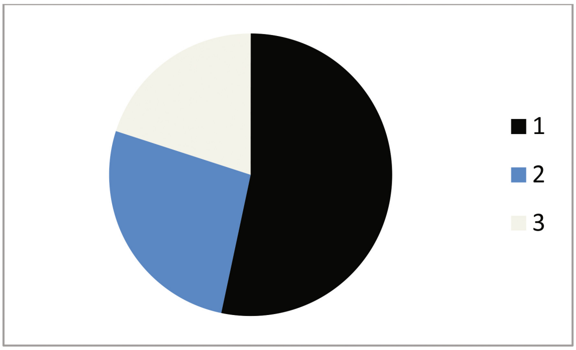


Fig. 8. After the experience, there was a significant change to the selection. The proportion of "black text" declined to 53%; "black text" was down to 27%; "white text" rose to 20%.

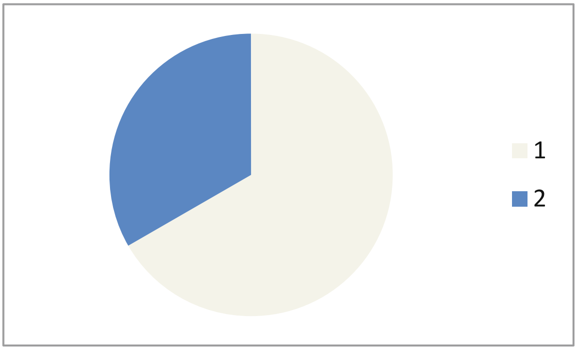


Fig. 9. In respect of background color, the proportion of “white background” was down from 93% to 80%, while “blue background” was up from 7% to 20%.

4 Discussion

This study aims to optimize the interface of the dashboard camera for the elderly, so that elder drivers can save some energy for reading the dashboard camera during the driving. Currently, there are too many color collocations on the interface of dashboard cameras, and drivers have to spend much time following the directions presented on the interface. In the second stage of the experiment, a few subjects made the color collocation according to the one on the interface. But after the simulation, they decided to change the previous color collocation to get the most energy-saving one. This indicates that the color collocation of existing dashboard cameras requires extra energy of the users and thus increases the risk in driving. After the adjustment, the optimal color collocation could lighten their visual load and enhance the driving safety and thus reduce the possibility of traffic accident.

Although the combination of “blue and white” ranked the highest in the first stage, the ones of “blue and white” and “black and gray” were preferred by the middle-aged and the elderly in the second stage who were in favor of simple and intuitive color collocation. All of the subjects had used a dashboard camera or a GPS, so they were very familiar with the product. Moreover, nearly all of them have drove for over 10 years. In the use of dashboard camera, the subjects expected that the sections worthy of attention would be presented in a simple and clear way and that the other sections should be made concise. This is consistent with [9] which argued that the elderly were less willing to see overlapped images than the young and that they preferred to simple and striking color collocation such as “black text and white background”.

The text environment of this study was not fully simulated, so it was impossible to simulate the emergencies in actual driving. In the experiment, the subjects merely looked to the interface. Second, such parameters as climate and light were not taken into consideration, and the influence of bright colors in different light was also neglected [16]. Third, the number of the samples was small, so other potential color collocations were not proposed. Hence, future studies can collect more data and consider the color collocation of hand-held smart devices. The elderly are reluctant to accept a digital product with a less friendly interface or color collocation, so the follow-up parts of this study can be taken as reference for the interface setting of hand-held smart devices for the elderly.

Acknowledgements. The authors would like to appreciate that this research is financially supported by MOST (Ministry of Science and Technology) at Taiwan, R.O.C, under the project number: MOST 105-2221-E-209-004-All invited subjects and designers are also appreciated very much.

References

1. Eby, D., Molnar, L.J., Kartje, P.S.: Maintaining Safe Mobility in an Aging Society. CRC Press, Boca Raton (2009)
2. O'Neill, D.: Towards an understanding of the full spectrum of travel-related injuries among older people. J. Transp. Health **3**(1), 21–25 (2016)
3. Bouma, H., Graafmans, J.A.M.: Gerontechnology. IOS Press, Amsterdam (1992)
4. Edmarlon, G., de Andrade, S.M., González, A.D., Mesas, A.E.: Professional experience and traffic accidents/near-miss accidents among truck drivers. Accid. Anal. Prev. **95**, 299–304 (2016)
5. Road Traffic Safety Portal Site. http://168.motc.gov.tw
6. Zajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid information services for distributed resource sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181–184. IEEE Press, New York (2001)
7. Rice, M., Alm, N.: Designing new interfaces for digital interactive television usable by older adults. Comput. Entertain. **6**(1), 6 (2008). doi:10.1145/1350843.1350849
8. Musselwhite, C.B.A.: Assessment of computer-based training packages to improve the safety of older people's driver behavior. Transp. Plan. Technol. (2017). doi:10.1080/03081060.2016.1238571
9. Thompsona, K.R., Johnsonb, A.M., Emersona, J.L., Dawsonb, J.D., Boer, E.R., Rizzoa, M.: Distracted driving in elderly and middle-aged drivers. Accid. Anal. Prev. **45**, 711–717 (2012)
10. Jenkins, A., Eslambolchilar, P., Lindsay, S., Hare, M., Thornton, I., Tales, A.: Attitudes towards attention and aging. Int. J. Mob. Hum. Comput. Interact. **8**(2), 47–68 (2016)
11. Nahm, E.S., Preece, J., Resnick, B., Mills, M.E.: Usability of health web sites for older adults: a preliminary study. Comput. Inform. Nurs. **22**(6), 326–334 (2004). doi:10.1097/00024665-200411000-00007
12. Evaluating websites for older adults: adherence to 'senior-friendly' guidelines and end-user performance. Behav. Inf. Technol. **27**(3), 191–199 (2004). doi:10.1080/01449290600802031
13. Ebnali, M., Ahmadnezhadb, P., Shateri, A., Mazloumi, A., Heidari, M.E., Nazerie, A.R.: The effects of cognitively demanding dual-task driving condition on elderly people's driving performance; real driving monitoring. Accid. Anal. Prev. **94**, 198–206 (2016)
14. Lee, C.-F.: Legibility of text-color combinations of PDA for the elderly. J. Des. **10**(2), 1–12 (2005)
15. Tsai, W.P.: Explore the Relationship of Color Discrimination Ability and Leisure Life for the Elderly—A Case Study in Hualien. Department of Recreation and Sports Management, Tajen University, Yanpu (2014)
16. Ou, L.-C.: Visual Comfort as a Function of Lightness Difference Between Text and Background: A Cross-Age Study Using an LCD and a Tablet Computer. Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taipei (2014)

System Interactions in Aviation and Remotely Piloted Aircraft Systems

Do Age and Experience Level Affect Views of Pilots' Towards Cockpit Automation

Aysen K. Taylor(✉) and T. Steven Cotter

Engineering Management and Systems Engineering, Old Dominion University, Norfolk, VA 23529, USA
aysenafk@cox.net

Abstract. Pilots trained over the last 20 years are more likely to have learned to use cockpit automation as an inherent part of their training when compared to pilots trained in earlier eras who first learned to fly and then were trained in the use of automation. Cockpit automation facilitated the elimination of the flight engineer from the crew and shifted the remaining cockpit crew member's focus from hands on control to programmers and monitors of automated systems that can control the vertical and horizontal flight path for most of the flight. Some airline companies are accepting applicants for pilot training with no previous flight experience, this trend is unprecedented in the history of commercial flight in the US. While automation has improved the safety and efficiency of commercial flights, it has created new concerns relating to loss of manual flying skills and issues of mode confusion and the phenomenon of automation surprises. This paper analyzes the correlation between the pilots' perspectives and their demographics. A total of 77 pilots who operate highly automated cockpit systems participated in this research. Each question was analyzed regarding correlation by using R software and the Spearman Correlation Analysis method. Some of the questions revealed there is a significant difference towards cockpit automation when we consider the pilots' perspectives. This survey also revealed the pilots perspectives about their companies. Older pilots are more likely to think that their companies do not allow them to fly manually often enough. Although airline company policies are similar, the survey responses reflected that older pilots are not happy with those policies as well. Three demographic specifications were used for the Spearman's Correlation Analysis. Age was one specification and the other two were Flight Time correlation and Seat correlation (Captain versus First Officer). The total number of questions for five dimensions was 37 and 9 of these questions resulted in significantly different answers when considering the demographics. The answers provided by end users (pilots) should be considered when engineers create automated systems that are to be used and monitored by humans. Airliners that are 100% autonomous are not possible in the foreseeable future. Automated aircraft control systems should be designed to exploit the best qualities of both the technology and the human operators across different demographics.

Keywords: Cockpit automation · Spearman's correlation · Pilots' perspectives

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_29

1 Introduction

A survey conducted amongst airline pilots revealed some of their concerns related to automated systems in the cockpit. Automation provides several benefits while creating some concerns in commercial aviation.

Pilots who took this survey ranged in age between 20–65 years old with only one respondent between 20–25 years old and one respondent between 56–60 while one was between 61–65. The majority of the pilots were between 26–55 years of age and only one pilot was retired. The remaining pilots are still actively flying. The total number of pilots participating in the survey was 77. Out of 77, 11 pilots were eliminated since they failed to answer some of the questions. The types of aircraft they were flying includes but was not limited to Boeing 777, 737, 757, and 767 and Airbus models 320, 330, 340, 380, as well as McDonnell Douglas MD11, and 80–88. Most of these planes have highly automated cockpits; some with fly by wire systems. All the Airbus aircraft and the Boeing 777 use flight envelope protection software. Thirty two percent of the pilots were captains in their current aircraft while 68% of them were first officers. The survey was designed around five dimensions to understand pilots' perspectives. The dimensions included trust, monitoring, interface, policies, and training.

Questions between 14 and 51 were built as multiple choice having the options of strongly agree, agree, neither/nor agree/disagree, disagree, and strongly disagree. Questions between 14–22 involved the trust dimension which evaluates trust of the pilots for automated systems, 23 to 32 were in the monitoring dimension, evaluating the pilots ability to monitoring the automated systems of the cockpit, 33 to 38 were in the interface dimension, evaluating the current interface and their interactions with pilots, 39 and 45 were in the policy dimension, built for evaluating the pilots' perspectives regarding current policies applied by regulatory systems and as well as their companies. Questions 46 to 51 were in the training dimension and evaluated pilots' opinion about their current and desired level of flight training. This paper presents the results from using Spearman analysis to distinguish some of the response results based on the demographics.

2 Correlations Between Questions and Demographics

A summary of the Spearman's Analysis results is listed below in Table 1. Although the majority of questions have no correlation to demographics, some of them show great significance such as question 41 with age correlation, question 46 with flight time correlation, in addition to other questions.

2.1 Age Correlation

There is no statistical significance in the answers for the questions related to Trust dimension and age group of pilots (Table 2).

Table 1. Spearman's Analysis results for demographics and survey questions

Question	Age correlation		Flight time correlation		Seat correlation	
	P value	rho value	P value	rho value	P value	rho value
14T	0.4435	−0.09649845	0.3614	−0.1148596	0.3318	−0.1220778
15T	0.3202	0.1250285	0.04319*	0.2519175*	0.8869	−0.0179602
16T	0.5583	−0.0737674	0.8297	−0.0271408	0.7818	−0.0349471
17T	0.5853	−0.0687781	0.9951	0.00080337	0.03171*	−0.267215*
18T	0.5481	−0.0756889	0.4077	0.1042358	0.938	−0.0098359
20T	0.5592	0.07360259	0.8818	0.01876821	0.6101	−0.0642827
21T	0.5879	−0.0682990	0.9369	0.01000136	0.5361	−0.0779754
22T	0.4776	−0.0894496	0.5731	0.07103051	0.1678	−0.172994
23M	0.04345*	0.2516151*	0.6483	0.05750804	0.2598	−0.1415972
24M	0.5449	0.07630008	0.3147	−0.1264407	0.1452	−0.1825885
25M	0.2558	0.1427997	0.7	0.04859951	0.1564	−0.1777282
26M	0.3989	0.1061991	0.3101	0.1276432	0.05153	−0.2428171
27M	0.446	0.09597444	0.1787	0.1687076	0.6148	−0.0634507
28M	0.4317	−0.099004	0.7136	−0.0463009	0.9408	0.009394
29M	0.353	−0.116861	0.0577	−0.2368299	0.3799	−0.1105084
30M	0.1669	−0.1733567	0.02365*	−0.281050*	0.6862	−0.0509563
31M	0.03671*	0.260065*	0.2806	0.1356485	0.1892	−0.1647623
32M	0.1534	−0.1789733	0.1522	−0.1794904	0.9465	0.0084826
33I	0.6624	0.0550644	0.99	0.0015979	0.09948	−0.206118
34I	0.6101	−0.0642900	0.2133	−0.1562617	0.6852	0.0511191
35I	0.588	−0.0682904	0.7578	0.0388974	0.2837	−0.1347652
36I	0.9289	−0.0112800	0.7835	−0.0346698	0.9727	−0.0043356
37I	0.87	0.0206588	0.5206	−0.0809387	0.3691	0.1130216
38I	0.6135	0.0636755	0.6829	−0.0515121	0.4697	0.0910665
39P	0.4703	−0.0909405	0.2352	−0.1491154	1	0
40P	0.1731	−0.1708912	0.1959	−0.1623196	0.7852	0.0343832
41P	0.01108*	−0.314239*	0.03674*	−0.260028*	0.007595*	0.329557*
42P	0.3224	−0.1244548	0.5532	−0.0747318	0.9632	0.0058482
43P	0.959	−0.0065038	0.4221	0.1010808	0.297	−0.1311213
44P	0.09568	0.2084284	0.09762	0.2072422	0.188	−0.165213
45P	0.7993	−0.0320841	0.355	−0.1163816	0.3255	−0.1236797
46R	0.5701	−0.0715828	0.00345*	−0.359411*	0.3839	0.1095972
47R	0.05738	−0.2371291	0.2914	−0.1326616	0.7442	−0.0411555
48R	0.7875	−0.0340155	0.02281*	−0.282718*	0.339	0.1202913
49R	0.9436	0.0089371	0.08947	−0.2123669	0.1349	−0.1873406
50R	0.3504	0.1174969	0.9684	−0.0050142	0.2981	0.130825
51R	0.9168	0.0132042	0.639	0.0591495	0.4091	−0.1039135

Table 2. Spearman's correlation for trust and age

Question	Age correlation with trust	
	P value	rho value
14T	0.4435	−0.09649845
15T	0.3202	0.1250285
16T	0.5583	−0.07376743
17T	0.5853	−0.06877815
18T	0.5481	−0.0756889
20T	0.5592	0.07360259
21T	0.5879	−0.06829907
22T	0.4776	−0.08944963

Question 23M; related to Monitoring dimension indicates statistical significance when question asked as "*I always check my next waypoint and I'm aware of pending heading changes*". The question reveals slight correlation with 0.04345 P value and 0.25161 rho value, between age groups and the answers for this question. The younger group of pilots more strongly agrees; they are always checking their next waypoints, and keep themselves aware of pending heading changes.

Question 31 was also statistically significant and answers to the question indicate that younger pilots more strongly agreed they could predict the behavior of automation with ease versus the older pilots. Although in the answers there was no disagree or strongly disagree present, we could make the correlation between age and only 3 choices of alternative answers; Strongly agree, agree or neither nor. The answers of this question show 12% of the participants strongly agreed, 79% agreed and 9% chose either/or option. Overall consensus reflected a positive response to automation behavior. Results of the spearman analysis for correlation between age and monitoring are shown above in Table 3.

Table 3. Spearman's correlation for monitoring and age

Question	Age correlation with monitoring	
	P value	rho value
23M	0.04345**	0.2516151**
24M	0.5449	0.07630008
25M	0.2558	0.1427997
26M	0.3989	0.1061991
27M	0.446	0.09597444
28M	0.4317	−0.09900467
29M	0.353	−0.116861
30M	0.1669	−0.1733567
31M	0.03671**	0.2600648 **
32M	0.1534	−0.1789733

There was no significant correlation in the answers between age and interface dimension as shown below in Table 4.

Table 4. Spearman's correlation for interface and age

Question	Age correlation with interface	
	P value	rho value
33I	0.6624	0.05506439
34I	0.6101	−0.06429001
35I	0.588	−0.06829038
36I	0.9289	−0.01128001
37I	0.87	0.02065882
38I	0.6135	0.06367553

Recent inflation of the automation [1] and the results of overstated benefits without the consequences, more and more airline companies force their crews to use automated systems mostly throughout their flights. Question related this subject brought the most variety of the answers in the survey.

Table 5. Spearman's correlation for policies and age

Question	Age correlation with policies	
	P value	rho value
39P	0.4703	−0.09094046
40P	0.1731	−0.1708912
41P	0.01108**	−0.3142389**
42P	0.3224	−0.1244548
43P	0.959	−0.006503845
44P	0.09568	0.2084284
45P	0.7993	−0.03208414

Question 41 was presented as; "*my airline doesn't allow me to fly manually, they force me to fly automated most or all of the time*" brought various answers; 20% of the participants chose not to answer (neither agree or disagree option), 43% agreed/strongly agreed while 37% disagreed/strongly disagreed. When we broke down the demographics we saw that older pilots were more likely to agree that their companies do not let them fly manually compared to younger pilots. This result may have reflected the different training experiences younger and older pilots had relating to cockpit automation. Newer pilots have automation emphasized in their training early on whereas older pilots typically focused on manual flight first and later added automation. Survey responses could have been impacted by this different training history and affected by pilots' experiences using automation to control their flight path for the great majority of their flight. The newer generation of pilots sees flying in highly automated modes most of the time as normal operations; they have little experience conducting flights in manual modes. Results of the Spearman analysis for correlation between age and policies are shown above in Table 5.

Crew Resource Management (CRM) is one the key component of the Aeronautical Decision Making (ADM) process and encourages pilots to speak up freely without concern for seniority or cultural differences [2]. Call outs and distributing the work amongst the crew (pilots, first officers) is critical for any flight. Even a student pilot has to comply with these basic procedures during their training in that a pilot flying cannot transfer the flight controls to the non flying pilot without doing several call-outs and confirmations. In contrast, the autopilot can disengage suddenly without giving proper output; a "ding" sound from an autopilot and it transfers all control to the pilots without confirmation from them. In light of these issues we thought it would be appropriate to get pilots' feedback regarding this. Question 44 directed as; *Regulations should require cockpit automation to adhere to the principles of CRM when feasible*. 72% of the participants agreed/strongly agreed, 23% chose not to answer (neither agreed or disagreed) only 5% disagreed.

We did not find any correlation between the answers and age groups in this question regarding if automated systems should adhere to CRM principles. We wanted to include this question since it is very important for the governance aspect of the human-machine interaction in the cockpit [3].

We also did not find any correlation regarding Training aspect of the survey with age groups, shown below in Table 6.

Table 6. Spearman's correlation for training and age

Question	Age correlation with training	
	P Value	rho value
46R	0.5701	−0.07158285
47R	0.05738	−0.2371291
48R	0.7875	−0.03401553
49R	0.9436	0.008937134
50R	0.3504	0.1174969
51R	0.9168	0.01320424

2.2 Flight Time Correlation

Question 15 shows a slight statistical significance when the question directed as "*Automation keeps me engaged throughout the flight*". The correlation between positive answers and flight time (hours) shows that pilots with less flight time agree more with this statement compared to pilots with more flight time. This means the pilots with less flying hours think automation keeps them engaged with their flight more strongly versus the pilots with more flight time. Fourteen percent of respondents strongly agreed while 43% of them agreed on this question. Thirty two percent of them neither agreed nor disagreed. It was surprising that 1/3 of the pilots did not want to give a clear answer on this. Current research warns that automation creates complacency [4] and therefore pilots get out of the loop and lose their situational awareness [5], this could be the reason why 1/3 of the respondents did not want to answer this question. Only 12% of them disagreed; there was no strongly disagreed as an answer. Spearman analysis results for the correlation between trust dimension of the survey and flight time shown below in Table 7.

Table 7. Spearman's correlation for trust and flight time

Question	Flight time and trust	
	P value	rho value
14T	0.3614	−0.1148596
15T	0.04319**	0.2519175 **
16T	0.8297	−0.02714085
17T	0.9951	0.0008033688
18T	0.4077	0.1042358
20T	0.8818	0.01876821
21T	0.9369	0.01000136
22T	0.5731	0.07103051

Question 30 shows significance with P value, 0.02365. It reveals slight negative correlation between the answers and flight time. Pilots with more flight time, less strongly agreed that their understanding of the picture does not diminish when they use HDG, SPD or ALT as autopilot control inputs versus pilots with less flight time. The overall answers for this question were positive regarding the automated systems; only 7% of the pilots agreed that their overall understanding of the big picture diminishes. Although 88% of the pilots gave positive input for automation in this question, it still shows pilots with fewer flight hours place more emphasis on automated systems in the cockpit. Results of the Spearman test for monitoring dimension and flight hours shown below in Table 8.

Table 8. Spearman's correlation for monitoring and flight time

Question	Flight time and monitoring	
	P Value	rho value
23M	0.6483	0.05750804
24M	0.3147	−0.1264407
25M	0.7	0.04859951
26M	0.3101	0.1276432
27M	0.1787	0.1687076
28M	0.7136	−0.04630093
29M	0.0577	−0.2368299
30M	0.02365**	−0.2810506**
31M	0.2806	0.1356485
32M	0.1522	−0.1794904

There was no correlation in the answers between flight and interface dimension, results shown below in Table 9.

Table 9. Spearman's correlation for interface and flight time

Question	Flight time and interface	
	P value	rho value
33I	0.99	0.001597952
34I	0.2133	−0.1562617
35I	0.7578	0.03889741
36I	0.7835	−0.03466979
37I	0.5206	−0.08093874
38I	0.6829	−0.0515121

Question 41 brought the most controversial answers as we discussed in the Age Correlation analysis. As the flying years increase, more pilots agreed that their companies do not let them fly manually; they are being required to fly under automated control most of the time. Flight time analysis confirmed the same findings in Age correlation with this question. As the flight hours (we may also call this experience) increase, pilots believe that they are being forced by their companies to fly more with automated systems. These answers are very relevant to the Department of Transportation's (DOT) January 2016 Audit Report findings that reveal the problems about manual flying skills and keeping them proficient. The FAA only has vague suggestions that airline pilots maintain manual flying skills by flying without automation when feasible and currently does not mandate any manual flight training protocol. Pilots accustomed to flying mostly in automated modes can have difficulties when manual flight control is suddenly required. The FAA does not have mandates that require airline pilots to have training designed to ensure their safe use and monitoring of automation while also maintaining manual flying skills. The FAA also has no oversight of air carrier tracking of pilots' manual mode flight hours or carrier's manual flight training programs. The FAA has no specific auditing programs that ensure airline training programs are effective in the maintenance of manual flying skills [6]. Results of the Spearman analysis regarding correlation between policies dimension of the survey and flight time are shown below in Table 10.

Table 10. Spearman's correlation for policies and flight time

Question	Flight time and policies	
	P value	rho value
39P	0.2352	−0.1491154
40P	0.1959	−0.1623196
41P	0.03674**	−0.2600284**
42P	0.5532	−0.07473181
43P	0.4221	0.1010808
44P	0.09762	0.2072422
45P	0.355	−0.1163816

Question 46 also provided one of the significant results in the survey. Question directed as "*A high level of competency in manual flying skills would benefit the industry*", 49% of the respondents strongly agreed while 33% of them agreed, 13% of them neither agreed or disagreed, while only 4% disagreed and 1% strongly disagreed. When the question was analyzed with Spearman correlation analysis, the pilots with more flight time more strongly agreed on this than the pilots with less flight time. Although the majority of the pilots agreed/strongly agreed in response, this analysis reveals that pilots with more flight time are more enthusiastic about having competent manual flying skills.

Another significant result was for question 48. Question directed as "*A number of recent airline accidents/incidents could have been avoided if the pilots had been more proficient in manual flying skills*". As we saw in question 46, question 48 brought the same result; pilots with more flying hours more strongly agreed on this question. Results of the Spearman analysis regarding correlation between training dimension of the survey and flight time shown above in Table 11.

Table 11. Spearman's correlation for training and flight time

Question	Flight time and training	
	P value	rho value
46R	0.003454**	−0.359411**
47R	0.2914	−0.1326616
48R	0.02281**	−0.2827178**
49R	0.08947	−0.2123669
50R	0.9684	−0.00501424
51R	0.639	0.05914953

2.3 Seat Correlation

Modern commercial airlines have 2 seat alternatives, Captain, with seniority for that particular aircraft and First Officer (F/O), often with less flight hours for that particular aircraft. In some cases a captain and first officer may have a similar number of total flight hours but, for that particular aircraft, a F/O might have significantly less hours.

In our survey regarding the trust dimension Question 17 was directed as; "*Through use of FMC (Flight Management Computer) for automated flight planning (e.g. planning of route, waypoints etc.) my overall workload is lower*". Spearman analysis results showed slight significance with 0.03171 P value, and the correlation is negatively related. This means captains don't believe as much as first officers that their workload is getting lower. Results of the Spearman analysis regarding correlation between trust dimension of the survey and seat alternative shown below in Table 12.

Monitoring, interface and training dimensions of the survey did not bring any significant correlations when considering seat positions and demographics. The only other significant dimension was with question 41 in policies dimension; P value 0.007595. Question 41 was directed as "*my airline doesn't allow me to fly manually, they force me to fly automated most or all of the time*". This question brought the most

Table 12. Spearman's correlation for trust and seat alternative

Question	Seat alternative and trust	
	P value	rho value
14T	0.3318	−0.1220778
15T	0.8869	−0.0179602
16T	0.7818	−0.0349471
17T	0.03171*	−0.267215*
18T	0.938	−0.0098359
20T	0.6101	−0.0642827
21T	0.5361	−0.0779754
22T	0.1678	−0.172994

significant results for every demographic we evaluated. In this case First Officers more often agreed their companies do not allow them to fly manually versus Captains. Results of the Spearman analysis regarding correlation between training dimension of the survey and seat alternative shown below in Table 13.

Table 13. Spearman's correlation for policies and seat alternative

Question	Seat alternative and policies	
	P value	rho value
39P	1	0
40P	0.7852	0.03438318
41P	0.007595**	0.329557**
42P	0.9632	0.005848193
43P	0.297	−0.1311213
44P	0.188	−0.165213
45P	0.3255	−0.1236797

3 Conclusions

Findings from Spearman test results aligned with current literature and reflect findings that the new generation of pilots is more automation friendly since they have not flown much in manual modes. A decline in manual flying skills is a trend caused by airline companies policies and even some organizations such as NASA. Research conducted by NASA encourages teaching cockpit automation in the "classroom". [7] Classroom teaching should be considered a small portion of the training. Recently more emphasis is placed on upset recovery training using real aircrafts not only simulators. Pilots are aviators first and they are responsible to fly their aircraft with or without automated systems. Including automation as a crew member in the cockpit but without sufficient feedback regarding what the automated systems are doing i.e. insufficient mode annunciation; creates problems regarding pilots cognitive model of their flights.

References

1. Taylor, A.K.: Human-machine interaction in the cockpit and applicable design changes towards better collaboration, conference presentation. Paper presented at the AHFE, Orlando, FL (2016, July)
2. Advisory Circular [FAA advisory circular AC No:120-51E]. Federal Aviation Administration, Washington, DC (2004)
3. Taylor, A.: Human-machine interaction in the cockpit and applicable design changes towards better collaboration. In: Nunes, I. (ed.) Advances in Human Factors and System Interactions. Advances in Intelligent Systems and Computing, vol. 497, pp. 271–279. Springer, Cham (2017)
4. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: an attentional integration. Hum. Factors **52**(3), 381–410 (2010)
5. Endsley, M.R.: Situation awareness in aviation systems. In: Garland, D.J., Wise, J.A., Hopkin, V.D. (eds.) Handbook of Aviation Human Factors, pp. 257–276. Lawrence Erlbaum Associates, Mahwah, NJ (1999)
6. Office of Inspector General Audit Report. Department of Transportation, Washington, DC (2016)
7. Casner, S.M.: Teaching Cockpit Automation in the Classroom. NASA, Hanover, MD (2003)

Performance Evaluation of Numeric Keypad Layout Based on Cognitive Workload

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

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

[2] National Key Laboratory of Human Factors Engineering, Astronaut Research and Training Centre of China, No. 26, Beiqing Road, Haidian District, Beijing 100094, China
huangshoupeng2005@163.com, tianzhiqiang2000@163.com, cctian@126.com

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

Abstract. A method for performance evaluation of numeric keypad layout based on cognitive workload is proposed and applied to model task's cognitive processes in detail and scrutinizes human workload at a cognitive level. To conduct experiment for performance evaluation of keyboard layout, numeric keypad test software whose layout is configurable is developed. The participant's keying behavioral data are collected during experiments. Cognitive model is constructed by using theories of ACT-R and verified by investigating the correlation coefficient between model and experiment. The cognitive workload is derived by summing the active time for each module of cognitive architecture as the model runs. The percentage of brain region's active time during the task is used to evaluate the task's performance with the mapping relationship between cognitive architecture and brain region. By comparing the percentage of the brain's active time with accomplishment of similar tasks, the performances of different numeric keypad layout are evaluated.

Keywords: Performance evaluation · Numeric keypad layout · Cognitive workload · Cognitive model · Cognitive architecture

1 Introduction

Numeric keypad is everywhere nowadays being utilized as an input device or calculator. In the meantime the study of the numeric keypad layout has attracted much attention in recent past [1–5]. Numeric keypad can be built in different key size, key space and key's layout for various purpose, and key's layout maybe telephone style, computer style, random key order style or even scramble key order style. For example,

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_30

it is expected that a fixed and familiar key layout could enhance entry speed, and however, with regard to accuracy rate, a random layout would force participants to visually search instead of recalling from memory for the location of the correct key, which might result in greater cognitive workload during key entry compared with a fixed key layout [6].

Key size and key layout are crucial factors affecting user performance (Celle and Hiszem 2004, Jin et al. 2007, Sesto et al. 2012, Tsang et al. 2013). Bender report in 1999 that a key size of 30 × 30 is better than 10 × 10 mm for keying a 4-digit numeric string, and Martin find in 1988 that the 13 mm square key size is better than a 6 × 13 mm or 13 × 6 mm key size in keying speed and accuracy. A study of Colle and Hiszan in 2004 shows that with the choices of square key size in 10, 15, 20 and 25 mm, the 20 mm square key results in superior performance and are most preferred by participants. While studies by Tricia S. Lowe in 1996 indicate that when utilizing a numerical data-entry keyboard, if the key layout has a poor/low level of Stimulus-Response (S-R) compatibility with normal expectations of the location of the digits, it needs more time for subjects to locate each key [7].

Due to the space limitation, manned spacecraft uses numeric keypad for data collection and control usually [8]. However, the spaceflight environment is characterized by temperature extremes, microgravity, solar and galactic cosmic radiation, lack of atmospheric pressure, and high-speed micrometeorites. While these factors induce a host of physiological, biomedical, and environmental stressors to flight crews, long duration spaceflight has revealed an additional group of stressors that impact crew performance and health. Workloads and task performance are critical to maintaining good working conditions for on-orbit crews. For the complexity of space environment and the change of astronaut's cognitive ability during flight, it is essential to consider ergonomic factors in design of numeric keypad layout for astronaut maintaining a suitable situation. Researchers have carried out studies on workload affects performance, and the result shows that excess workload can result in human performance issues such as slower task performance and errors such as slips, lapses or mistakes and even under workload can lead to human performance issues such as boredom, loss of situation awareness and reduced alertness [9–11]. An assessment of workload should be required if one want to keep an astronaut good situation during space flight. Workloads and task performance evaluation are critical to maintaining good working conditions for on-orbit crews [12].

In fact, most of the previous studies on keying performance of keypad layout were on the base of mean entry speed, mean accuracy rate, and mean completion rate. Nevertheless, studies on effect of keypad layout on cognitive workload are relatively less proliferous, while cognitive workload is the direct factor that really determines human cognitive ability. In this study, we examine the effects of key size and layout on keying performance with workload in a level of cognitive detail, evaluate cognitive workload and task performance for numeric keypad operation of spaceflight and provide the basis for numeric keypad design for space capsule and improvement of improving astronauts' performance.

2 Method and Experiment

The purpose of this study is to test the numeric keying performance for different numeric keypad layout. The cognitive active time of the task process is taken as the measures for the workload assessment, and then the cognitive workload is applied to the task's performance evaluation. The key methods involved in this study include software development and experimental design, model verification and single-key entering time calculations formula etc.

2.1 Experiment Design and Test Software Implementation

Experiment Design. A total of twenty-seven subjects were recruited and participate the test experiment and each of them finishes 10 tasks, and we use a digital with width of six digits in the experiment. Each experiment consists of multiple tasks, each task applying different keyboard layout and/or key sizes with several digital string and every task involve two stages of practice and test.

Software Platform. ACT-R (Anderson et al. 2004) is a highly sophisticated computational architecture designed to support modeling of human cognition and performance at a detailed temporal grain size, has been used successfully to modeling a variety of behavioral phenomena and has proven particularly successful at modeling tasks with a demanding cognitive component [13]. Therefore, ACT-R cognitive architecture is chosen to build cognitive model in this study. For the visual studio is the most diverse coding application and the most intuitive development platform, Visual Studio 2013 C# is selected as the development tools for the configurable keypad entering software.

Test Software Implementation. The software is designed to perform a key ergonomics test for digital input, and its main function include layout user-configurable keypad, recording function for keying behavior and cognitive behavior analysis and cognitive behavior simulation of the task. Figure 1 shows the parameters configuration interface for numeric keypad layout.

2.2 Model Keying Time Empirical Formula

Experience shows that for the non-first digit's input, the keying time satisfies the following empirical formula.

$$T_0 = T_1 + T_2 \tag{1}$$

where T_0 represents the basic reaction time, the basic reaction is the most basic human's ability, T_0 almost does not contain scanning, attention and other cognitive procedural, as for a specific subject it is a fixed value and its default value

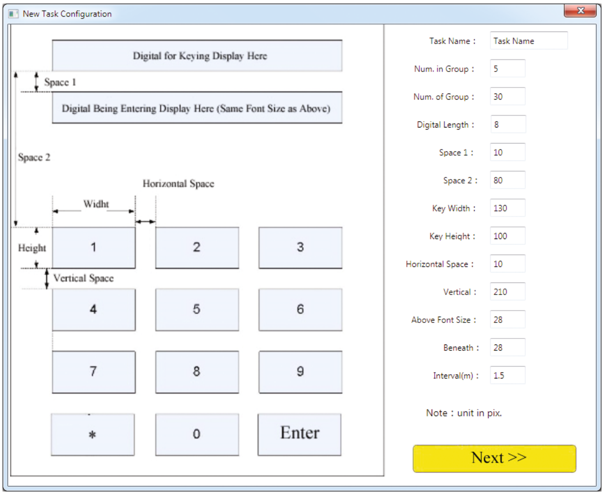


Fig. 1. The parameters configuration interface for numeric keypad layout.

$T_0 = 150$ ms. T_1 represents finger movement time and it can be calculate by following formula.

$$T_1 = 10 * \sqrt{s} \tag{2}$$

where s denotes the finger move distance from keying key to the previous keying key, and its unit is millimeter (mm).

T_2 in Eq. (2) represents finger alignment time and it is calculate by following formula (3).

$$T_2 = 10 * ((FP/w)^2 + (FP/(w+d))^2 + 1) \tag{3}$$

where FP (FingerPad) denotes the radius for the circle touch area of the finger and the key, it is a fixed value for specific subject and its default value FP = 11 mm. The w denotes key width (mm), and d denotes key space (mm).

The formula (1) is the keying time for the non-first key without concerning the scanning, grouping, and memory encoding times. As for the first digit, the keying time should take these time and others into account.

2.3 Cognitive Model Verification Methods

In order to evaluate the cognitive workload, it is important to clarify the cognitive process. Whether the cognitive process is consistent with the experiment directly reflects whether the cognitive workload meets the *real* cognitive workload. The experimental data and the model data are used to determine the workload of the model and the experimental data. To avoid singularity sample in experimental data, this study checks the experimental data by using standard deviation.

Mean Value of Keying Time for Experiment (MKE). Figure 2(a) shows the flow chart diagram for calculating mean keying time value of the subject in the experiments. MKT is used to calculate standard deviation of the experimental data for investigating the regularity of sample data.

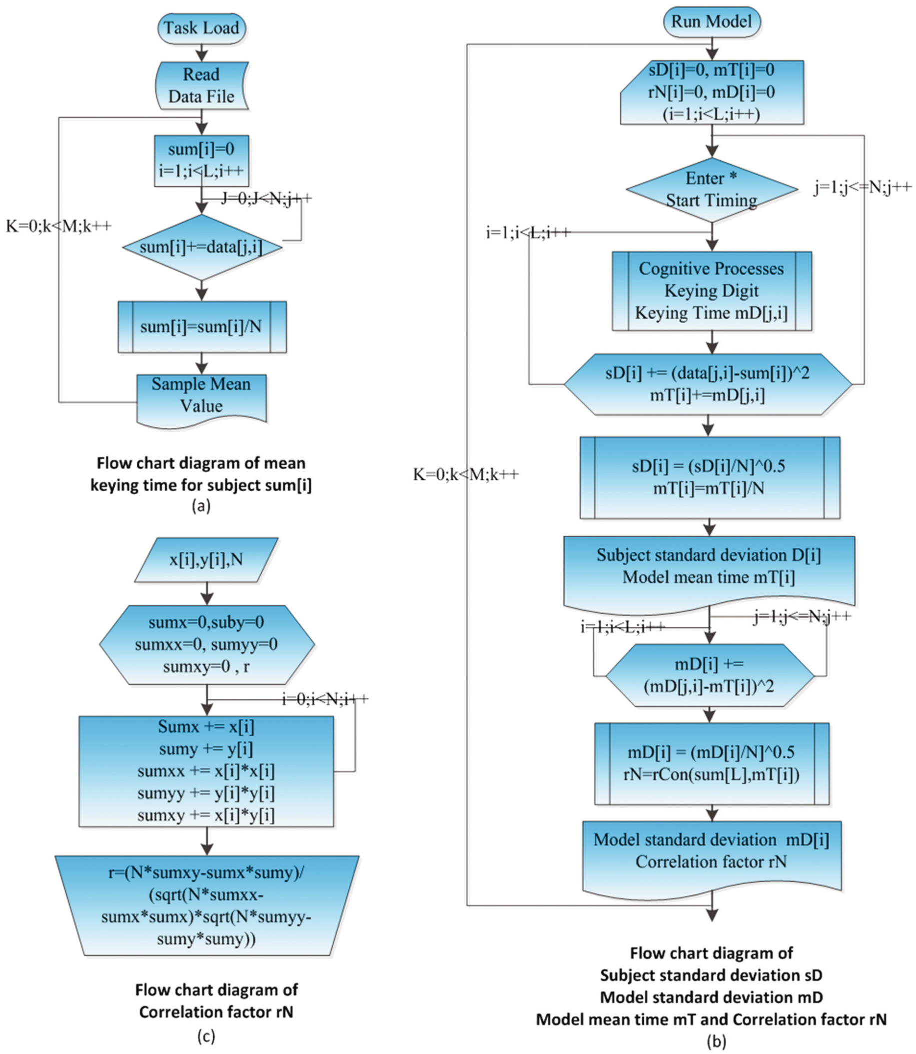


Fig. 2. Calculating diagram for keying time mean value of experiment, keying time mean value of model, standard deviation and data correlation coefficient between experiment and model result.

Standard Deviation and Mean Value of Keying Time for Model (MKM). Figure 2 (b) gives the flow chart diagram of calculating standard deviation for subject keying time, the flow chart diagram of calculating model keying time, the flow chart diagram of calculating standard deviation of the model, and the flow chart diagram of calculating data correlation coefficient between experiment and model.

Correlation between Experiment and Model. Figure 2c shows the algorithm subroutine diagram of correlation coefficient between experiment and model.

3 Cognitive Processing Analysis

As mentioned above, width of 6-digit string used as a case study for layout performance evaluation. A cursory glance at the test experiment one will find almost all of subject split the six digits into two groups, i.e. they read three letters keying them and then read all the others then finish the numeric keying task. Therefore, the cognitive behavior of keying a width of 6-digit string follows.

(a) Start keying task
(b) Scan whole digit string
(c) Group the whole string into two groups
(d) Scan first group
(e) Enter the first letter of first group
(f) Enter the second letter of first group
(g) Enter the third letter of first group
(h) Scan the second group
(i) Enter the first letter of second group
(j) Enter the second letter of second group
(k) Enter the third letter of second group
(l) Finish numeric string input

These cognitive behavior processes will be the main sources of procedural knowledge for cognitive model.

3.1 Cognitive Model

Aim to build an ACT-R cognitive model for numeric keypad keying task is to make clarify of the cognitive processes and then to evaluate the cognitive workload for each cognition segment. A running cognitive model can vividly descript the cognition details in every cognitive cycle 50 ms. The activity and its active time in each module of the cognitive architecture will be displayed and can be recorded. For the reason of cognitive architecture maps its each module to a corresponding human brain region, one may take it for granted that the module's activity may reflect human's brain region's activity. Therefore, the total time of the module in cognitive architecture may represent human brain's workload.

The main job to create an ACT-R based cognitive model is to build the procedural knowledge fired consecutively during model run, and to abstract model's parameters for the purpose of modeler's requirement. Based on the cognitive processes mentioned above, the main procedural knowledge of the numeric keying task's cognitive model involves (*a*) *scanning full-length digit*, (*b*) *grouping the whole digit*, (*c*) *scanning first group*, (*d*) *keying first digit of the first group*, (*e*) *keying second digit of the first group*, (*f*) *keying third digit of the first group*. Then the procedural knowledge of the other groups is dealt with as well as the first group and so on. The relevant keying time for each digit in the group can be calculated by formula (1).

3.2 Model Verification and Simulation

This study verifies the model at two stages. The first stage is the model successfully complete the model's task which model built based on. The second stage is the every model's single cognitive behavior coincides with actual in details. The first stage verification is easy to achieve and the second stage verification we apply simulation method to implement. According to ACT-R based cognitive processes this study simulate numeric keypad keying task appeared in experiment and examine the simulation output compliance extend with experimental data by using data standard deviation and data correlation to verify what a satisfaction extend for cognitive model. Figure 3 shows the processes of cognitive model, standard deviation/data correlation, and keying behaviors' simulation outputs.

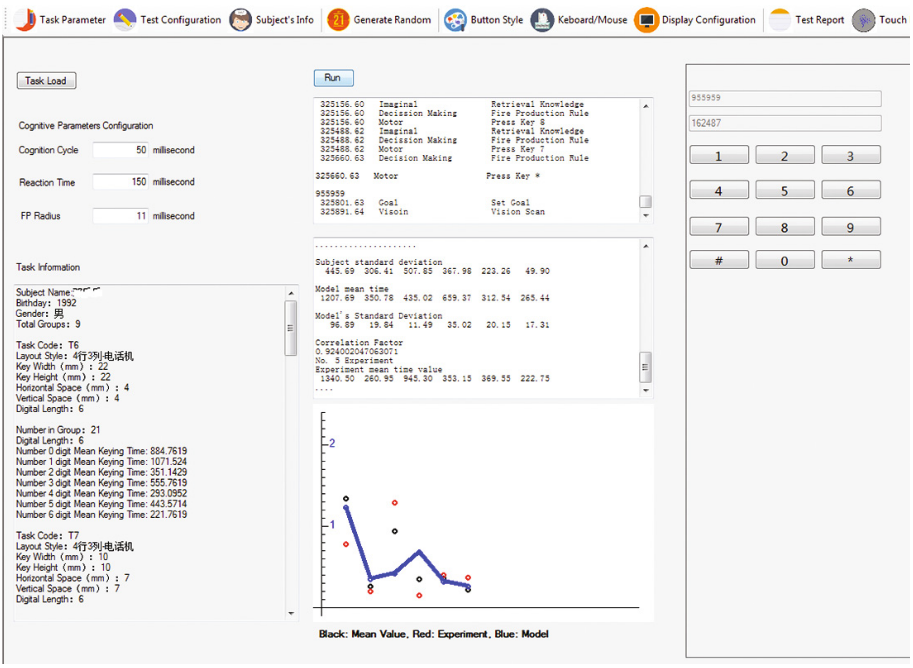


Fig. 3. The processes of keying cognitive mode, data standard deviation for both experiment and model, data correlation between experiment and model, and keying task's simulation.

4 Workloads and Task Performance

Cognitive abilities are brain-based skills we need to carry out any task. ACT-R implements its each module explicit mapping to a certain human brain functional regions. While model running, ACT-R shows and records active buffer status that corresponding to relevance module of that cognitive period. It is reasonable to say that while the module is active it is the brain is working, and the brain working is the human has workload. While summing the brain active time for a task one can consider the total time as the workload of finishing the task.

4.1 Workloads

Since the cognitive workload can be represented by the brain activation time of the task-related cognitive activity, we calculate the total amount of time that cognition related brain activation of the task. By comparing the total amount of time that task cognition related brain activation one can investigate workload for different numeric keypad keying task. As an example, Fig. 4 shows the workload of a group keying tasks in a numeric keypad keying test experiment for 10 digital strings.

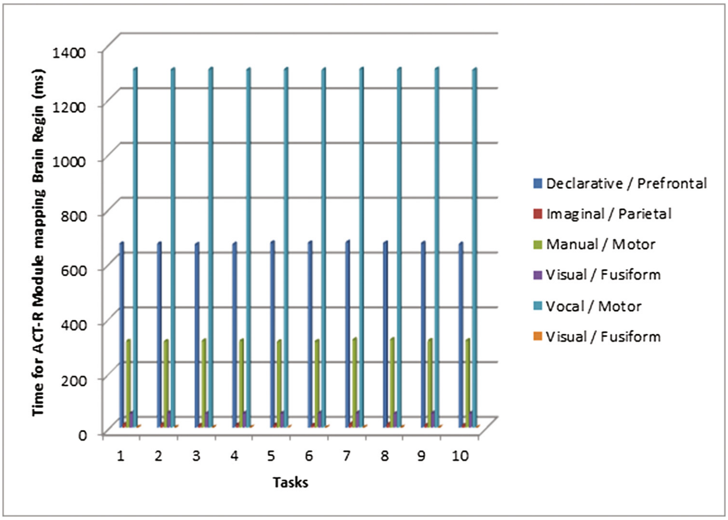


Fig. 4. The workload of a group tasks of a numeric keypad keying test experiment.

4.2 Task Performance

The workload is evidenced by brain's cognition time with different keypad layout. A rational layout may result in low workload and high performance and vice versa. Therefore, to investigate task performance with different numeric keypad's layout, one need to figure out cognition related brain region are active at what time and what are

Table 1. The task's performance of group keying tasks.

Task	T1	T2	T3	T4	T5	T6	T7	T8	T9
Declarative/Prefrontal	3.78%	3.15%	2.73%	2.49%	2.20%	18.53%	9.23%	6.39%	4.73%
Imaginal/Parietal	0.20%	0.16%	0.16%	0.16%	0.16%	0.31%	0.27%	0.19%	0.20%
Manual/Motor	1.79%	1.51%	1.29%	1.19%	1.05%	8.82%	4.44%	3.07%	2.23%
Goal/ACC	0.29%	0.26%	0.22%	0.20%	0.18%	1.44%	0.74%	0.51%	0.38%
Visual/Fusiform	6.91%	5.83%	5.00%	4.55%	3.95%	17.36%	17.36%	12.00%	8.79%
Vocal/Motor	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Others	87.03%	89.09%	90.60%	91.40%	96.41%	35.03%	67.96%	77.85%	83.68%
Task total time (ms)	381387	457820	534369	614448	693553	76271	150995	228957	305585
Correlation coefficient	0.84	0.81	0.82	0.74	0.96	0.89	0.86	0.92	0.92

their activation percentage in the overall workload. Table 1 shows the task's performance of a nine groups keying task in a numeric keypad keying test experiment.

5 Discussion

We observed that the mean keying time of the third digit is longer than that of the second one in the first group of a six digits test experiment. Theoretically, there should be less different between them. What is the reason?

It also can be found from experimental data that there is a maximum value with standard deviation of mean keying time for the third digit. This indicating that there exists disagreement data or the true situation is changeable. Theoretically, inference, the cognitive processes of the second and third digit should be same as the fifth and sixth digit. At the beginning, the procedural knowledge involved in numeric keying are the following four categories: (*a*) the first digit (need scan and group before), (*b*) the fourth digit (first digit of a group, need scan before), (*c*) previous digit's repeat (besides the first in a group), and (*d*) other. These are corresponding to the basic knowledge obtained from above analysis for cognitive modelling. The construction of the initial model is mainly depended on all these procedural knowledge, however it is found that the model result are not fully fit the actual experimental data while model running. Careful analyzing, we found that we assume the subjects conduct a second scans with the forth digit. However, some subjects carry out scan again before keying the third digit and implement the second group's scan after the first group, but they almost never scan again with the second digit. This is the reason why the keying time of the third digit is longer than that of the second one and why the mean keying time standard deviation of the third digit is larger. After revising the model cognitive processes, the results of the model are consistent with the experimental data.

We analyzed the reason why the mean keying time of the third digit is higher than that of the second one in the experiment above. We also found that the mean keying time of the sixth digit is higher than that of the fifth one, and they are not as well as that of relationship between the third digit and the second digit, while theoretically the two processes are the same. Observation reveals that the two cognitive processes are definitely not alike, and there is no re-scanning behavior with the sixth digit at all during experiment as well as the third digit.

6 Conclusion

This paper presents a method to evaluate the task performance for touch screen numeric keypad layout in spacecraft on the base of cognitive workload. In this study, keypad layout configurable software is developed, the procedural knowledge of the cognitive model is constructed by analyzing specific task, and the cognitive model are built for numeric keypad keying task. The study also explores model's effective verification method and brilliantly adopt of cognition activity time as an important index of cognitive workload evaluation. By mapping the cognitive processes to the brain region activities through ACT-R typical cognitive architecture, accordingly the quantitative evaluation of mental workload achieved. Finally, according to the quantitative of the workload and the percentage of the whole task this paper implements the evaluation of the task performance. Taking a six digits width digital as a study case the study analyzes the cognitive process, the cognitive workload and the task performance for numeric keying task. The results show that the key size, key space, and layout style all affect the numeric keying workload and performance.

Acknowledgements. This work is supported by the Feitian Foundation of China Astronaut Research and Training Center (No. FTKY201505), the China Advanced Space Medico-Engineering Research Project (No. 2012SY54A1705) and Zhejiang Provincial Natural Science Foundation (No. LY12C09005).

References

1. Armand, J.T., Redick, T.S., Poulsen, J.R.: Task-specific performance effects with different numeric keypad layouts. Appl. Ergon. **45**(4), 917–922 (2014)
2. Marteniuk, R.G., Ivens, C.J., Brown, B.E.: Are there task specific performance effects for differently configured numeric keypads? Appl. Ergon. **27**(5), 321–325 (1996)
3. Jin, Z.X., Plocher, T., Kiff, L.: Touch screen user interfaces for older adults: button size and spacing. In: International Conference on Universal Access in Human-Computer Interaction, pp. 933–941. Springer, Heidelberg (2007)
4. Bakelman, N., Monaco, J.V., Cha, S.H., Tappert, C.C.: Keystroke biometric studies on password and numeric keypad input. In: 2013 European Intelligence and Security Informatics Conference (EISIC), pp. 204–207. IEEE (2013)
5. Oladimeji, P., Thimbleby, H., Cox, A.: Number entry interfaces and their effects on error detection. In: IFIP Conference on Human-Computer Interaction, pp. 178–185. Springer, Heidelberg (2011)
6. Tsang, S.N.H., Chan, A.H.S. Chen, K.: Effects of key size and key layout on numeric data entry performance with touch screens. In: IAENG Transactions on Engineering Sciences: Special Issue of the International MultiConference of Engineers and Computer Scientists 2013 and World Congress on Engineering 2013, vol. 279. CRC Press (2014)
7. Tsang, S.N.H., Chan, A.H.S. Chen, K.: A study on touch screen numeric keypads: effects of key size and key layout. In: International MultiConference of Engineers and Computer Scientists (2013)

8. Tan, L.F., Zhou, S.H., Tian Y., Wang C.H.: Research on display-control layout of haptic interaction interface based on character input task. In: Proceedings of the 15th International Conference on Man–Machine–Environment System Engineering, pp. 457–465. Springer, Heidelberg (2015)
9. Gore, B.F.: Workload as a performance shaping factor for human performance models. In: 0th Behavior Representation in Modeling and Simulation BRIMS Conference, Sundance, Utah (2011)
10. Grussenmeyer, W., Folmer, E.: Accessible Touchscreen Technology for People with Visual Impairments: A Survey (2017)
11. Michelon, P.: What are Cognitive Abilities and Skills, and How to Boost Them. SharpBrains (2006). http://sharpbrains.com/blog/2006/12/18/what-are-cognitive-abilities
12. Liu, Y., Tian, Z., Liu, Y., Li, J., Fu, F.: Cognitive architecture based platform on human performance evaluation for space manual control task. In: Advances in Neuroergonomics and Cognitive Engineering, pp. 303–314. Springer (2017)
13. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychol. Rev. **111**(4), 1036–1060 (2004)

Motor Imagery Brain–Computer Interface for RPAS Command and Control

Rosa Arnaldo(✉), Fernando Gómez Comendador, Luis Perez, and Alvaro Rodriguez

School of Aeronautics and Space Engineering, Universidad Politecnica de Madrid, Plaza de Cardenal Cisneros 3, 28040 Madrid, Spain
{rosamaria.arnaldo,fernando.gcomendador,l.perez, alvaro.rodriguez.sanz}@upm.es

Abstract. Nowadays, technology is evolving towards the development of new controlling methods based in signals produced by our brain (Brain Computer Interfaces BCI). Applications of Brian computing Interfaces are also being explored in the field of aeronautics. This paper presents the initial steps of a work focused on the evaluation of brain patterns that occur when an individual excites the brain to perform an action. The final goal of this project is to implement this method in a real time program that is capable of filtering the signals obtained by brain measurement system; treating these signals to obtain the amplitude accumulation at the indicated frequencies and sending the control commands to a RPAS in order to be able to control it.

Keywords: Brain Computer Interface · Motor system · Signal processing · EEG

1 Introduction

Brain-computer interfaces (BCI) are devices that facilitate communication without movement through the conduction of signals outside the nervous system without engaging muscles [1].

According to the most extended taxonomy [2], BCIs imply direct measures of brain activity, feedback to the user, on-line operation and intentional control. Intentional control means that the user has to execute a mental task each time he want the BCI to generate a command.

Figure 1 presents the high-level operational concept of a BCI. It illustrates the processing of signals in a brain-computer interface. Firstly, brain activity is registered by an EEG (signal acquisition). Later, changes and alterations in frequencies and signal patterns are extracted from registered signals (feature extraction). After that, observed signal changes are correlated with the reactions of the device, using classification algorithms (feature classification). Finally, the identified features are converted into commands to control the device, (e.g. into parameters of flight in a flight simulator). Feedback about the reactions of the system is provided to the user who can consequently modify its brain activity.

More than 150 companies worldwide are currently concerned with BCI, from the marketing and technology companies to the aviation industry. The roadmap and strategy

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_31

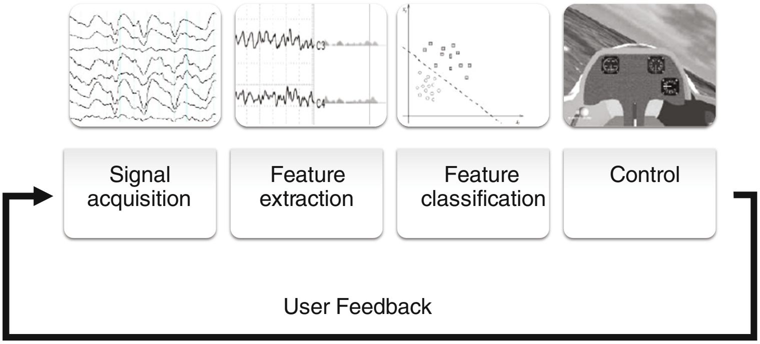


Fig. 1. Processing of signals in a Brain Computer Interface

for Horizon 2020 research on brain-computer interfaces (BCI) [3] gives a global outlook on BCI research, determines potentials and challenges, provides a sight of the state of art and BCI trends. A broad range of brain-controlled applications are envisaged by the year 2025, mostly in therapeutic treatment and health monitoring. Most of these applications are designed for handicapped people, with limited or nulled ability to communicate with the environment [4, 5]. Among the uses of BCIs are prostheses [6], wheelchairs [7], enhanced communication [8] virtual reality [9]. BCIs applications are also oriented for healthy people, for instance, assistant to steer a car [10], a humanoid robot [11], a drone [12] or an aircraft [13].

One of the fields in which BCI can be used, is aviation and astronautics. Potentials and challenges of applying BCI to aircraft control have been analyzed by several research programs [14]. Some attempts to use BCIs to steer drones or simulators have been made [12, 13]. Direct mental teleoperation of semi-automatic manipulators in the hostile external environment from within the safety of the spacecraft is consider as an example of how brain signals may be practically applied in space operations [15]. The work done by Coffey [15] highlights three main areas for BCIs application in astronautics: (a) alteration of interactions between the user and a steered device; (b) assessment of systems ergonomics and utility; (c) gathering of data for human performances assessment, particularity when performing tasks under extreme conditions.

This paper presents the first steps of a work focused on the evaluation of brain patterns that occur when an individual excites the brain to perform an action. The final goal of this project is to implement this method in a real time program that is capable of filtering the signals obtained by the brains measurement system; treating these signals to obtain the amplitude accumulation at the indicated frequencies and sending the control commands to a RPAS in order to be able to control it.

Based on different previous neurological studies, it is concluded that when a person wishes to perform an action, the part of the brain that is excited is the motor zone that resides in the front part of the head. This part of the brain will produce waves, known as "mu" or "μ" waves, which have different characteristics from the waves that occur when an individual is performing other kind of actions. The development of the project resides in the identification and processing of different brain patterns based upon the intended movement actions of an individual.

Initially, the study focuses on identifying the waves produced by measuring the energy of the brain when the individual is thinking about making movements of the right or left part of their motor system. After performing different tests, it has been shown that when the individual concentrates in making a right-hand movement the difference of the amplitude accumulation of these signals, in the frequency range between 8 and 12 Hz (frequencies associated with μ waves), is positive. Whereas on the other hand if this difference of values is negative, it implies that the individual is thinking of a movement related to his left zone of the motor system. The measurements are made through three sensors located in the central, left and right side of the head, thus obtaining two signals, which are measured simultaneously when the person is thinking of taking an action.

This paper discuss the theoretical background behind the study, the experimental set up, the instrumentation and signal processing used and the results and conclusions achieved.

This work has been carried out as a cooperative classroom project along a whole semester by students of the Aerospace Engineering Master program from the ETSIAE (Aeronautical and Space Technical Engineering School) at the UPM (Universidad Polytechnic de Madrid). This Masters' Degree is oriented at research and professional activities in the areas of aeronautics and space, particularly in the analysis and architectural design of air transport systems. Its main aim is research applied to the improvement in operational safety, efficiency and the capacity of air transport and its infrastructure, by using information technologies and the analysis of systems centered on the human factor.

2 Fundamentals

The command of a mobile device trough a BCI implies the translation of the sensorimotor rhythms (SMR), that reflects the subject's motor intention, into a command signal through real-time detection of motor imagery states, e.g. imagination of left and right hand movement.

During motor imagery, (observation, imagination or performance of a motion), activity is observed over the areas of the sensorimotor cerebral cortex. The motor cortex areas register activity in three frequency ranges, μ (8–12 Hz), β (18–30 Hz) and γ (30–200+ Hz) [16]. Signals detected in the encephalogram are described in Table 1.

When observing, imagine or performing a motion, SMR signal presents a power decrease, which is named event-related desynchronization (ERD). During relaxation periods after conclusion of the movement, SMR signal presents a power increase, which is referred to as event-related synchronization (ERS) [17]. ERD/ERS is more noticeable in the contralateral electrodes (Fig. 2).

The first motor imagery BCI was based on the detection of EEG ERD/ERS power changes on μ and β rhythms during imagination of left and right movements. Previous works have build BCIs that translate difference ERD/ERS between ipsilateral and contralateral electrodes into movements of the cursor on the screen [1].

Table 1. Signals detected in an encephalogram.

Alpha waves (α)	Mu (μ) waves	Beta waves (β)	Delta waves (δ)	Gamma waves (γ)
8–13 Hz	8–13 Hz	12–30 Hz	0.1–3.9 Hz	40–100 Hz
Occipital lobe Periods of relaxation	Motor cortex	Subject is awake and fully aware	The slowest waves	The fastest in the brain
Eyes closed but awake	Attenuated with movement or even with the intention of moving	Associated with alert, active or stress situations	Sleep deeply, without dreams or deep meditation	Associated with increased mental activity
Appropriate to identify the visual activity of a person	Of great interest to identify movements			

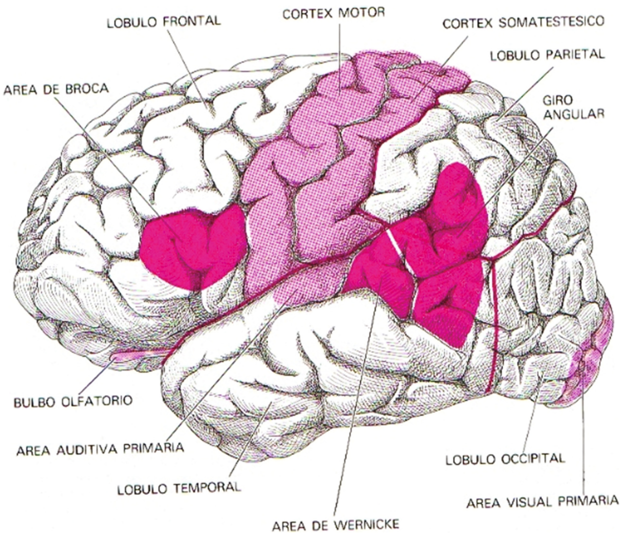


Fig. 2. Motor cortex area.

Process of steering objects by imaging movements has a great potential because is natural for the humans. However current algorithms for classifying motor imagery relay only on the ERD/ERS on the primary sensory-motor cortex. Because of that, usefulness of SMR-BCI-driven control depends on the individual's ability to imagine sensorimotor phenomena [18]; they lack of attention [19], and they need to be preceded by training sessions [1].

Complementarily motor imagery also activates the supplementary motor cortex area (SMA) [20]. The combination of ERD/ERS on the primary sensory motor cortex with activity in the supplementary motor cortex area (SMA), can bring potential improvements of SMR-BCI.

Wang et al. have observed a contralateral increased synchronisation of μ rhythms between the SMA and the primary motor cortex (M1) [21]. Consequently, the use of the signal in the supplementary motor cortex area (SMA), as reference, can augment the power difference between M1 areas. This facilitates the design of an electrode layout with two bipolar leads, C3-FCz and C4-FCz, that offers optimal performances in recognizing motor imagery states.

3 Configuration of the Electrodes

The electrodes layout compares the electrical activity in the left part of the brain with the electrical activity in the rigth part when the individual is thinking of making a movement (right or left).

The brain activity of the right region is related to the actions of the left side of the body. That is, if you are thinking of moving to the right, the predominant activity in the brain will be that of the left region.

The nomenclature and positioning of the electrodes in the trails follows the international 10–20 system. This is a standard method to locate scalp electrodes in an EEG test. It use ensures reproducibility, e.g. comparison of subject's studies over time and to each other. In this actual distances between contiguous electrodes are 10% or 20% of the total front–back or right–left distance of the skull. Figure 3 shows a schematic positioning of electrodes according to the guidelines for standard 10–20 electrode position nomenclature.

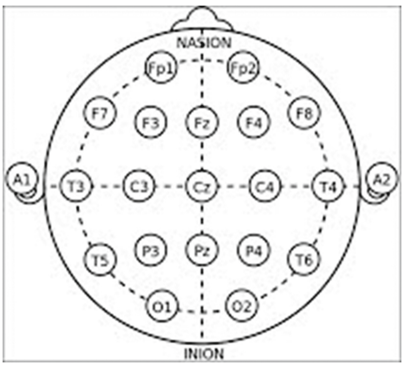


Fig. 3. International 10–20 electrode system.

For the placement of the electrodes, as SMA in synchronized with the ERD region, it is considered as a zero-phase reference. The electrode FCz is place in the SMA area,

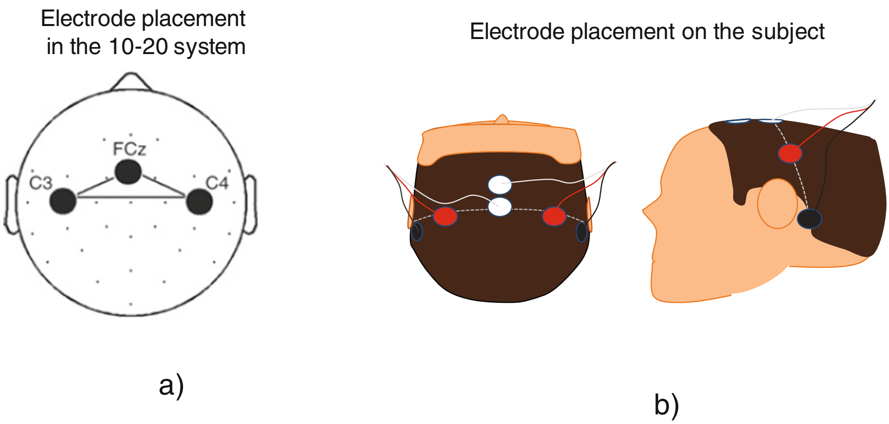


Fig. 4. Placement of three electrodes (C3, C4, and FCz) in the 10–20 system (a) and on the subject of the experiment (b).

while electrodes C3 and C4 correspond to the area M1. Using FCz as the reference electrode, as indicated in Fig. 4a, will make the power differences between M1 areas more substantial.

The Fig. 4b illustrate how the electrodes where configured on the subject of the experiment. In total 6 electrodes are placed symmetrically. Each side has a ground electrode under the ear (black electrode), a positive electrode on the side of the head (red electrode) and a negative electrode in the middle (white electrode).

4 Program Settings

To record the activity in both zones it is necessary to have at least two channels in the microprocessor, one for every three electrodes. The signals taken by channel one (CH1) will be those corresponding to the right side, while the signals recorded in channel two (CH2) are those corresponding to the left side.

The channels (CH1 and CH2) will be composed of the three electrodes already mentioned:

- Positive electrode (red wire).
- Negative electrode (white wire).
- Ground electrode (black wire).

Data collection was done with the program Biopac Student Lab PRO. Figure 5 shows how the two channels are configured to capture signals μ with a frequency range from 8 to 13 Hz.

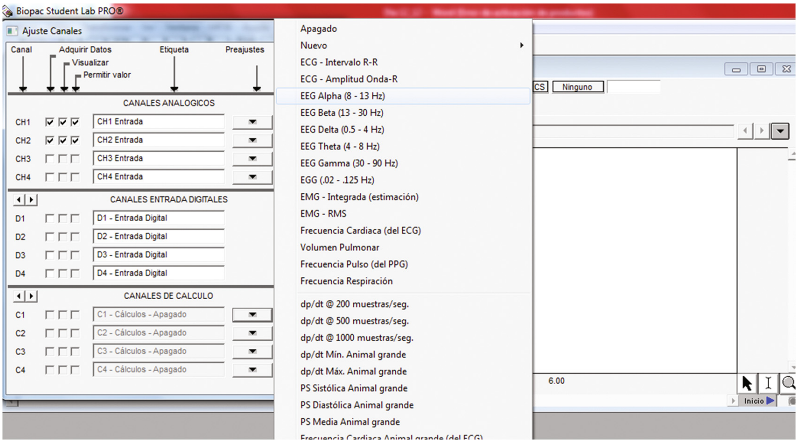


Fig. 5. SW configuration.

5 Population of the Experiment

The experiment was carried out by testing five subjects: three males and two females. The aim of the test was to detect the shape of the signals sent by the brain when an individual thinks of a right or left movement.

Two types of experiments were performed with eyes closed: the first consisting of movements with the right or left hand, and the second one consisting only in to think of the movement but without executing it.

Figure 6 shows an image taken from an individual during measurements with eyes closed.

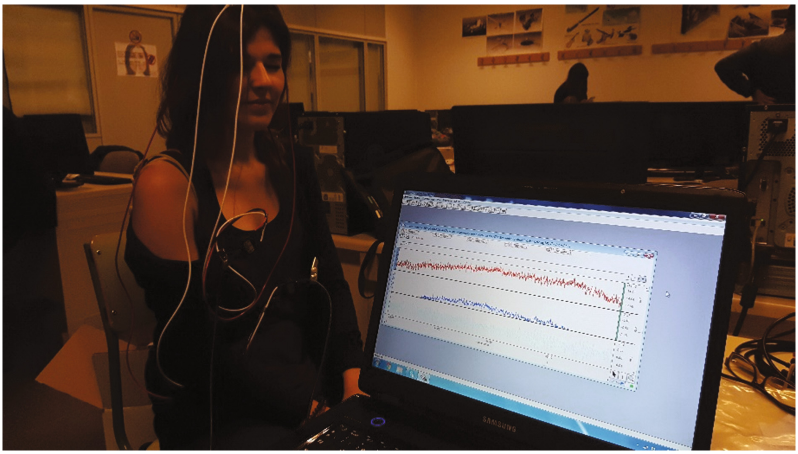


Fig. 6. Experimental set up with individuals.

Each of the tests is performed in different time intervals:

- 0–10 s: This interval corresponds to a process of relaxation with closed eyes in which the individual tries to abstract from any thought during a period of 10 s.
- 10–20 s: The subject begins to move or think about movement with eyes closed for an interval of 10 s.

This sequence is repeated 3 times in each test.

Figure 7 present an example of the signals obtained for one of the measurements.

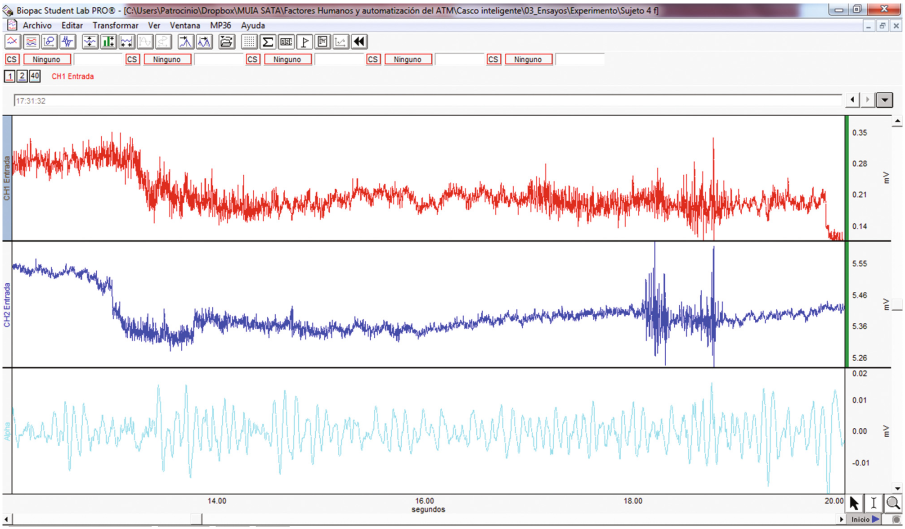


Fig. 7. Example of signals measured.

6 Data Processing and Results

Signals processing has been performed using MATLAB R2013a. The feature extraction and classification algorithm process a given vector that contains the readings of the electrodes, to discern whether the subject is performing actions with the right part of his body or with the left. Two different codes have been developed.

Code 1, illustrated on Fig. 8, follows the next steps. First the signals of each of the channels electrodes are recorded. Second, the intensity readings from the two channels are stored in a vector. A discrete Fourier transform is realized in order to know the frequencies that make up the signal obtained as well as the amplitude associated to each one of them. Within the frequency spectrum, those between 8 and 13 Hz are selected.

Finally, the amplitudes of the signals are subtracted. If the result of this difference is negative, it means that the right side of the brain is being used more indicating that the subject wants or is moving the left arm. If, on the other hand, the result of this difference

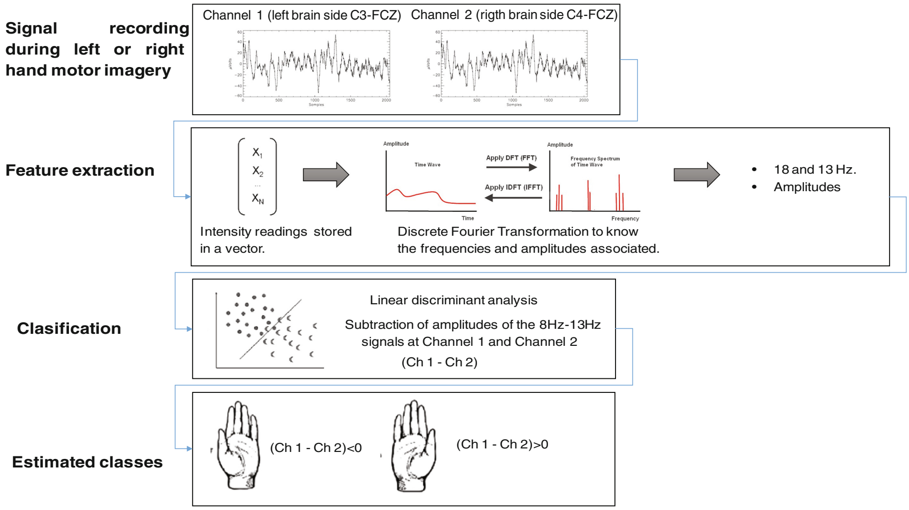


Fig. 8. Code 1.

is positive, it means that the subject is using more of the left side of the brain and, therefore, indicates that the subject wants or is moving the right arm.

Code 2 illustrated on Fig. 9, follows the next steps. The signals recorded in a first steps, are subsequently filter with a band pas filter that cut-off frequencies higher than 12 Hz and lower than 8 Hz.

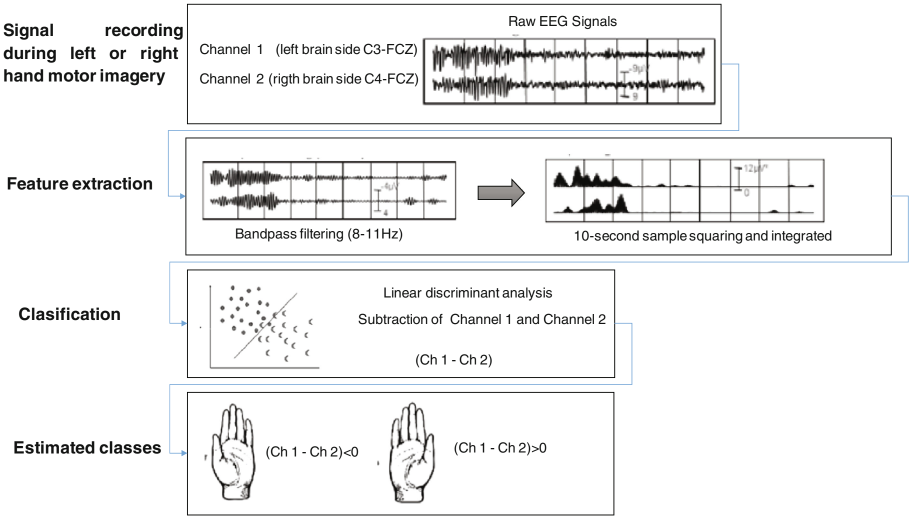


Fig. 9. Code 2.

Next, the total signal is separated in the different samples corresponding to the 10-second tests. Finally, the intensity of each sample is squared and integrated over time. Finally by subtracting the intensity of channels 1 and 2 it can be detected on which side of the brain there is greater cerebral intensity. If there is a greater signal strength in the right channel, it corresponds to the move/think of the left, whereas if the signal strength is greater in the left channel, it is intended to move/think left.

Code 1 has shown positive results in 91% of the cases, with a 90% success rate in the trials corresponding to men, and 92% in the trials corresponding to the women who participated in the study.

The code 2 employed showed a success rate of 94.4%, with a 96.6% success rate in the trials corresponding to men, and 92.6% in the trials corresponding to the women who participated in the study.

7 Conclusions and Further Work

This paper have presented the first step in the development of a BCI for RPAS command and control as developed by a group of 15 students of the Aerospace Master of the UPM. The paper have covered the theoretical background behind the study, the experimental set up, the instrumentation and signal processing used and the results and conclusions achieved.

The experience have confirmed the initial idea that a simple electrode set up can capture and discriminate motor imagery, particularly right and left movement intentions.

Students of the Aerospace Master of the UPM will further continue with the development of the project which final aim is to develop a Brain Computer Interface to control and command an RPAS.

References

1. Wolpaw, J., Birbaumer, N., McFarland, D.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. **113**, 767–791 (2002)
2. Pfurtscheller, G., Allison, B., Bauernfeind, G., Neuper, N.: The hybrid BCI. Front. Neurosci. **113**, 4–30 (2010)
3. Brunner, C., Birbaumer, N., Blankertz, B.: BNCI Horizon 2020: towards a roadmap for the BCI community. BCI J. (2015)
4. Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T.M.: Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. Neurology **64**(10), 1775–1777 (2005)
5. Marchetti, M., Priftis, K.: Brain-computer interfaces in amyotrophic lateral sclerosis: a metanalysis. Clin. Neurophysiol. **126**(6), 1255–1263 (2015)
6. Pfurtscheller, G., Guger, C., Neup, C., Krausz, G., Müller, G.: Brain oscillations control hand orthosis in a tetraplegic. Neurosci. Lett. **292**(3), 211–214 (2000)

7. Huang, D., Qiank, D., Fei, W., Jia, X., Chen, O.: Electroencephalography (EEG)-based brain-computer interface (BCI): a 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control. IEEE Trans. Neural Syst. Rehabil. Eng. **20**(3), 379–388 (2012)
8. Perdiki, S., Leeb, R., Williamson, J., Ramsay, A., Tavella, M.: Clinical evaluation of BrainTree, a motor imagery hybrid BCI speller. J. Neural Eng. **11**(3), 036003 (2014)
9. Leeb, R., Lee, F., Keinrath, C., Scherer, R., Bischof, H.: Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. IEEE Trans. Neural Syst. Rehabil. Eng. **15**(4), 473–482 (2007)
10. Bi, L., He, T., Fan, X.: A driver-vehicle interface based on ERD/ERS potentials and alpha rhythm. In: IEEE International Conference on Systems, Man and Cybernetics Systems, Man and Cybernetics (SMC), San Diego (2014)
11. Bell, C., Shenoy, P., Chalodhorn, R., Rao, R.: Control of a humanoid robot by a noninvasive brain–computer interface in humans. J. Neural Eng. **5**(2), 214 (2008)
12. LaFleur, K., Cassady, K., Doud, A.: Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface. J. Neural Eng. **10**, 4 (2013)
13. Fricke, T., Zander, T., Gramann, K., Holzapfel, F.: First Pilot-in-the-Loop Experiments on Brain T. Control of Horizontal Aircraft Motion. de Deutscher Luft- und Raumfahrtkongress, Augsburg (2014)
14. Holzapfel, T., Fricke, T.: Applying Brain Machine Interfaces to Aircraft Control. de The Man Between Automation, Competence and Responsibility (DGLR Report 2014-01), Bon, Special print from M. Grandt and S. Schmerwitz (2014)
15. Coffey, E., Brouwer, A., Wilschut, E., van Erp, J.: Brain–machine interfaces in space: using spontaneous rather than intentionally generated brain signals. Acta Astronaut. **67**, 1–11 (2010)
16. Ja, P.: The functional significance of mu rhythms: translating “seeing” and “hearing” into “doing”. Brain Res. Rev. **50**(1), 57–68 (2005)
17. Pfurtscheller, G., Lopes da Silva, F.: Event-related EEG = MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. **110**, 1842–2857 (1999)
18. Vuckovic, O.B.: Using a motor imagery questionnaire to estimate the performance of a brain-computer interface based on object oriented motor imagery. Clin. Neurophysiol. **124**(8), 1586–1595 (2013)
19. Curran, S.M., Stokes, M.J.: Learning to control brain activity: a review of the production and control of EEG components for driving Brain-Computer Interface (BCI) systems. Brain Cognit. **51**(3), 326–336 (2003)
20. Roland, P.E., Larsen, B., Lassen, N.A., Skinhoj, E.: Supplementary motor area and other cortical areas in organization of voluntary movements in man. J. Neurophysiol. **1**, 118–136 (1980)
21. Wang, Y., Hong, B., Gao, X., Gao, S.: Phase synchrony measurement in motor cortex for classifying single-trial EEG during motor imagery. In: Proceedings of 28th International IEEE EMBS Conference, New York, USA (2006)

Human-System Interactions: Practical Applications

A Comparative Analysis of Human-Mediated and System-Mediated Interruptions for Multi-user, Multitasking Interactions

Nia Peters[1(✉)], Griffin Romigh[2], George Bradley[3], and Bhiksha Raj[1]

[1] Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA
nbradley@ece.cmu.edu, bhiksha@cs.cmu.edu
[2] Air Force Research Laboratory, Wright-Patterson Air Force Base, Dayton, OH 45433, USA
griffin.romigh@us.af.mil
[3] Walden University, 1000 S Washington Ave., Minneapolis, MN 55401, USA
george.bradley@waldenu.edu

Abstract. When to send system-mediated interruptions within multitasking environments has been widely debated in the development of interruption management and notification systems primarily from the perspective of single-user interactions. Scholars illustrate that task structure is a useful predictor in determining when to send interruptions. However, these works do not address when to send system interruptions in multi-user, multitasking scenarios and do not address predictors of interruptibility within communication tasks. This paper addresses the issue of predicting interruptibility within multi-user, multitasking communication interactions with special attention to leveraging human interruption techniques as predictors of interruptibility. Specifically, in our project, we will be looking at how task structure and speech information influence human interruption strategies. These strategies could potentially be modeled and integrated into interruption management systems for multi-user, multitasking interactions. We will discuss human interruption strategies and juxtapose them against random interruption strategies, to reveal an intelligent technique for modeling interruptions. We argue that humans use task structure and speech cues to make more informed decisions about when to interrupt that are distinct from more random strategies. An analysis of variance showed that the effect of the interruption strategy (human versus system) on the proximity of an interruption to task boundaries was significant for start and end boundaries, $F(2,5938) = 17.46$, $p = 0.001$, $F(2,5938) = 7.46$, $p = 0.006$ respectively. This project sheds light on the use of task structure as a predictor of interruptibility within multi-user, multitasking environments via techniques used by human interrupters.

Keywords: Interruption management · Collaborative communication · Human factors · Human-human-computer interactions

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_32

1 Introduction

Within human-human-machine collaborative communication tasks, it is imperative that machines adhere to appropriate communication strategies that do not hinder the overall task goals. One form of communication within these tasks are *interruptions*. The work in interruption and notification management system development primarily explores interruptions in human-machine tasks from the perspective of single-user, multitasking environments. This research area is motivated by the reality that as users increasingly multitask among proactive systems, their tasks are being interrupted more often. Though proactive delivery of information can benefit users, studies show that interrupting primary tasks can negatively impact productivity [1–4] and affective state [5, 6]. Within these contexts there have been proposed methods for information dissemination and intelligent system-mediated interruptions. There is empirical research dedicated to manipulating time of the delivery [1–3] of system-mediated interruptions [7] in multi-task environments [8]. There is also literature that explores immediate interruption or notification dissemination [9–11] within dual-task scenarios. Studies have shown that delivering interruptions at random times can result in a decline in performance on primary tasks [1, 9, 11–13]. Other studies show similar results [6, 9, 14, 15] and the differences in cost of interruptions are typically attributed to differences in workload at the point of interruption [1]. Additionally, studies have illustrated that interrupting users engaged in tasks has a considerable negative impact on task completion time [2, 9, 12, 16–18].

There is an immense amount of literature that recommends breakpoints within task execution as appropriate points of interruptibility. A breakpoint can be defined as a time instance between two moments of task execution. Studies have shown that deferring delivery of notification until a breakpoint is reached can meaningfully reduce costs of interruptions [1, 2, 5, 19]. Interrupting tasks at random moments can also cause users to take up to 30% longer to resume the tasks, commit up to twice the errors, and experience up to twice the negative affect than when interrupted at boundaries [1, 5, 20]. In the work done by Miyata and Norman [21], the timing of interruptions is also defined by the structure of a task which is presented as having a beginning, middle, and end. If interruptions occur between any of these points in a task or subtask, they are deemed appropriate.

The aim of this work is to augment this area of research by investigating appropriate interruption timings and their effect within multi-user, multitasking environments. In such environments, an interruption is a machine's action to deliver information relevant to a human-machine task that may halt an orthogonal human-human task. Prior to developing an interruption management and notification system for multi-user multitasking environments, it is necessary to determine what constitutes interruptibility within these interactions. Since an extensive amount of work has been proposed in modeling task structure to predict appropriate interruption points in single-user, multitasking environments, we aim to explore this claim with respect to multi-user, multitasking environments.

If we define an *appropriate interruption* as an interruption decision made by a human, then it is necessary to determine whether humans make more informed interruption decisions and discriminate these from random interruption strategies with respect to task structure and speech information. We compare two classes of interruption decisions, human-mediated (HM) and system-mediated (SM) interruptions, based on when these decisions occur with respect to task boundaries and speech characteristics preceding the interruption decision. We hypothesize HM interruptions will occur closer to the start and end of the interrupted task compared to SM interruptions. This hypothesis assumes that humans wait for the logical beginning or conclusion of a task to interrupt compared to a system interruption that have more random characteristics. Additionally, we hypothesize that the speech characteristics prior to HM interruptions will be illustrative of a lag in communication in comparison to speech characteristics prior to a SM interruptions, which could be a bit more random.

In Sect. 2 the methods and procedures are presented. In Sect. 3 the results are presented and suggest that humans interrupt closer to the beginning and end of a task compared to a system, specifically closer to end task boundaries. Finally, in Sect. 4 we conclude that since human interruption strategies are different from random interruption strategies with respect to the task structure, modeling task structure to predict interruptibility may be useful in the design of an intelligent interruption system that can support alleviating the cost associated with system-mediated interruptions in multi-human-machine multitasking environments.

2 Methods and Procedures

We simulate a collaborative human-human-machine communication task with a primary human-human task and secondary human-machine task where interruptions are only relevant to the human-machine task. A HM interruption is an interruption decision made by a human listening to the primary task with information relevant to the secondary task. A SM interruption occurs every 15 s unless impeded by a human interruption. Both HM and SM interruptions are presented via a synthesized voice and only the interruption timing is compared. A more detailed description of the entire experimental design and corpus is described in [22].

2.1 Data

From our data collection, we leverage the audio and system-state information from all sessions of each interrupter. Each session includes two participants working on a collaborative matching task and a third participant (interrupter) sending information to both participants regarding a secondary monitoring task. Hand-crafted prosody and task-breakpoint features are processed from the audio and system-state information across all sessions for each interrupter.

Prosodic Features. To model the difference between HM and SM interruptions based on how people speak, we use the acoustic content prior to an interruption. The intuition behind these features is that the human may be listening for a pause in the current conversation which can be illustrated via declining energy and pitch contours. A 750 ms window of audio is extracted prior to the interruption. Using this window segment, we partition the audio into 25 ms frames with 10 ms overlap. For each window segment, we calculate the energy, pitch using both the autocorrelation function (ACF) and subharmonic summation (SHS) methods, probability of voicing (ACF and SHS), and loudness of each frame. From the prosody of these frames we extract the linear regression coefficients for each prosody category. This results in a 12-dimensional feature vector.

Task Breakpoint Features. The intuition behind these features is that humans interrupt when they perceive a task will begin or end. A start task boundary is a timestamp associated with the beginning of the primary human-human task and the end task boundary is the timestamp when both users have pressed *DONE* or concluded the primary human-human task. We compute the absolute distance between an interruption and the start task boundary (d_n) and the absolute distance between an interruption and the end task boundary (d_{n+1}). From d_n and d_{n+1} we calculate the d_{min}, d_{max}, d_{diff}, d_{per}, and time of completion (TOC) which is the overall task duration for that interruption as indicators how close interruptions are sent relative to a task boundary.

$$d_{\min} = \min(d_n, d_{n+1})$$

$$d_{\max} = \max(d_n, d_{n+1})$$

$$d_{\text{diff}} = d_{\max} - d_{\min}$$

$$d_{\text{per}} = \frac{d_{\min}}{d_{\min} + d_{\max}}$$

3 Methods

From the data collection, there are 18 human interrupters. Each interrupter participated in a total of 10 experimental blocks on average. Each experimental block lasted 15 min. From the total number of blocks, there is approximately 45 h of audio data in addition to system state information such as subtask start and end times, collaborative task scores, and push-to-talk start and end timestamps. A human-mediated (HM) interruption point is a timestamp associated with a human interrupter sending an interruption relevant to a secondary "human-system task" while interrupting the primary "human-human task." Similarly, a system-mediated (SM) interruption is the timestamp associated with a human's inability to make an interruption decisions so an interruption is automatically sent after a pre-defined time elapse. In total, there are 5939 interruptions across all interrupters (4847 HM and 1092 SM). With the 80-20% split of data, we down sampled the number of HM interruptions to achieve a 65-35% split to avoid prior bias during the modeling process. The final number of interruptions is 3207 (2115 HM and 1092 SM) and only these were used for feature extraction and classification.

Three methods were used to compare HM and SM interruptions.

1. To get a global interpretation of how task boundary information was used by humans and the system, a one-way analysis of variance (ANOVA) was conducted to evaluate the relationship between how close an interruption is to a task breakpoint and the interruption dissemination method. The independent variable is the interruption method: human-mediated and system-mediated. The dependent variable is the absolute distance in time an interruption is from the start of a task and the absolute distance in time an interruption is from the end of a task.
2. To access what information streams are salient in individual human interruption strategies compared to system-mediated interruption strategies, we use a two-sample t-test feature selection algorithm across all prosodic and task boundary features for each class (HM versus SM). If the p-value for a feature is less than 0.005, we extract the feature.
3. Finally, we use the selected features and a Naïve Bayes classifier to model the two classes. We used a hold-one-out cross-validation method to generalize the model and evaluate the performance of the model based on the accuracy measured by Area Under the Curve (AUC) which illustrates the trade-off between the true-positive and false-positive rate.

4 Results and Discussion

From Table 1, on average, human-mediated interruptions are closer to the start task boundary. The ANOVA showed a significant difference at the 0.05 level for that start time boundaries, $F(2, 5938) = 17.46$, $p = 0.001$. Also from Table 1, on average, human-mediated interruptions are closer to the end boundary. The ANOVA showed a significant difference at the 0.05 level for the end time boundaries, $F(2, 5938) = 7.46$, $p = 0.006$. These results additionally illustrate that on average human interruptions are closer to end task boundaries in comparison to start task boundaries.

Table 1. System-mediated versus human-mediated distance comparison from start and end boundaries

	Interruption type	N	Mean	Std. error mean
Start task boundary	System-mediated	1092	20.54	0.34
	Human-mediated	4847	15.36	0.14
End task boundary	System-mediated	1092	12.81	0.31
	Human-mediated	4847	9.60	0.13

To get a better illustration of where humans are interrupting relative to the end boundary, a histogram of all human interruption distances from the end task boundary was plotted. Figure 1 illustrates that the distance between most human interruptions and the end task boundary is between 0–5 s. These human interruption times could very well correspond to the period between when a user hears a grounding confirmation

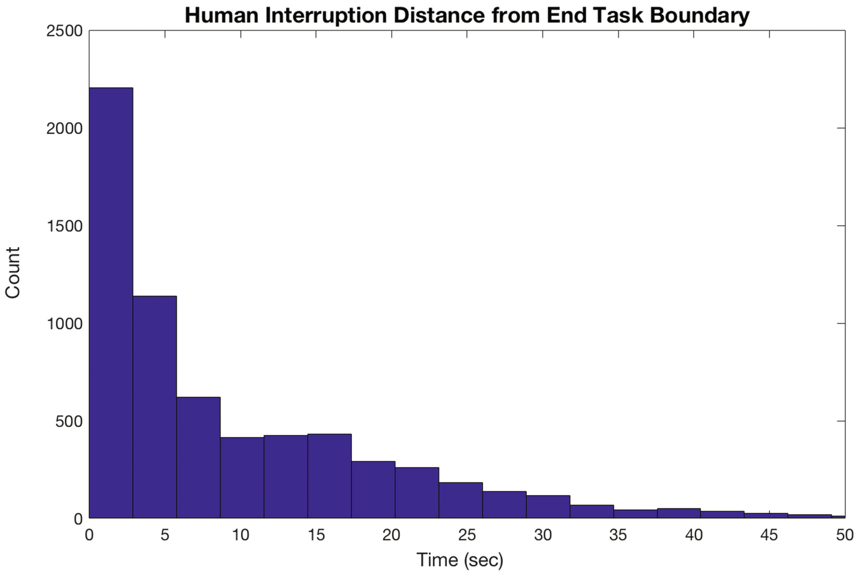


Fig. 1. Histogram of human interruption distance from end task boundary

of “ok got it” or “done” and the actual clicking of the button *DONE* where the system latency allows for an interruptible point in the exchange. This result is very useful in giving us more information about human interruption strategies and task structure.

Using the feature selection algorithm results, we calculate the percentage of occurrences a feature was selected across all interrupters. TOC was selected 100% of the time, d_{min} was selected 94%, and d_{diff} was selected 89% in comparison to prosody features where probability of voicing (ACF) and F_0 (ACF) features were selected only by 44% of the interrupters. The TOC feature was possibly salient because a system-mediated interruption was sent every 15 s unless preempted by a human interruption. If a task exceeds 15 s, users may let the system send an interruption. This may explain why longer task times are a discriminator of system and human interruptions. These results also give some indication that the subjects made decisions closer to task boundaries than the system did.

Since task boundaries seem to be an informative stream to evaluate human interruptions, we wanted to know how much more information prosody could contribute for discriminating human-mediated and system-mediated interruptions. Figure 2 compares the task boundary only model with a prosody/task boundary model.

If we define an AUC of 0.65 as the random baseline associated with the model detecting all interruptions as human-mediated, then across all interrupters both models outperform the baseline model. The average AUC over all the interrupters is 0.76 for the task boundary only model which is a 16.9% increase over the baseline model and augmented to 0.78 for the combined models which is a 20% increase over the baseline model. These results give some indication that prosody may offer some additional information, but not a significant amount with only a 2.6% increase in performance

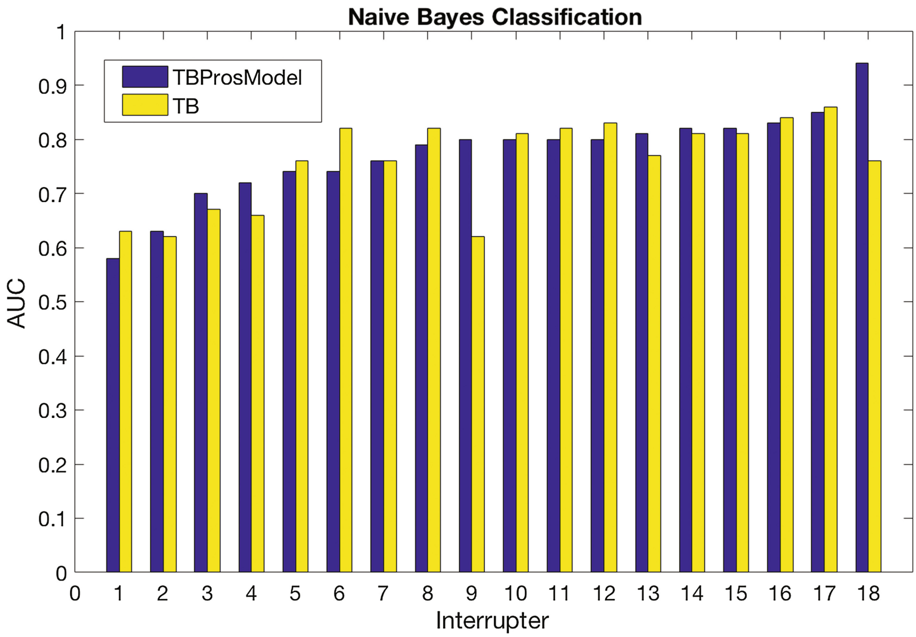


Fig. 2. Task boundary vs. task boundary and prosody model

over the task boundary only model. Speech characteristics did not seem to be useful in discriminating human and system interruption strategies. This could potentially be attributed to the stream of acoustic data we used to extract features prior to interruptions. We extracted features from the entire audio stream prior to an interruption rather than the push-to-talk audio information. If users are talking to themselves or rehearsing other segments of the task, this is captured within the audio stream. This way of extracting features is also not consistent with what humans would hear when listening in on the primary task. In continuing work, we focus on only the push-to-talk audio information stream and evaluate human interruption strategies from this perspective.

What these results suggest is that within multiuser, multitasking environments, humans interrupt closer to task boundaries and their interruption techniques are distinct from less intelligent interruption strategies. What does this mean? Humans are not random when interrupting in these interactions and if we wanted to replace that human with a system, we could potentially model human interruption strategies to build an intelligent interruption and notification management system for multiuser, multitasking collaborative communication interactions. These results are consistent with current literature in interruption and notification management systems for single-user, multitasking environments that propose task boundaries as predictors of interruptibility. Additionally, although we were not able to illustrate speech information as a predictor of interpretability, further exploration of speech will be explored in further work since the proposed interruption management systems could be applied to communication tasks.

5 Conclusion

In conclusion, we could discriminate human-mediated and system-mediated interruption strategies via task structure information within a simulated multi-user, multitasking environment. Within our setup, we illustrate that humans not only interrupt closer to task boundaries, but also tend to interrupt closer to the end of a task. This conclusion is consistent with previous work that proposes task structure as indicator of appropriate interruptibility points within single user, multitasking interactions. We can use our results and the current literature in interruption and notification management systems to make a case for developing an intelligent interruption and notification management system for multi-user, multitasking environments by modeling task breakpoints. In future work we would like to design an intelligent interruption system based on task boundaries, integrate this system into multi-user, multitasking environments and evaluate users' performance on primary tasks based the proposed intelligent interruption system's strategies.

References

1. 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**, 685–708 (2006)
2. Czerwinkski, M., Cutrell, E., Horvitz, E.: Instant messaging: effects of relevance and timing. In: People and Computers XIV: Proceedings of HCI. British Computer Society (2000)
3. Monk, C.A., Boehm-Davis, D.A., Trafton, G.J.: The attentional costs of interrupting tasks performance at various stages. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2002)
4. Cutrell, E.M., Horvitz, E.: Effects of instant messaging interruptions on computing tasks. In: CHI (2000)
5. Adamczyk, P.D., Bailey, B.P.: If not now, when? The effects of interruption at different moments within task execution. In: Proceedings of the SIGCHI Conference on Human Factors in Computer Systems (2004)
6. Zijlstra, F., Roe, R., Leonora, B., Krediet, I.: Temporal factors in mental work: effects of interrupted activities. J. Occup. Organ. Pyschol. **72**, 163–185 (1999)
7. McCrickard, S.D., Chewar, C.M., Somervell, J.P., Ndiwalana, A.: A model for notification systems evaluation assessing user goals for multitasking activity. ACM Trans. Comput. Hum. Interact. (TOCHI) **10**, 312–338 (2003)
8. McFarlane, D.C., Latorella, K.A.: The scope and importance of human interruption in human–computer interacting design. Hum. Comput. Interact. **17**, 1–61 (2002)
9. Czerwinski, M., Cutrell, E., Horvitz, E.: Instant messaging and interruption: influence of task type on performance. In: Proceedings of OZCHI 2000 Conference (2000)
10. Dabbish, L., Kraut, R.E.: Controlling interruptions: awareness displays and social motivation for coordination. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (2004)
11. Latorella, K.A.: Investigating interruptions: implications for flightdeck performance (1996)
12. Kreifeldt, J.G., McCarthy, M.E.: Interruption as a test of the user–computer interface. In: JPL Proceedings of the 17th Conference on Manual Control (1981)

13. Robinstein, J.S., Meyer, D.E., Evans, J.E.: Executive control of cognitive processes in task switching. J. Exp. Psychol. Hum. Percept. Perform. **27**, 763 (2001)
14. Altmann, E.M., Trafton, J.G.: Task interruption: resumption lag and the role of cues. In: CogSci (2004)
15. McFarlane, D.A.: Coordinating the interruption of people in human–computer interaction. In: Human–Computer Interaction, INTERACT (1999)
16. Cutrell, E., Czerwinski, M., Horvitz, E.: Notification, disruption, and memory: effects of messaging interruptions on memory and performance (2001)
17. McFarlane, D.C.: Interruption of People in Human–Computer Interaction: A General Unifying Definition of Human Interruption and Taxonomy. Office of Naval Research, Arlington (1997)
18. Bailey, B.P., Iqbal, S.T.: Understanding changes for interruption management. In: ACM Transactions on Computer–Human Interaction (TOCHI) (2008)
19. Isbal, S.T., Bailey, B.P.: Leveraging characteristics of task structure to predict costs of interruption. In: Proceedings of CHI (2006)
20. Isbal, S.T., Bailey, B.P.: Investigating the effectiveness of mental workload as predictor of opportune moments for interruption. In: CHI (2005)
21. Miyata, Y., Norman, D.A.: Psychological issues in support of multiple activities. In: User-Centered System Design, Hillsdale
22. Peters, N.S., Romigh, G., Raj, B., Bradley, G.: When to interrupt: a comparative analysis of interruption timings within collaborative tasks. In: Advances in Human Factors and System Interactions, Orlando (2017)

Recognition of Physiotherapeutic Exercises Through DTW and Low-Cost Vision-Based Motion Capture

Yves Rybarczyk[1,2,3,4(✉)], Jan Kleine Deters[5], Arián Aladro Gonzalo[1,2], Danilo Esparza[1,2,6], Mario Gonzalez[1,2], Santiago Villarreal[1,2], and Isabel L. Nunes[3,4]

[1] Intelligent and Interactive Systems Lab (SI2 Lab), Universidad de Las Americas, Av. de los Granados E12-41 y Colimes esq., Quito EC170125, Ecuador
{yves.rybarczyk, arian.aladro, danilo.esparza, mario.gonzalez, santiago.villarreal}@udla.edu.ec
[2] Research Laboratory in Physiotherapy and Human Movement, Universidad de Las Americas, Av. de los Granados E12-41 y Colimes esq., Quito EC170125, Ecuador
[3] Department of Electrical Engineering, UNINOVA/CTS, Nova University of Lisbon, Lisbon, Portugal
{y.rybarczyk, imn}@fct.unl.pt
[4] Department of Mechanical and Industrial Engineering, UNIDEMI, Monte de Caparica, Portugal
[5] University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
j.kleinedeters@student.utwente.nl
[6] Pontificia Universidad Católica del Ecuador, Quito, Ecuador

Abstract. Telemedicine is a current trend in healthcare. The present study is part of the ePHoRt project, which is a web-based platform for the rehabilitation of patients after hip replacement surgery. To be economically suitable the system is intended to be based on low-cost technologies, especially in terms of motion capture. This is the reason why the Kinect-based motion tracking is chosen. The paper focuses on the automatic assessment of the correctness of the exercises performed by the user. A Dynamic Time Warping (DTW) approach is used to discriminate between correct and incorrect movements. The classification of the movements through a Naïve Bayes classifier shows a very high percentage of accuracy (98.2%). Models are built for each individual and reeducation exercise with only few attributes and the same accuracy. Due to these promising results, the next step will consist of testing the algorithms on patients performing the exercises in real time.

Keywords: Telerehabilitation · Movement assessment · Dynamic Time Warping · Machine learning · Kinect-based motion tracking

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_33

1 Introduction

A current trend in medicine is home therapy systems. This concept consists of enabling patients to carry out part of the rehabilitation at home and to communicate through the Web the evolution of the recovery process. Thus, health professionals can proceed with a remote monitoring of the patient's performance and an adaptation of the treatment accordingly. In Ecuador, telerehabilitation systems are still little developed. However, this technology could bring several advantages for the individual and the society in terms of healthcare (improvement of the recovery process by the possibility to perform rehabilitation exercises more frequently), economy (reduction of the number of medical appointments and the time patients spend at the hospital), mobility (diminution of the transportation to and from the hospital) and ethics (healthcare democratization and increased empowerment of the patient).

By taking into account these considerations, the project ePHoRt proposes to develop a Web-based platform for home motor rehabilitation. The tool is developed for patients after hip arthroplasty surgery. This orthopedic procedure is an excellent case study, because it involves people who need a postoperative functional rehabilitation program to recover strength, function and joint stability. In addition, due to the condition of the patients, it is difficult to carry them to the hospital. The project intends to tackle three main issues. First, the motion capture technology must be low-cost, in order to be used worldwide. Second, the system should automatically detect the correctness of the executed movement, in order to provide the patient with real time feedback. Third, new computational approaches have to be researched, in order to promote patient's motivation to regularly complete the rehabilitation tasks (e.g., by the use of affective computing and serious games paradigms to detect difficulties and stimulate effort in the patients, respectively).

This paper focuses on the two first parts of the project, which are the development of a program to automatically assess the correctness of the movements by using an information provided by a low-cost motion capture device. The manuscript is divided into five sections. The first part is a state of the art in terms of technologies for home rehabilitation and telemedicine. The second is a description of the architecture that will support the platform. The third exposes the methodology used to evaluate the movement. The fourth part consists of a presentation of the results. Finally, some conclusions and perspectives are drawn up regarding the midterm progress of the project.

1.1 State of the Art

In any kind of rehabilitation, repeated exercises of an impaired limb maximize the chances of recovery [1]. In practice, medical and economic situations limit the number of therapeutic sessions the patient can take at the hospital or medical centers. This fact justifies the increased trend in developing telehealthcare systems to enhance home therapy. However, this new approach of healthcare copes with several other obstacles. First, the used technology must be both reliable and affordable in order to significantly reduce the costs. Second, the system has to provide a rigorous and real time monitoring of the patient's movements to make sure that the rehabilitation protocol is properly executed. For instance, it has to detect the patients' tendency to compensate their diminished limb with other functional part of their body, which makes rehabilitation progress slower [2]. Third, providing a therapeutic framework at home does not automatically implicate a speedy recovery, because of the lack of motivation among patients to exercise for sustainable period of time.

Home-based rehabilitation has gained prominence over the recent years through the development of exercise platform [3], virtual reality-based system [4], gaming console [5] and the widely popular Kinect camera-based system [6]. For the purpose of this project it is necessary to make a survey on three critical points regarding home therapy. First, the existing technologies used in telemedicine for patient's rehabilitation. Second, systems and methods developed to perform a real time analysis and recognition of the movement. Third, the main approaches implemented to enhance motivation in patients to regularly complete the exercises. In the context of this paper that focuses on the motion capture and assessment parts of the project, the state of the art is based on the first two points, only.

Motion Capture Systems. Motion capture is used in a wide range of areas, from the entertainment industry (e.g., three-dimensional animation) [7] to scientific studies, such as biological motion analysis [8]. However, it can be an expensive and complex technology that it is not always usable at home. For instance, one of the most famous motion capture systems for professional use is the Vicon motion system. This kind of technology is based on light-reflecting markers positioned on the individual body and infrared-camera sensors that enable a precise analysis of movements. The main limitations of this tool are its high-cost and the fact that it only works in very controlled conditions (usually in laboratories). A quite more affordable equipment is the popular Kinect camera from Microsoft. This system has several advantages over the other technologies. It is a cheap motion capture device (around 100USD). It is quite easy to install and use at home. And it provides an automatic reconstruction of the three-dimensional coordinates of the body main joints, with a reasonable spatial and temporal resolution [9]. A study comparing the Vicon and the Kinect system for measuring movement in people with Parkinson's disease shows that the Kinect has the potential to be a suitable equipment to capture gross spatial characteristics of clinical relevant movements [10]. Also, a computer vision approach based on the use of CCD webcams can be implemented, but the apparatus usually requires two optical sensors in order to calculate the 3D-coordinates by triangulation and it needs the development of additional image processing algorithms [11]. An attempt to reconstruct body postures and 3D movements from monocular video sequences was proposed by [12], but the true accuracy and applicability for real-time tele-rehabilitation are not yet demonstrated. Finally, an alternative to vision-based monitoring is a wearable system. However, the use of complicated wearable devices makes it tiresome and, consequently, tends to decrease the effectiveness of rehabilitation exercises [13]. A preliminary experiment for the ePHoRt project shows a significant correlation between the accuracy of a Kinect and an accelerometer sensor, which demonstrates that a vision-based motion capture is an excellent alternative to an inertial-based motion capture [14].

Movement Recognition and Assessment. Human action recognition is a very challenging topic. It presents more degrees of freedom with respect to system design and implementation when compared to language processing [15]. Different approaches are used regarding the sensor technologies and computational algorithms. The first kind of sensors are wearable devices, which are based on pervasive and mobile computing [16, 17]. A less invasive technique consists of using a vision-based recognition. The Kinect camera has largely contributed to the growth of this approach, due to the fact that it facilitates the extraction of the pertinent features for gesture recognition [18, 19]. Movement recognition has to be carried out through a process flow that usually involves raw data recording and filtering, feature extraction and selection, and classification by the use of machine learning models. Techniques of time series analysis are applied to compare the similarity between two temporal sequences, such as Dynamic Time Warping [20]. A wide range of classifiers have been used for action recognition in the last decades [21]. One of the most successful methods to achieve recognition of daily activity is the discriminative approach. This classification is based on the construction of decision boundaries in the feature space, specifying regions for each class. The main classifiers that implement this type of method for activity recognition are the k-Nearest Neighbour [22], Support Vector Machines [23], the Naïve Bayes [24], and C4.5 Decision Tree [25]. The last two are by far the most popular algorithms, because they generally enable a high classification accuracy [26] and the hierarchical tree representation makes the model easier to read than a sequence of rules [27].

1.2 Platform Architecture

The ePHoRt project is a Web application based on a three layers Client-Server architecture (Fig. 1). The client layer (browser) will be developed in JavaScript by the use of two possible frameworks: jQuery or AngularJS. The application layer (domain server) will be implemented either in Django framework from Python or Java framework. The data server layer (database) will be developed in MySQL.

A different user interface will be implemented for each of the three kinds of user: patient, physiotherapist and physician. After login, the patients will have an access to the programmed exercises they have to achieve. The patient's interface will be composed with two dynamic frames: (i) one to display a video example of the exercises to be completed and (ii) another to display a 3D-avatar that will mimic in real-time the patient's movements captured by the Kinect. The patient will receive a real-time feedback regarding the correctness of the movements, thanks to the assessment module, and game scores. In addition, questionnaires will be available on this interface, in order to record qualitative information. The main functionalities of the physiotherapist's interface will be: (i) accessing to the patient's performance for each exercise, (ii) consulting the answers to the self-reported functions, (iii) watching the three-dimensional reproduction of the patient's movement through an animated avatar and (iv) updating (to more challenging parameters of the same exercise or to new kinds of exercise) the rehabilitation program according to the patient's progress and/or medical advices. Finally, an interface for the physicians will enable them to supervise all the recovery process and communicate with the physiotherapist to authorize or not certain

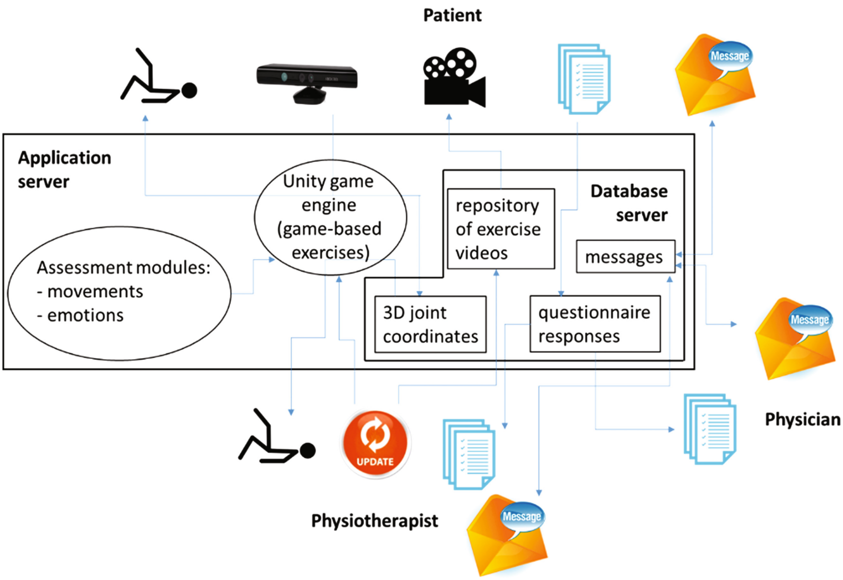


Fig. 1. ePHoRt platform architecture.

movements according to the specific condition of each patient. The communication between the three stakeholders will be supported by an exchange of messages.

The application layer will contain the logical structure of the platform. This intermediate layer will be connected to the client, through Internet, and to the database server. It will receive and process the requests from the three types of user. Here, the Unity 3D game engine will be used to develop the game-based exercises and animate the avatar. Also, the affective and movement assessment modules will be integrated at this level.

The database layer will be connected to the application layer. Heterogeneous data will be stored into the database, such as: (i) quantitative data about the 3D-coordinates of the movements, (ii) qualitative data about the responses to the questionnaires, (iii) videos of the exercises and (iv) comments made by the physiotherapists and physicians.

2 Automatic Assessment of the Correctness of the Movement

2.1 Material and Method

Experimental Protocol. Eight subjects participated in the experiment. They were asked to execute four different rehabilitation movements. Each movement was repeated eleven times: six times correctly and five times imperfectly. The correctness of the exercises was labeled by a physiotherapist. The first movement was a hip abduction (HA). The second movement was a hip extension (HE). The third movement was a slow flexion of hip and knee (SFHK). And the last movement was a sequence, in which the subject had to do one step forward, one step sideways and one step backward (FSB). All of these movements were performed on the right side, only. For these rehabilitation exercises the main mistakes that an individual can do are: (i) an inappropriate amplitude of the movement (too short or too large), (ii) an additional flexion of joints not involved in the exercise (e.g., trunk flexion or extension), (iii) an execution of the movement in the wrong spatial plane, (iv) an incorrect positioning of the center of mass. All of these errors were used as imperfect trials of the experiment. During the execution of the movements, subjects were in stand up position and at approximatively 2.5 m from a Kinect camera. The Kinect height was aligned with the xiphoid apophysis of the subjects. A program was developed to record the 3D-coordinates (X,Y,Z) of each joint of the Kinect Skeleton. Thus, twenty joints were analyzed. The frame rate of the motion capture was 33 Hz, approximately.

Time Series Analysis. A Dynamic Time Warping (DTW) analysis was used to assess the correctness of the movement. This method measures the similarity between two temporal sequences which may vary in speed [28]. Thus, similarities in rehabilitation exercises can be detected, even if the referential movement (the movement correctly executed by the patient) is not executed at exactly the same velocity when performed at home. This technique is based on the calculation of the distance between all pair of points in two signals (distance(i, j)). Lesser distance implies that these points may be candidates to be matched together. In order to find the minimum distance, an accumulated cost matrix (D) is calculated as follows:

$$\mathrm{D(i,j) = min\{D(i-1,j-1),\ D(i-1,j),\ D(i,j-1)\} + distance(i,j)} \qquad (1)$$

For movement assessment, the signal of the movement to be assessed is compared to the signal of the movement correctly executed (correct trials). In the present protocol six trials per exercises were evaluated as correct by the therapist. Thus, the DTW algorithm was applied by the use of six different referential signals. At the end, forty-five distances were calculated for each of the three axes of the whole joints. Thirty of these observations were labeled as incorrect movements and fifteen as correct movements. Then, these data were used as input of a machine learning algorithm of classification, in order to evaluate if the DTW is an accurate technique to discriminate the correct from the incorrect execution of the movements.

Machine Learning Algorithm for Classification. The assessment of the DTW approach is carried out through a machine learning classification. If a high percentage of correct classification is obtained, it will mean that the DTW is an appropriate technique to differentiate the right from the wrong movements. As described in the introduction, several classifiers are possible. Here, the Naïve Bayes classifier is chosen because we make the assumption that all of the features (coordinates X, Y, Z of each joint) contribute equally and independently to the decision [29]. This method is based on the “Bayes Theorem” of the probability (Pr) of an event H (class of an instance) given an evidence E (attribute values of the instance), such as:

$$\Pr(H|E) = \Pr(E|H)\Pr(H)/\Pr(E) \tag{2}$$

Pr(H) is called the priori, or baseline, probability of the hypothesis H. That is the probability of the event before any evidence is seen. Pr(H|E) is a posteriori probability of H, after an evidence is seen (E). The naïve assumption is that the evidence splits into parts that are statistically independent. The parts of the evidence in the movement assessment are the sixty coordinates (20×3) in the joints motion dataset. When the events are independent the probabilities multiply, such as:

$$\Pr(H|E) = \Pr(E_1|H)\Pr(E_2|H)\ldots\Pr(E_{60}|H)\Pr(H)/\Pr(E) \tag{3}$$

There are two hypotheses (H) for the movement assessment: correct vs. incorrect. To get a probabilistic value (between 0 and 1), the likelihood of each hypothesis (or class) is normalized. The main possible issue in using Naïve Bayes method is in case of redundant attributes. In this situation it is possible to use additional methods for feature selection, in order to select a subset of fairly independent attributes.

Feature Selection. Not all of the sixty attributes are essential to proceed with the assessment of the movement. Thus, a selection of the relevant attributes for each kind of exercises could be performed, in order to improve the classification (correct vs. incorrect) of the movement (elimination of redundant features) and to get a simplified model of assessment of the correctness of the movement (only based on the most pertinent features). Several techniques are available to automatically perform this selection. The method used in this study is the "wrapper" attribute selection [30]. This method can be applied backward, forward or bi-directional. The backward searching consists of removing one attribute (the worst one) at each search step. On the contrary, the forward searching start with a zero attribute subset and add the best attribute each time. The bi-directional is a combination of backward and forward searching. In all of the cases, the search stops when the classification performance gets worse. The "wrapper" method uses cross-validation to select the best attribute to add or to drop at each stage. To sum up, two components have to be defined to apply this technique in practice: a search method and an attribute evaluator. The search method defines the searching direction and the search termination criteria. The attribute evaluator evaluates feature sets by using a learning scheme and classifier. In this study, the setup used is backward searching and Naïve Bayes classifier.

2.2 Results

Experimental Results. Overall, the average classification accuracy of the movements is 98.2% (SD = 1.1). The mean of accurate assessments for each movement is higher than 97%. Table 1 shows that the accuracy of this classification is almost the same between the different exercises and the different subjects. These results suggest that the DTW is an appropriate technique to discriminate between correct and incorrect execution of the rehabilitation movements.

Table 1. Percentage of accuracy in the assessment of the movements.

Subjects	HA	FSB	HE	SFHK	Mean
1	.96	1	1	1	.99
2	1	1	.98	.96	.985
3	.91	1	1	.93	.96
4	.96	1	.98	.98	.98
5	1	1	.96	1	.99
6	1	.96	1	1	.99
7	.98	.93	1	.96	.968
8	1	1	.96	1	.99
Mean	.976	.986	.985	.979	.982

After feature selection the overall percentage of accurate classification of the movements is 98%. As previously, none of the movements have a classification lower than 97%. The similarity between the results with and without attribute selection is confirmed by a T-test analysis that shows no significant differences between the classification performance on these two datasets ($p = .7$). It is to note that the value of the standard deviation is slightly lower when a selection of features is performed ($SD = .71$). This fact suggests that the inter-individual differences in the assessment of the movements are reduced when they are based on a selection of the most relevant attributes for each exercise and individual. Figure 2 shows a comparison of the classification accuracy, for each movement and subject, between the two datasets. Even on this detailed analysis it is possible to confirm that the quality of the assessment with the selected features is as good as with the totality of the features. This result suggest that it is preferable to build a model based on a selection of the most relevant attributes than to use the whole features, because we will get a simple model as accurate as a complex one. This characteristic will be fundamental when the model will have to assess the correctness of the movements in real time.

Models. Different models were created for each subject and exercise through a feature selection based on a wrapper technique. Table 2 shows a synthesis of the main joints involved in the assessment of the rehabilitation movements. The main features that enables the algorithm to discriminate between a correct and incorrect movement are related with the “Right Foot”, the “Right Ankle” and the “Right Knee”. This is not surprising considering that all of the exercises asked to the participants were designed for the rehabilitation of the “Right Hip”. Taking together, these three joints represent 78% of the features used for the assessment. The other 22% are represented by different kind of joints according to the exercise and the individual. With 62.5% of “Other Joints”, “Hip Abduction” is the movement with the largest inter-individual variability. On the contrary the “Forward, Sideways and Backward” sequence is the only exercise that can be exclusively assessed on the base of the lower limbs (100%) involved in the movement (mostly foot kinematics). Also, “Hip Extension” and “Slow Flexion of Hip and Knee” are mainly evaluable through an analysis of the lower limbs in movement with 80% and 90% of the recognition based on these joints, respectively.

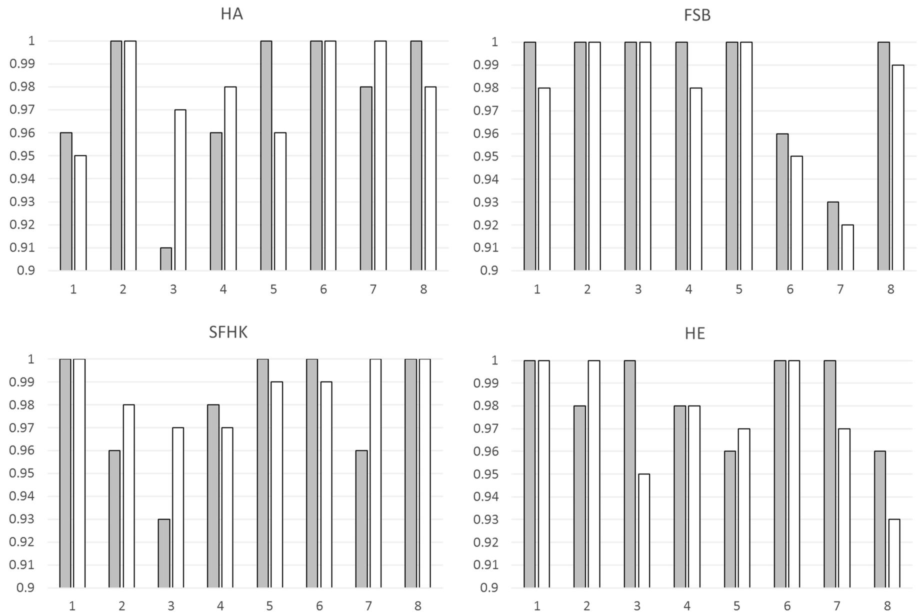


Fig. 2. Comparison of classification accuracy (in %) between the whole feature dataset (grey bars) vs. the selected feature dataset (white bars) for each exercise and individual.

Table 2. Percentage of the joints used to assess each exercise.

Exercises	R Foot	R Ankle	R Knee	Other joints
HA	.375	0	0	.625
FSB	.625	.1	.25	0
HE	.4	.2	.2	.2
SFHK	.5	.3	.1	.1
Mean	.48	.15	.15	.22

To sum up, half of the assessment of the exercises depends on the kinematic of the "Right Foot". Then, the "Right Ankle", "Right Knee" and the rest of the joints contribute almost equally for the evaluation of the correctness of the movement. However, it is difficult to get a generic model that could be used to assess anybody, because more than 20% of the features represent joints that are not directly involved in the movement. Thus, the best and easier way to assess the movement is to create a specific model for each individual and exercise. For instance, Table 3 shows the models obtained for the subject nº5.

In the case of this individual the algorithm only needed two features to discriminate between a correct and incorrect "Hip Abduction" ("Head" in the y-axis and "Right Foot" in the x-axis). For the "FSB" movement a single attribute was sufficient for the discrimination ("Right Foot" in the z-axis). Again, only two attributes were necessary

Table 3. Models for the subject no. 5 to assess the whole therapeutic exercises.

Exercises	Attributes	Classes			
		Incorrect		Correct	
		Mean	SD	Mean	SD
HA	Head - Y	2.67	1.92	.06	.07
	R Foot - X	2.08	.92	1.11	.19
FSB	R Foot - Z	267805	93179	63722	15659
HE	R Elbow - Z	37712	24606	5758	1857
	R Foot - Y	2.57	1.33	1.39	.29
SFHK	R Foot - Z	151958	69724	29849	4175

for the assessment of the "HE" exercise ("Right Elbow" in the z-axis and "Right Foot" in the y-axis). And the "SFHK" was assessed by a single attribute, as well ("Right Foot" in the z-axis). The value of the mean of each feature corresponds to the distance between the referential movement and the tested movement calculated by the DTW technique. Data in Table 3 confirm that the mean and the standard deviation are always lower for the correct movements than the incorrect movements.

Through these models it is possible to estimate the correctness of a new movement executed by this subject, based on the calculation of the two probabilities of this movement for being correct and incorrect. The example of the "Hip Abduction" is used to explain the methodology applied to assess the movement from its model. The method is based on the use of the distribution of the numerical variable (attribute values) to have a good guess of the frequency. One common practice is to assume normal distributions for the numerical variables. The probability density function for the normal distribution is defined by two parameters: mean (μ) and SD (σ). These two values are provided by the model in Table 3 for each selected feature of a determined movement. This probability is obtained through the calculation of the value of the normal distribution of an attribute value (x), by the use of the Eq. 4.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

Thus, in the case of the "Hip Abduction" for subject no. 5, the probabilities to assess a new movement as correct and incorrect by the use of the "Head distance in the y-axis" (d_{headY}) will be obtained by the calculation of the Eqs. 5 and 6, respectively.

$$\Pr(d_{headY}|correct) = \frac{1}{0.07\sqrt{2\pi}} e^{-\frac{(d_{headY}\ -\ 0.06)^2}{2(0.07)^2}} \tag{5}$$

$$\Pr(d_{headY}|incorrect) = \frac{1}{1.92\sqrt{2\pi}} e^{-\frac{(d_{headY}\ -\ 2.67)^2}{2(1.92)^2}} \tag{6}$$

The same calculation has to be applied to get the probabilities with the feature "Right Foot distance in the x-axis" (d_{rFootX}), as described by the Eqs. 7 and 8.

$$\Pr(d_{rFootX}|correct) = \frac{1}{0.19\sqrt{2\pi}} e^{-\frac{(d_{rFootX} - 1.11)^2}{2(0.19)^2}} \tag{7}$$

$$\Pr(d_{rFootX}|incorrect) = \frac{1}{0.92\sqrt{2\pi}} e^{-\frac{(d_{rFootX} - 2.08)^2}{2(0.92)^2}} \tag{8}$$

Then, the value of each probability is used to calculate the likelihood (LL) for a movement classification as correct and incorrect. The calculation of the two likelihoods is described in the Eqs. 9 and 10.

$$\mathrm{LL(correct)} = \mathrm{Pr(d_{headY}|correct)Pr(d_{rFootX}|correct)Pr(correct)} \tag{9}$$

$$\mathrm{LL(incorrect)} = \mathrm{Pr(d_{headY}|incorrect)Pr(d_{rFootX}|incorrect)Pr(incorrect)} \tag{10}$$

Finally, these likelihoods have to be normalized to get the value of the probabilities for the new movement to be correct and incorrect (see Eqs. 11 and 12). 50% is usually used as a threshold to take the final decision to classify the trial in one or the other class.

$$\mathrm{Pr(correct)} = \mathrm{LL(correct)/(LL(correct) + LL(incorrect))} \tag{11}$$

$$\mathrm{Pr(incorrect)} = \mathrm{LL(incorrect)/(LL(correct) + LL(incorrect))} \tag{12}$$

3 Conclusions and Perspectives

The presented study is part of the ePHoRt project, which is a telerehabilitation platform for patients after hip replacement surgery. It focuses on the automatic assessment of the correctness of the exercises performed by the user. The results show that a DTW approach permits a high level of discrimination between correct and incorrect execution of the movements (98.2%). The implementation of a feature selection technique allows us to build simple models only based on few attributes to evaluate the movement. Future work will consist of (i) a DTW analysis on the joint angles and (ii) testing the algorithms in real time with patients.

References

1. Feys, H., De Weerdt, W., Verbeke, G., et al.: Early and repetitive stimulation of the arm can substantially improve the long-term outcome after stroke: a 5-year follow-up study of a randomized trial. Stroke **35**(4), 924–929 (2004)

2. Cramp, M.C., Greenwood, R.J., Gill, M., et al.: Effectiveness of a community-based low intensity exercise programme for ambulatory stroke survivors. Disabil. Rehabil. **32**(3), 239–247 (2010)
3. Mavroidis, C., Nikitczuk, J., Weinberg, B., et al.: Smart portable rehabilitation devices. J. Neuroeng. Rehabil. **2**, 18 (2005). doi:10.1186/1743-0003-2-18
4. Holden, M.K., Dyar, T.A., Dayan-Cimadoro, L.: Telerehabilitation using a virtual environment improves upper extremity function in patients with stroke. IEEE Trans. Neural Syst. Rehabil. Eng. **15**(1), 36–42 (2007)
5. Rand, D., Kizony, R., Weiss, P.T.L.: The Sony PlayStation II EyeToy: low-cost virtual reality for use in rehabilitation. J. Neurol. Phys. Ther. **32**(4), 155–163 (2008)
6. Oikonomidis, I., Kyriazis, N., Argyros, A.A.: Efficient model-based 3D tracking of hand articulations using Kinect. In: Proceedings of the 22nd British Machine Vision Conference. University of Dundee (2011)
7. Gameiro, J., Cardoso, T., Rybarczyk, Y.: Kinect-sign: teaching sign language to listeners through a game. In: Rybarczyk, Y., et al. (eds.) Innovative and Creative Developments in Multimodal Interaction Systems, pp. 141–159. Springer, Heidelberg (2014)
8. Rybarczyk, Y., Santos, J.: Motion integration in direction perception of biological motion. In: Proceedings of the 4th Asian Conference on Vision, Matsue (2006)
9. Dutta, T.: Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace. Appl. Ergon. **43**, 645–649 (2012)
10. Brook, G., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., Rochester, L.: Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. Gait Posture **39**(4), 1062–1068 (2014)
11. Rybarczyk, Y.: 3D markerless motion capture: a low cost approach. In: Proceedings of the 4th World Conference on Information Systems and Technologies, Recife (2016)
12. Remondino, F., Roditakis, A.: 3D reconstruction of human skeleton from single images or monocular video sequences. In: Proceedings of Joint Pattern Recognition Symposium, Magdeburg (2003)
13. Krukowski, A., Vogiatzaki, E., Rodríguez, J.M.: Patient Health Record (PHR) system. In: Maharatna, K., et al. (eds.) Next Generation Remote Healthcare: A Practical System Design Perspective, Chap. 6. Springer, New York (2013)
14. Rybarczyk, Y., Deters, J.K., Aladro Gonzalvo, A., Gonzalez, M., Villarreal, S., Esparza, D.: ePHoRt project: a web-based platform for home motor rehabilitation. In: Proceedings of the 5th World Conference on Information Systems and Technologies, Madeira (2017)
15. Bulling, A., Blanke, U., Schiele, B.: A tutorial on human activity recognition using body-worn inertial sensors. ACM Comput. Surv. **46**(3), 33 (2014). doi:10.1145/249962
16. Chen, L., Hoey, J., Nugent, C.D., et al.: Sensor-based activity recognition. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(6), 790–808 (2012)
17. Choudhury, T., Consolvo, S., Harrison, B., et al.: The mobile sensing platform: an embedded activity recognition system. IEEE Pervasive Comput. **7**(2), 32–41 (2008)
18. Galveia, B., Cardoso, T., Santos, V., Rybarczyk, Y.: Towards the creation of a gesture library. EAI Endorsed Trans. Creat. Technol. **15**(3), e3 (2015)
19. Jost, H.: Kinect-based approach to upper limb rehabilitation. In: Vogiatzaki, E., Krukowski, A. (eds.) Modern Stroke Rehabilitation Through e-Health-Based Entertainment, pp. 169–193. Springer, Heidelberg (2016)
20. Antón, D., Goñi, A., Illarramendi, A.: Exercise recognition for Kinect-based telerehabilitation. Meth. Inf. Med. **54**(2), 145–155 (2015)
21. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Mach. Intell. **22**(1), 4–37 (2000)

22. Kunze, K., Barry, M., Heinz, E.A., et al.: Towards recognizing tai chi: an initial experiment using wearable sensors. In: Proceedings of the 3rd International Forum on Applied Wearable Computing. Bremen (2006)
23. Banos, O., Damas, M., Pomares, H., et al.: Daily living activity recognition based on statistical feature quality group selection. Expert Syst. Appl. **39**(9), 8013–8021 (2012)
24. Maurer, U., Smailagic, A., Siewiorek, D.P., Deisher, M.: Activity recognition and monitoring using multiple sensors on different body positions. In: Proceedings of International Workshop on Wearable and Implantable Body Sensor Networks, Cambridge (2006)
25. Parkka, J., Ermes, M., Korpipaa, P., et al.: Activity classification using realistic data from wearable sensors. IEEE Trans. Inf. Technol. Biomed. **10**(1), 119–128 (2006)
26. Besson, P., Richiardi, J., Bourdin, C., Bringoux, L., Mestre, D.R., Vercher, J.L.: Bayesian networks and information theory for audio-visual perception modeling. Biol. Cybern. **103**(3), 213–226 (2010)
27. Rybarczyk, Y., Zalakeviciute, R.: Machine learning approach to forecasting urban pollution: a case study of Quito, Ecuador. In: Proceedings of IEEE Ecuador Technical Chapters Meeting, Guayaquil (2016)
28. Salvador, S., Chan, P.: Toward accurate dynamic time warping in linear time and space. Intell. Data Anal. **11**(5), 561–580 (2007)
29. John, G.H., Langley, P.: Estimating continuous distributions in Bayesian classifiers. In: 11th Conference on Uncertainty in Artificial Intelligence, San Mateo (1995)
30. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artif. Intell. **97**, 273–324 (1997)

HFE Application in Human System Interface Design of Nuclear Power Plant

Xiaoyang Ming(✉), Guangwei Yu, and Shiguang Deng

China Nuclear Nuclear Engineering Corporation, Heneng Building, #117 Xisanhuan Road, Haidian District, Beijing, China
mingxiaoyang1990@126.com, yugw@cnpe.cc, dengshiguang@qq.com

Abstract. Human System Interface is the medium that the operators monitor and control the Nuclear Power Plant. To improve the operating performance for plant safety, Human Factor Engineering needs to be adopted in the new plant HSI design. An implementation methodology for the application of HFE in the HSI designing is proposed here based on NUREG-0711 which is established as the review guideline for HFE in the new plant construction. The methodology mainly contains three parts as: Functional Requirements Analysis and Function Allocation, Task Analysis and Human-System Interface design. With systematic analysis of HFE, the HSI would be more scientific and rational.

Keywords: Human factors · Human system interface · Nuclear power plant

1 Introduction

Nuclear Power Plant (NPP) Human-System Interface (HSI), through which operators interact with the plant, includes the alarms, displays and so on. To improve the operators' manipulation efficiency, the engineers spent more and more time on the design of HSI.

Operators monitor and control plant systems and components to verify their proper functioning. The Human Factors Engineering (HFE) focused on the improvement of the operating performance on both normal conditions and emergency conditions of the plant to keep the plant safety operation. The HFE aspects of the plant should be developed, designed, and evaluated on the base of a structured analysis using accepted HFE principles. The American Nuclear Regulatory Commission (U.S. NRC) made the NUREG-0711 [1] as the design and review guideline for HFE. Here, we introduce one HFE implementation plan which can be used in the HSI design of NPP.

2 Human Factor Engineering Program

In NUREG-0711, the HFE program is sorted into twelve elements arranged in four general activities as shown in Fig. 1.

For the design of HSI, system and operations analysis is the basic tool used to establish design requirements, which is accomplished by systematically defining equipment,

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_34

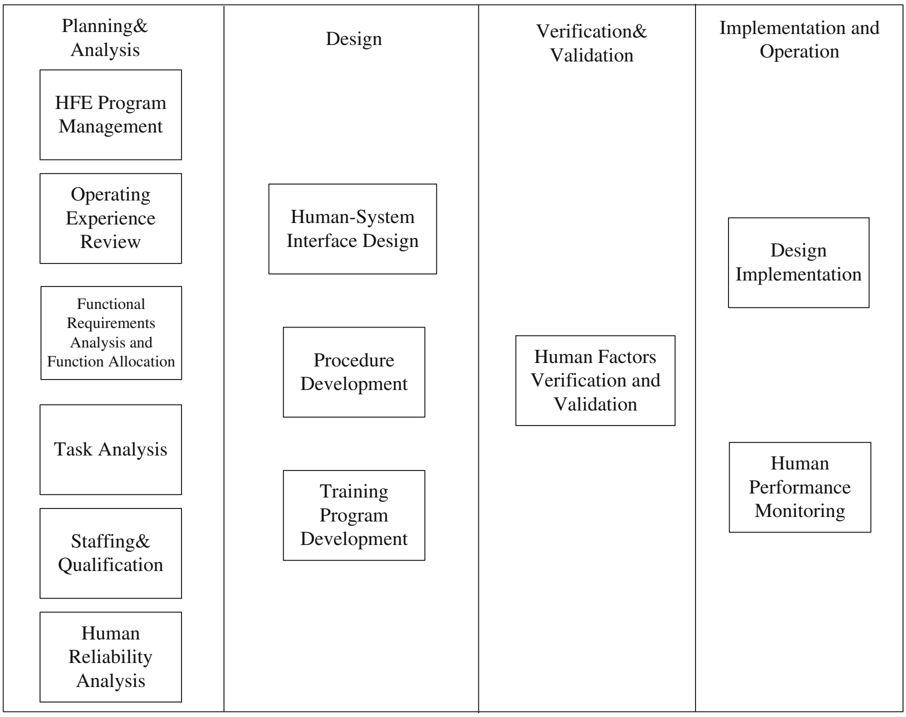


Fig. 1. HFE program review model

software, personnel and procedural data requirements to meet all functional objectives of each system and its operating crew, including safety operation of the plant. Above all, Functional Requirements Analysis and Function Allocation (FRA&FA), Task Analysis (TA) and HSI are the three basic steps to design the HSIs with HFE.

FRA&FA: The object is to verify that the applicant has (1) defined the plant's functions that must be performed by operator to satisfy plant safety objectives, and (2) that the allocation of those functions to human and system resources has resulted in a role for personnel that takes advantage of human strengths and avoids human limitations.

TA: The object of TA is to verify that the applicant's task analysis identifies the specific tasks that are needed for function accomplishment and their information, control and task-support requirements. The task analysis results should provide input to the design of HSIs, procedures, and personnel training programs.

HSI: The HSI should be designed using a structured methodology that should guide designers in identifying and selecting candidate HSI approaches, defining the detailed design, and performing HSI tests and evaluations. The HSI should be satisfied with the requirements of functional and task analysis about alarms, displays, controls and other aspects of the HSI.

In the following chapter, a methodology for the HSI design is proposed with three steps as FRA&FA, TA, and HSI.

3 Implementation Plan for HSI Design with HFE Principle

3.1 Function Requirement Analysis and Function Allocation

To analyze the applicability of a plant system with the requirement of HFE guideline, the first step is to identify the functions that should be performed to satisfy plant safety objectives. On the other hand, verifying that the allocation of those functions to human and system resources has resulted in a role for personnel that takes advantage of human strengths and avoids human limitations [2].

Figure 2 illustrates the exact process for dealing with functions analysis (safety or not).

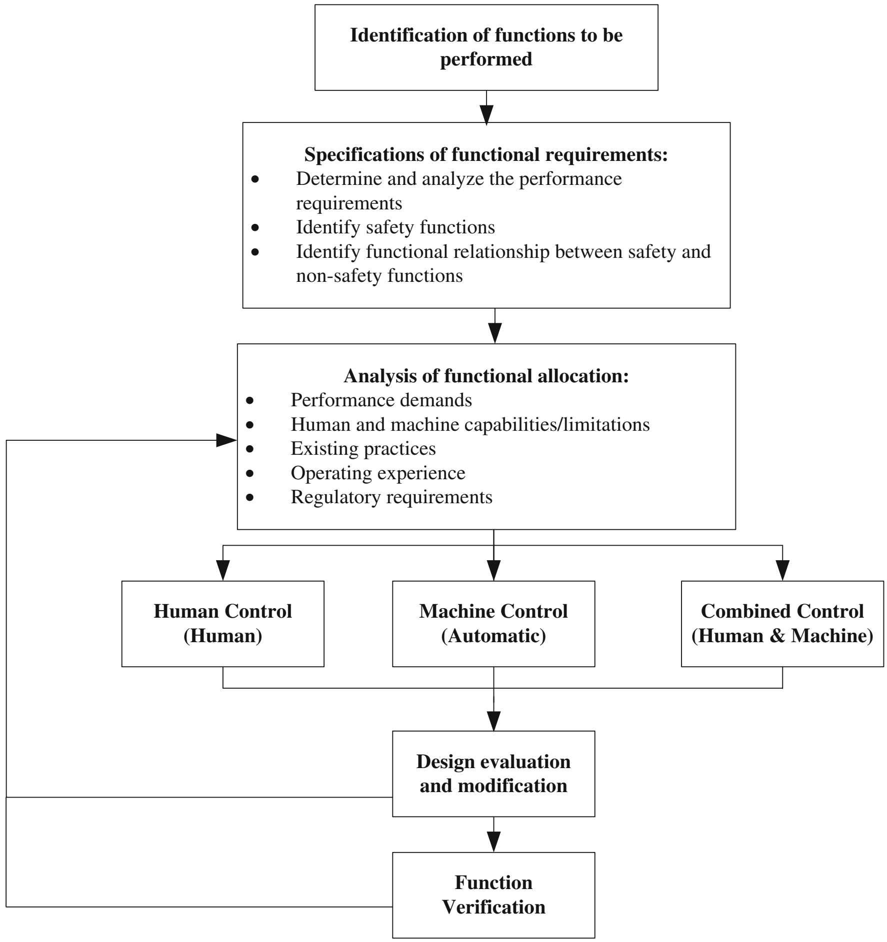


Fig. 2. Process for dealing with functions analysis

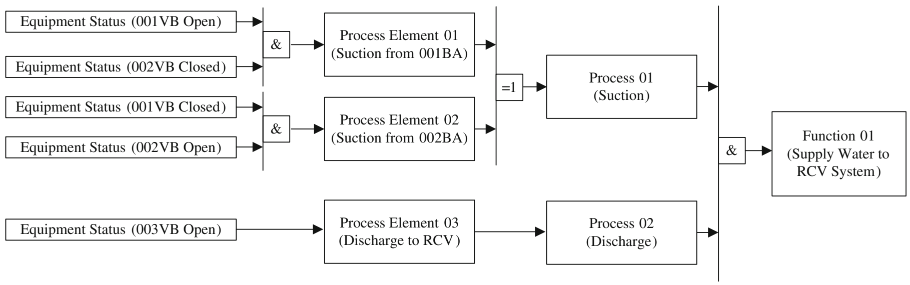


Fig. 3. Four level hierarchical structure of function

It is required to identify the safety functions defined in system design input documents and verify they support the plant critical safety functions (CSF). The CSFs are those functions that are essential to prevent a direct and immediate threat to the health and safety of the public. These are the accomplishing or maintaining of: reactivity control, reactor core cooling, reactor coolant system integrity, primary reactor containment integrity and radioactive effluent control. It will be stated if the function is safety-related or not. For the requirement of NUREG-0711, designer should provide the purpose of the function, conditions that indicate that function is needed, parameters that indicate the function is available, parameters that indicate the function is operating, parameters that indicate the function is achieving its purpose, parameters that indicate the function can be terminated.

In order to accomplish these objectives, each system can be analyzed according to a four level hierarchical structure to identify: Function, Process, Process element, Component status, which can be illustrated in Fig. 3.

Function: Function is the definition of what the system does. It can be summarized from the system design input documents of each system. Different from the function in input documents, the summarized function is based on the components and parameters that can be controlled or be checked in the main control room.

Process: The systems can be decomposed into smaller parts called processes. The processes which contain system components should have a significant effect on the function either individually or in combination.

Process Element: The valid status alignment of all the components in the process composes Process Element.

Component Status: Component has different status when it is working at different conditions. For example, the status of an isolation valve can be open or closed; while the status of a pump can be running or stopped.

Figure 3 has shown an example of the hierarchical structure for the function "Supply Water to Chemical and Volume Control System (RCV)". It's pretty clear that this function is to suck water from 001BA or 002BA, and then discharge to RCV system. Two processes named as "Suction" and "Discharge" are included in this function. Two valves (001VB and 002VB) are included in the process "Suction"; and one valve (003VB) belongs to the other process "Discharge". Although there are two suction paths, they are not considered as different functions but different operating modes (OM) of the same function (Supply Water to RCV System).

OMs are those different components alignments which can be used to perform a function. OMs will be determined as the different combinations of process elements that perform the function. A certain OM can only be composed of one process element from each process. There are two OMs for the function “Supply Water to RCV System”, named as “Supply With Water from 001BA” and “Supply With Water from 002BA”. When the suction source is changed from 001BA to 002BA, or vice versa, it is considered as a mode change.

When the function is defined, the next step is to identify the FA. The FA will be the evaluation of every function from FRA that can be allocated to a human, machine or combination of human and machine. The FA should be based on the system design and the performance demands, the analysis process is shown in Fig. 4.

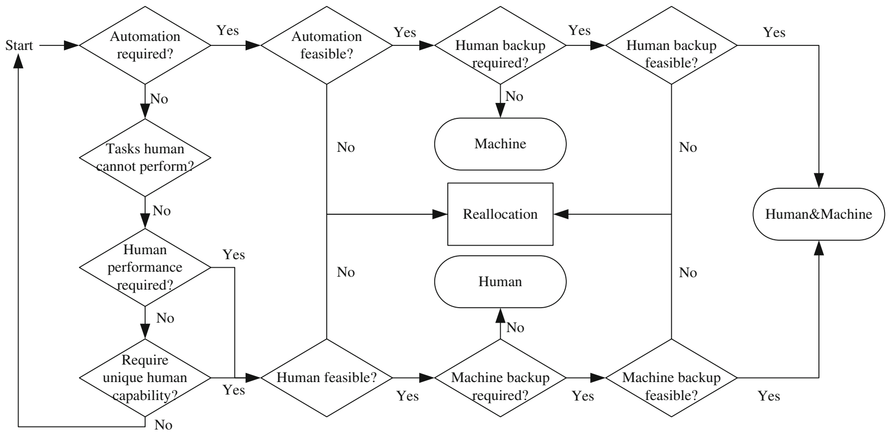


Fig. 4. Evaluation of function allocation

Hypothetical allocation of functions is made based on the “Overall Control Actions” associated to all the operating modes and mode changes rather than the whole functions. Overall control actions are the initiation, performance, verification and termination of each operating mode or mode change. These control actions will be allocated either to the machine, human, or a combination of both, according to the evaluation process in Fig. 4. In Table 1, it is an example for the function allocation of mode change from “Supply With Water from 001BA” to “Supply With Water from 002BA”. For the valves are operated manually in the MCR, so the initiation of this mode change is Human while termination is not available for a mode change.

Once function analysis and allocation are defined, the operator task is to be analyzed for the function fulfillment.

Table 1. Function allocation example.

Operating mode/mode change	Overall control	Machine	Human	Combination
Mode change from "Supply with water from 001BA" to "supply with water from 002BA"	Initiation	N/A	Yes	N/A
	Performance	N/A	Yes	N/A
	Verification	N/A	N/A	Yes
	Termination	N/A	N/A	N/A

3.2 Task Analysis

To accomplish the functions allocated to plant personnel that define their roles and responsibilities, human actions are performed. Human actions can be further divided into tasks which are groups of related activities that have common objective or goal. Task analysis is the identification of requirements for accomplishing these tasks, i.e., for specifying the requirements for the HSI, data processing, controls, and job support aids needed to accomplish tasks. Use of explicit task analysis therefore leads to more efficient and effective integration of the human element into the system design and operations in two principal areas: Safety and Availability.

TA can be divided into three steps.

Operating Sequence Scenarios (OSS): In order to develop the descriptions of the operating sequences, it is necessary to define scenarios that include all the operations that can be performed with the system. OSS is defined for the initiation and termination of every operating mode or mode change (normal, abnormal or emergency operation). For one OSS, the included elements are: Initial Conditions (The condition of the plant and the system before initiating the scenario), Sequence Initiator (The reason or situation that requires initiating the scenario), Final Conditions (Situation of the plant or the system after terminating the scenario), Operation Classification and Standards (Type of plant operation).

Tasks: sets of activities that the operator has to perform in order to accomplish the system functions for each defined scenario.

Activities: actions that the operator has to perform sequentially in order to accomplish each one of the defined tasks.

For the mode change example in Table 1, one OSS can be created with "Manual Transfer from Supply With Water from 001BA to Supply With Water from 002BA". For the OSS, we have to open 002VB and close 001VB, 003VB should be kept open. So the corresponding task could be named as "Place suction from 002BA in service".

So the OSS and task for the example mode change can be shown in Table 2:

For ST01, we have two activities to perform:

A01: Open isolation valve 002VB;

A02: Close isolation valve 001VB (Table 3).

In order to provide a general view of each task, a diagram that is called Operational Sequence Diagram (OSD) will be developed for each task, in which the operations to be carried out will be presented. Tasks can be explained sequentially by OSD, by using

Table 2. Operating sequence scenario example

OSS title	Manual transfer from supply with water from 001BA to supply with water from 002BA
Initial conditions	Supply with water from 001BA
Sequence initiator	001BA short of water
Final conditions	Supply with water from 002BA
Classification and standards	Normal operation
OSS task	Manual transfer from supply with water from 001BA to supply with water from 002BA
T01 (ST01)	Place suction from 002BA in service

Table 3. Task and activity example

Task	ST01: Place suction from 002BA in service	Component
ST01 A01	Open isolation valve 002VB;	002VB
ST01 A02	Close isolation valve 001VB	001VB

symbols to indicate actions, inspections, data transmitted or received, data stored, and decisions made, the OSD shows the flow of information through a system. The information flow is shown in relation to both time sequence and location. In the OSD, the interrelationships between operators and equipment (man-machine interfaces) are easily displayed. The OSD symbols are shown in Fig. 5, which is consistent with the American Society of Mechanical Engineers (ASME) flow chart standards, the corresponding OSD for the example task ST01 is illustrated in the Fig. 6.

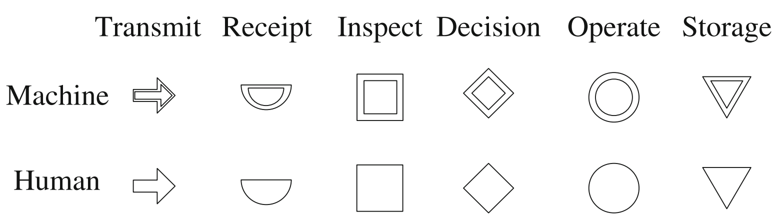


Fig. 5. Symbols for OSD

When the activities are analyzing, we have to gather all the control and display requirements from the MCR as the design input for HSI design.

3.3 Human-System Interface

The HSI design process represents the translation of function and task requirements into HSI characteristics and functions. The HSI should be designed with a structured

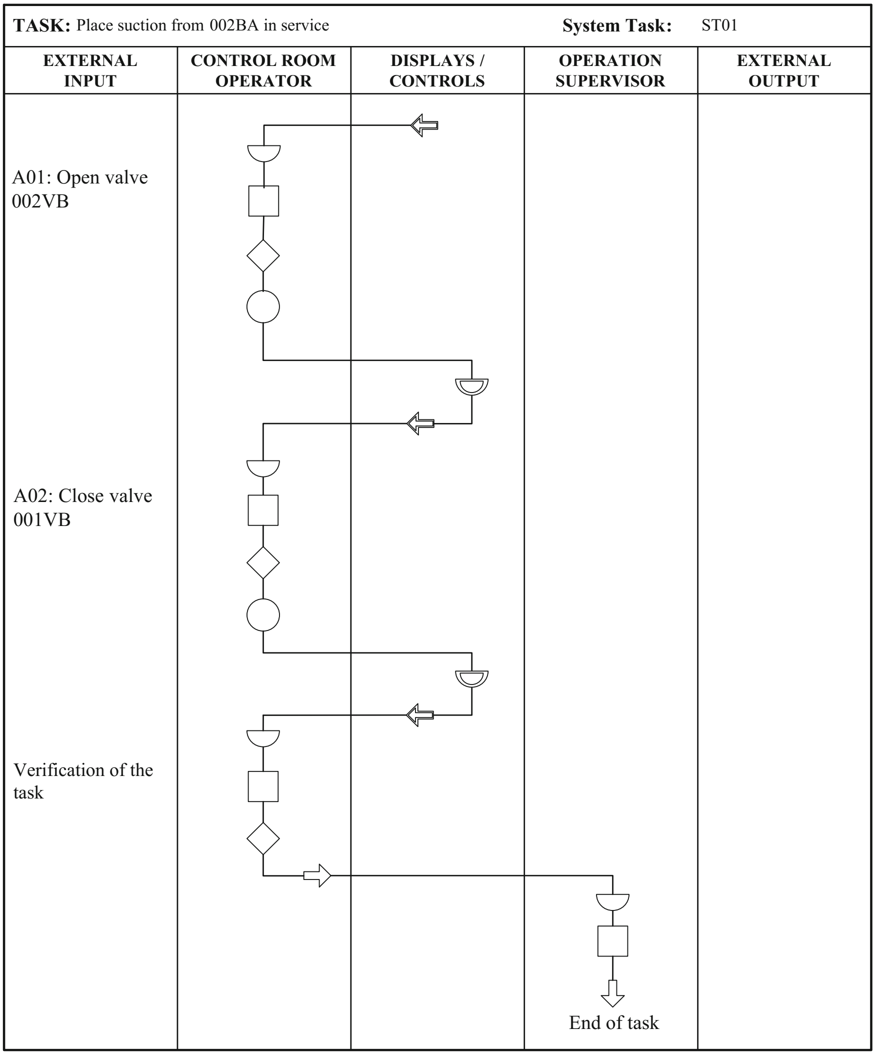


Fig. 6. OSD for task ST01: place suction from 002BA in service

methodology that should guide designers in identifying and selecting candidate HSI approaches, defining the detailed design, and performing HSI tests and evaluations [3].

The HSI detailed design should support personnel in their primary role of monitoring and controlling the plant while minimizing personnel demands associated with use of the HSIs. NUREG-0700 describes high-level HSI design review principles that the detailed design should reflect.

After all, one small example for the HSI of “Supply Water to RCV System” is shown in Fig. 7.

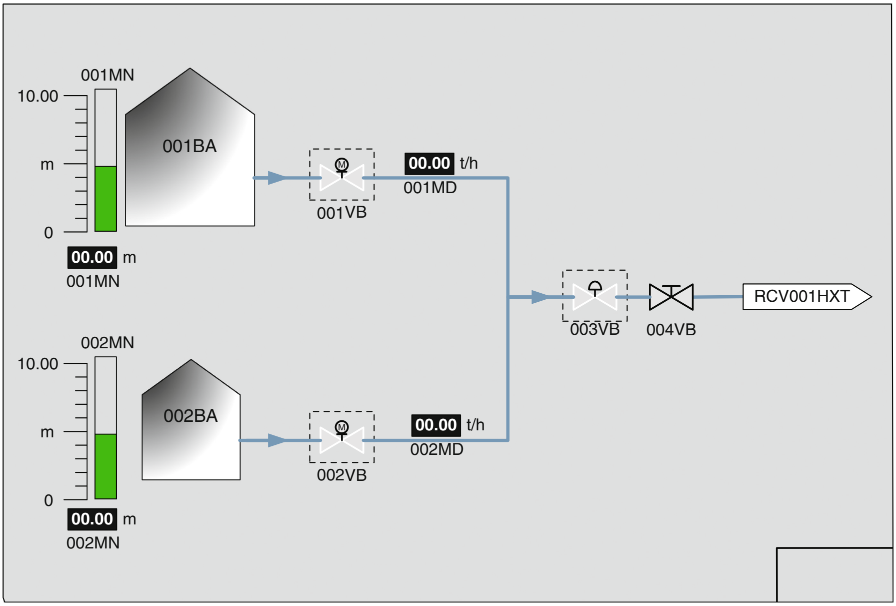


Fig. 7. HSI example

In Fig. 7, as the basic methodology of NUREG-0700, the design has general rules as follows:

General flow direction: from left to right, from up to down;
Width for pipes: 3 pixels for main pipe, 2 pixels for bypass pipe, and 1 pixel for auxiliary pipe;
Color for pipe: the pipe color shows the differentiate fluids that flow along the pipe, here blue represents water;
Distinction between control and indication representation: components with black dashed line means control capabilities from HSI while the white represents indication abilities in HSI for this component, which are defined in Task Analysis.

Although the above rules are parts of the design rules, they are universal for all the HSI even though the layout is much more complicated.

4 Conclusions

To meet the requirements of HFE for the HSI, an implementation methodology is proposed for the HSI design based on NUREG-0711. The detail methodology which is composed of three steps as FRA&FA, TA and HSI can put the HFE analysis into practice for the HSI engineering design. Now it is used in the new NPP HSI design of China which is well received by customers.

References

1. USNRC: Human Factors Engineering Program Review Model. S. NUREG-0711 (2012)
2. Wang, X., Zhu, P.: The analysis of human-system interface's human factor engineering design in nuclear power plant's main control room. J. EIC **20**, 83–85 (2013)
3. IEEE: Recommended Practice for the Application of Human Factors Engineering to Systems, Equipment, and Facilities of Nuclear Power Generating Stations and Other Nuclear Facilities. S. IEEE Std. 1023-2004

Creative Activity Mediated by Technology – Artifacts, Technology and Ideas that Could Shape Our Lives

Arminda Guerra Lopes(✉)

Polytechnic Institute of Castelo Branco and Madeira Interactive Technologies Institute, Madeira, Portugal
aguerralopes@gmail.com

Abstract. This paper attempts to associate creativity and design thinking with the mediated technology. It represents the results from creative thinking experiences of human computer interaction (HCI) students within the design of interactive artifacts. The objective is to demonstrate how to engage students in learning by doing and to link theoretical concepts with design practices during the process of producing artifacts. Our findings suggested that creativity is not taught and students when they are not provided with the work statement, they get the opportunity to be more creative and engaged during the learning process by discovering the idea and planning its execution.

Keywords: Human computer-interaction · Creative activity · Mediated technologies · Design thinking · Learning by doing

1 Introduction

In retrospect, great discoveries may seem like single, definable unique moments, which tended into view slowly. Over the past 600 years, the way that great inventions and discoveries are made seems to have dropped gradually away from individual inventors and toward networks of people. Lienhard in ‘How Invention Begins’, merges the ends of invention with the individual dives upon which they are built, enlightening the enormous web of individual inspirations that lie behind whole technologies [1]. Conversely, the Japanese art of inventing ingenious everyday gadgets that, on the surface, seem like an ideal solution to a particular problem, are not exactly useful, but somehow not altogether useless. These created inventions “didn’t quite work…but were nonetheless fun”.

There have been three main concepts by which creativity has been defined in literature: the creative process [2, 3] the creative person [4, 5] and the creative product [6]. Previous research has tended to focus on the embodiment of creative ideas in subsequent products, viewing the latter as the creative product. Creativity is the production of the ideas themselves – the product of design, rather than the product of implementation [7].

Technology does not quell creativity, in fact, there’s a great deal of evidence that suggests that technology enhances creativity [8]. People expectations are that we are more creative in our working lives than a generation ago. The increasing possibilities

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_35

and computerizing parts of the creative process allow us to be more creative and productive. The cleverly designed tools at our disposal provide us to generate everything. Conversely, “the best way to create value in the 21st century is to connect creativity with technology” (Steve Jobs).

This paper presents the results of a study carried out on a public high education institution. The learning domain was human computer interaction and our goals were to motivate students to engage deeply on the subjects, and to produce, creatively, artifacts with mediating technologies.

2 Background

The necessity to reflect on Human Computer Interaction (HCI) history has to do with the rapid change on the field focus. By understanding the reasons for different outcomes, we can assess today’s visions more realistically. HCI as a cover to form disciplines: human factors, information systems, computer science, and library, and information science. HCI wherever it is studied it will be in its early days [10]. The reasons for that are the changes on influences: new waves of hardware enable different ways to support the same activity; email changed the way we communicate; social networking came along; the desktop computer has lost the spotlight to portable devices; government and industry are investing on parallel computing; different patterns of technologies use emerge in different cultures, different industries; accessibility and sustainability are development areas; digital technologies changed people’s behaviors.

There are still cultural barriers that separate the Conference on Human Factors in Computing Systems (CHI) and Information Systems (SI): CHI discovered the limitations of laboratory studies and surveys for understanding discretionary use of methods and IS research focus on the economic, organizational, and marketing theory and practice than CHI [11]. The way design is done and who is doing it changed based on a move from the designing of things to interactions to systems, and from designing for people to designing with people and by people [12].

Recent issues about the disciplinary HCI have been debated. Kostakos based on the CHI publications over the past 20 years showed that the field seemed to follow technical fashions rather than long-term research themes. He considers that HCI does not seem to have a solid intellectual or methodological core [13]. Then, Stuart Reeves recommends that HCI should be thought not as a discipline but rather as an interdiscipline [14], which is agreed by Rogers and Blackwell [15]. Blackwell goes deeply on the discussion stating that there is a hole at the center of HCI research. To explain this hole he is based on two case studies he did: one was a systematic analysis of 180 collaborative projects on research in interdisciplinary design to understand the insights into human behavior. The second case study was a survey of interdisciplinary innovation. These studies allowed reflecting on the use of theories and methods on other disciplines to do HCI work. And this was the problem: is HCI a discipline or not? It might be not. Kim advised for interdisciplinary cooperation in HCI. For him HCI field is not a discipline but being an interdisciplinary field it might be called as generality [16].

Pan et al. are concerned about HCI becoming a fashion driven discipline. They propose to examine and explore what might happen if HCI becomes a fashion driven discipline [17]. Several aspects of HCI research such as user experience, aesthetics, design thinking that is more difficult to research in a traditionally scientific way. Fashion in relation to interaction design has a place in HCI and it becomes an influence on the decisions and judgments made by HCI practitioners and researchers [18]. Other authors subscribe this approach that fashion affects interaction designers 'design thinking in relation to functionality, appearance, user experience, and visual experience.

3 Mediated Technology and Creativity

Our society has become a more technological, which leads to a more creative society. As a result, the routine tasks that used to take up a lot of our time and effort have become automated. Technology increases our potential to engage in the types of experiences that lead to greater creativity. Creativity is the essence of technology we have today, without technology we would be without the most basic of equipment in our modern day lives. It would be difficult to imagine a life without cars, computers, and even houses. Technology does not control creativity technology enhances creativity by the way users dealt with it. Digital technologies have an important role on digital information and communications technologies (ICT). They are to be appropriately applied as set of tools in the creative process. Conversely, the characteristics of ICT can also make a distinctive contribution to those processes, providing new tools, media and environments for learning to be creative and learning through being creative.

In our experience, students of HCI are, in general, able to use a variety of skills and techniques with particular technologies. They are also able to understand the reasons why and whose digital technologies might be appropriate, for particular tasks and situations. Students made learned choices with technologies, they evaluated its impact and they were opened to new developments and possibilities.

3.1 Creativity

We investigate existing theories and models of creativity. There is an overabundance of creative theories and models and no one is generally accepted [19, 20]. The main reason is probably due to the multidisciplinary of the subject.

Creativity is becoming an intrinsic part of working life. It is a quality that is highly regarded, but not always well understood [21]. Creativity denotes a person's capacity to produce new or original ideas, insights, inventions, or artistic products, which are accepted by experts as being scientific, aesthetic, social or technical value [22]. Creative activity grows out of the relationship between an individual and the context of his or her work, as well as out of the relations between an individual and the other human being. Much human creativity arose from activities that took place in the context in which interaction and the artifacts that embodied group knowledge were important providers to the process. Moreover, creative thinking is not so much an individual feature but rather a social phenomenon involving interactions among people within their specific group or cultural settings [23]. "Creativity isn't just a property of individuals; it is also a property of social groups" [24]. And the more divergent the thinking of a group, the more creative the group be considered. So, the changes on the environment where students started working will enhance creativity instead of trying to make them think more creatively [25].

Other approaches about creativity reflect on creativity as the tendency to generate or recognize ideas, alternatives, or possibilities that may be useful in solving problems, communicating with others, and entertaining ourselves and others [26]. Creativity can occur through any mode of thought, most often pursued through intentional, relational and formative thought. However, creativity often results from relating cognitive objects in such a way as to generate new interpretations or expressions [27].

As there are many definitions of creativity, there are also many studies in this area in different fields, for example: Kim offers a case study of interdisciplinary student design teams [16]. Mahdi et al. studied creativity through innovation and engagement in science and technology education [28]. Lemos et al. presented an application of creativity approaches to support the requirements engineering [20]. Cherry et al. developed a creativity support Index, which is a survey for evaluating the ability of a creativity support tool to assist a user engaged in creative work [30]. A Creativity support tools (CSTs) can make a substantial impact on both individuals and society by improving scientific, engineering, humanist, and artistic endeavors [7, 31]. A creativity support tool serves as an instrument that can be used by people in the open-ended creation of new artifacts.

We outline four essential stages within our study in creativity: the process of getting the idea; the discussion until team members agreed with the most creative; the challenge of the distribution tasks per members; the excitement during the output presentation to the class and teacher.

4 The Study

4.1 Goals

The main goal of this study was to motivate students using their skills and abilities concerning creativity and technology use. Conversely, we were interested on students' engagement on HCI practices (application of theories, methods, concepts) during the design process of producing and artifact.

4.2 Participants

The participants in this study were undergraduate students of informatics and technologies degrees. Students should have succeeded on human computer interaction course, level I, and be attending HCI, level two. Twenty-two participants participated in the study. Students were invited to form a team with three participants. They had three months to accomplish the challenger. They could work during laboratory classes' timetables or at home. The work in progress should be presented and discussed weekly to the classmates. At the final deadline, they presented the whole work: the prototypes and the written report.

4.3 Teaching/Learning Processes

The method of teaching/learning processes adopted pointed to a student-centered, supported by incentives for self-learning and integration of knowledge, preparing them for lifelong learning. The student was confronted with a set of theoretical and practical problems, based on real cases, for which they would need to present solutions. They analyzed scientific and technical papers on the themes presented. Besides these works, there should present a mini project, target multiple stages of evaluation.

4.4 Constraints

During several years students were used to attend theoretical classes, paying attention to the teacher exposition and presentation. Each course had two hours for theoretical exposition and three hours for practical work. At the laboratory classes, usually, they were asked to practice the theoretical concepts previously taught through proposed exercises. Assignments and projects were made out of classes. We realized that they had a low participation at theoretical classes, and at practical ones, they were working just for the demanding tasks or exercises. These students had a computer science background. For them, HCI is not strictly software development class, but a multi-disciplinary course rooted on social sciences, e.g. a course to learn cultural subjects.

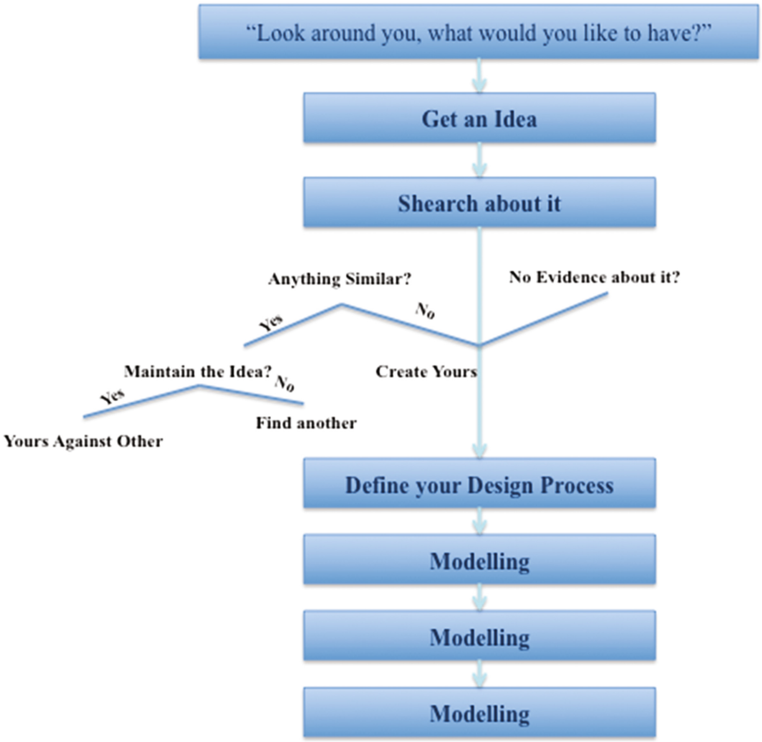


Fig. 1. Design settings

4.5 The Challenger

A challenge was proposed to the students: look around (search) and invent/re-design things that you would want to use yourselves. The final objective was the design of an interactive artifact supported on creative activities, - HCI theoretical and design-based research – and to develop the prototype or final product. Corroborating Goldschmidt the specific design goals were neither to prescribe specific design goals nor to refer to knowledge categories to be attended to [4]. Students defined their own design goals. The teacher's role was the design settings definition (Fig. 1). Creativity is not taught at best, it can be identified, encouraged and supported [2].

5 The Creative Design Thinking Process

The first stage of the creative design thinking process was the problem definition. They defined the idea to be developed. In the second stage, students got information to understand contexts, the existence, or not, of the proposed interactive artifact; they searched, sketched, they took photographs, they interviewed colleagues and other people about the relevance of the proposal. Then, in the analysis stage they did diagrams and other sketches and they searched about technologies to be used for the prototype development. They listed the limits on the design according the available resources and constraints. At the development stage, students created solutions to share with other students, and they discussed ways to solve the problem each team had. Then, they started the technology integration on the design idea. Finally, the prototype was examined, evaluated and the constraints were clearly identified (if they existed) to produce the final artifact.

Moreover, they completed each of the previous steps and they documented their work, in a report format; they presented the state-of-the-art about each of the subjects related with its main idea; they described the similar interactive artifacts on the market; they made the conceptual design research carried on their work; they presented the arguments about the technological choices and decisions about the materials used. A section with the description of the design process was integrated and the whole sketches and draws were included. Finally, they revised the work, they stated the identified problems and, they described the proposed solutions. Figure 2 presents the creative design thinking process.

The consideration of four Cs (collaboration, creativity, culture, communication) within the design process was based on the first impressions gathered from the observation of students doing their work. Examining them using communication as a means of sharing knowledge and ideas; watching their preoccupation on building artifacts carrying messages of their own cultural backgrounds; observing the students collaborative environments where each task was executed by sharing interests; examining their common values, meanings and feelings which were a constant influence among the teams participants. The design process stages were developed within the four Cs concepts [23]. Each concept could be attached to one of the stages but it could also make part of other stages. The phases of the creative design thinking process were derived either during the design brief or the data analysis. The data were analyzed

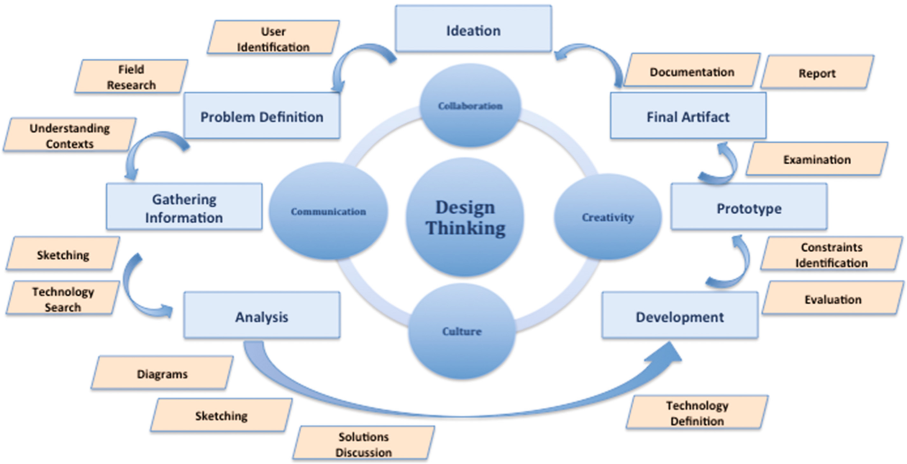


Fig. 2. Creative design thinking process

based on the final report submitted by students, and also on the group discussion during the artifact presentation.

6 Discussion

The meaningful interactions between students and digital products were seen as a means of asking questions through design, which means that they considered how to use technologies to facilitate the design process or how to apply technologies as a means to allow other kind of interactions.

Through the artifacts' analysis, as well as, by the design process and the observations carried on along this study, some reflections about creativity and technology's use were suggested.

Creativity can be affected/supported by technology: when students did not know how to use a software package, or when they had not enough practices with it. In the later situation, sometimes they rejected it and they used 'ad-hoc' support system. Moreover, creativity was supported by technology when students were confident using the software. In this case, they speeded the task on hands, and they got better output designs. The technology was also used as the artifact instead of a tool. They used those that they knew better to use.

Our model of creativity was composed by the design thinking phases carried on by students: Searching to get the idea, searching literature and digital sources to be inspired; Debating data results to understand and find relationships; Collaborating with colleagues and teacher to support their knowledge; Thinking to get combination of ideas and arguments to explain the chosen option; Exploring solutions either making use of technology as a tool or developing its own tools; developing artifacts step by step; presenting the results in engaged way.

Some of the design guidelines were to understand the materials, the people, the interaction, the context, the designer (student) himself and then the use of the artifact. The student pretends to behave as a final user. They were involved on the design process, the environment, using their imagination and creativity. The interactions between them and artifacts were consequently, fundamental to the creative processes. The better understanding of these should be valuable in the design of interactive systems for creative activities.

7 Results

Students' approach demonstrated that design thinking is a creative process, which focuses on problems and solutions simultaneously. It combines contextualizing problems, generating insights and executing solutions. As stated before, the design thinking process involved problem definition, field research, idea generating, storyboarding, frequent prototyping and narratives as a mode for engaging the participants and motivating the action. In our method, the design thinking process helped students to take as many ideas as they had, generating and figuring out how to determine which ones were more likely to produce the specific kinds of outcomes they wanted. The design thinking creative process differed from teamwork to teamwork.

The final artifacts were very creative. For example: an Electronic shopping cart, a smart cart developed to help people controlling their expenses. It registers the amount to be spent at the end of the shopping activity. An interactive mirror providing a different kind of user interaction through the information it can get on it: meteorology information, agenda, video or voice call from email or Skype and information about the user weight. Another artifact was a medical bracelet for patients' clinical history, allowing measures of cholesterol, glucose and blood strain. To search objects inside a handbag they created a search engine. They designed an interactive mattress with the shape and main functions of an ordinary mattress, but able to interact with users. The device shall interpret the users' rest position and subsequently temperature is generated appropriately according the users desires. Basically, the artifact interprets when the user has cold and it rapidly generates a comfortable temperature, the same will happen when the user has hot. These were some of the examples developed by the students. To illustrate one of them we present Fig. 3, which is a Central Process Unit (CPU), inserted on a microwave. The main idea is the concept of reuse and of 'modding' applied to computers, the reuse of unused appliances. The aesthetics is observed on the storage of computer components and cables management inside a microwave.

The presented interactive artifacts could make part of those that we have at our disposal in our daily life, however, for different reasons they don't. The students made the role of designers and they accomplish the proposed challenger. They were satisfied of their ability to innovate, to develop creativity, and to present powerful ideas. In the 'real' world sometimes these ideas win: design prizes, academic results, motifs for discussion, and topics for research. Conversely, the rate at which these ideas achieve commercial success is low. Many of the ideas die within the place where they were developed (companies, universities, home) never becoming a product. Among those that become products, a good number never reach commercial success. The discussion

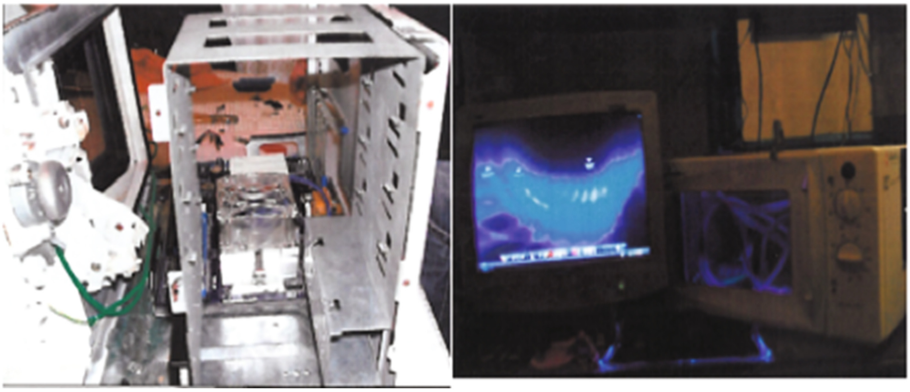


Fig. 3. CPU on microwaves

about this complex situation is out of the scope of this paper. Some of the ideas seemed foolish ones.

Although people may believe that their behavior or what they express in an artifact is creative, others may not accept either as such. However, if we consider great ideas of the past and how they were generated, and if we judge them by appearance we will arrive to the conclusion that some of the ideas were engines of creativity. Nevertheless, those design ideas and discovers permitted the development and evolution of our society and the artifacts contributed to improve some of our quality of daily life.

We are satisfied with the achievements on this study: students were persuaded to be creative. They did the work in an environment of relaxation and pensiveness, since although the challenger was compulsory, they could choose the sequence of tasks and they could work on the idea they proposed. As stated, students were free to work either at home or at classes. They improved their collaboration and communication skills. In collaborative work, participants are encouraged to select the best in each idea, proposal, or plan, build on the ideas of others; they can integrate and correlate ideas. Students were facing the process of learning by doing, applying the theoretical concepts, by free order, presented in the tutorial classes. They presented the creative design thinking process using the recommended tools or others they found adequate. Other presented ideas they did pass from the prototype phase and have been developed and applied in specific environments. They used any medium or technology to express and communicate meaning appropriate to their task and the people involved. It was found that they used, as support, several dimensions of creativity: exploration, self-expression, engagement, enjoyment, the results, and collaboration.

7.1 Lessons Learned

Technologies for creative interactions are challenges in terms of spaces, time, portability, connectivity and flexibility. Students can engage in a range of activities to working together to build ideas and artifacts.

According our observations, technologies were tools, which had two different purposes, either to help students to draw, sketch, programming or as inspirational artifact to apply/develop the new idea of the future artifact. They used technology according their skills either as a tool to develop and present their work or as a creative artifact technology based. When they were less able to use the technology they created other strategies based on their creative skills: they switch media.

Students may learn with the development of personalized technology systems by looking at creativity in action. There are enormous opportunities to create innovative technologies expanding the repertoire of tools that increase the creative process.

Analyzing the output artifact, we concluded that, in general, students are concerned with software development applications or prototypes that aim to sort social aspects of life in the area of health, social care. Therefore, they wanted to create something unexpected that challenges its creative activity trying to predict the future concerning user needs: things that they could imagine and that were not yet created either because of missing resources or missing skills to do that. But in anyway they presented the design process of it. Students improved their collaborative and communication skills along the task on hands.

The result of this case study shown us that the changes occurred fundamentally in the students' creative thought, the way they found to sort problem, attained goals using more or less the technology. The change did not occurred neither with the participants nor with the techniques or even with the design process.

8 Conclusion

In our experience we found that, in general, students in technologies and computer science courses, at a polytechnic institution, studying HCI subjects were not motivated enough to attain the goals of creativity within those courses. We proposed students a new methodology, putting them at the center of the process, asking for their creativity skills, finding a team common idea, developing it, applying the theories and concepts learned on HCI course, level I, and coming back with an output. We find this challenge very interesting, since the results shown student more engaged on the process and even working for completion aiming with colleagues. We consider that this study is a starting point of our knowledge acquisition and understanding of the situations that worried us. The study convinced us that this method should be continued with other teams looking at the creative process, the creative person and the creative product. The creative process did not always evolved in a sequence of states. The process was seen as series of clarification from idea presentation to the achievement.

The work goals were: looking at something that surrounded the participants, an event, an artifact, a problem someone had to sort in a team fashion. They should do design research applying their creative practices to present an output. Activities to support creativity and exploit the features of digital technologies: developing new ideas, making connections, creating and making, collaboration, communication and evaluation.

The results even without great novelty, they permitted to improve students' motivation and to obtain data which was compared with oldest one (teaching methods in previous years) and arrive to the conclusion that students are more engaged with a course work when they design the content of their own assignment.

Acknowledgements. We thank all the volunteers who participated in this study.

References

1. Lienhard, J.H.: How Invention Begins: Echoes of Old Voices in the Rise of New Machines, 1st edn. Oxford University Press, New York (2006)
2. Boden, M.: The Dimensions of Creativity. MIT Press Cambridge, London (1994)
3. Koestler, A.: The Act of Creation. Dell, New York (1964)
4. Gough, H.G.: A creative personality scale for the adjective check list. J. Pers. Soc. Psychol. **37**(8), 1398–1405 (1979)
5. Guilford, J.P.: Creativity. Am. Psychol. **5**, 444–454 (1950)
6. Amabile, T.M.: The Social Psychology of Creativity. Springer Press, New York (1993)
7. Tanaka, A.: Interaction, experience and the future of music. In: O'Hara, K., Brown, B. (eds.) Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies, CSCW, vol. 35, pp. 267–288 Springer Press Cham (2006)
8. Terry, M., Mynatt, E., Nakakoji, K., Yamamoto, Y.: Variation in element and action: supporting simultaneous development of alternative solutions. In: Proceedings of CHI 2004, pp. 711–718. ACM Press (2004)
9. Satell, G.: The Creativity Post. http://www.creativitypost.com/tag/design-thinking (2014). Accessed 27 Oct 2016
10. Grudin, J.: Three faces of human-computer interaction. IEEE Trans. 2–18 (2005)
11. Grudin, J.: A moving target—the evolution of human-computer interaction. In: Jackson, J. (ed.) Human-Computer Interaction Handbook, 3rd edn., p. 40 (2012)
12. Sanders, L., Stappers, P.J.: Designing to co-designing to collective dreaming: three slices in time. ACM Interact. **21**(6), 24–33 (2014)
13. Kostakos, V.: The big hole in HCI research. Interactions **22**(2), 48–51 (2015)
14. Reeves, S.: Human-computer interaction as science. In: Proceedings of 5th Decennial Aarhus Conference (Critical Alternatives). ACM Press (2015)
15. Rogers, Y.: HCI theory: classical, modern and contemporary. Synth. Lect. Hum. Cent. Inform. **5**(2), 1–129 (2012)
16. Kim, S.: Interdisciplinary cooperation. In: Laurel, B. (ed.) The Art of Human-Computer Interface. Addison-Wesley, Reading (1990)
17. Pan, Y.: Fashion thinking: fashion practices and sustainable interaction design. Int. J. Des. **9**(1), 53–66 (2015)
18. Pan, Y., Roedl, D., Blevis, E., Thomas, J.: Re-conceptualizing fashion in sustainable HCI. In: Proceedings of the DIS 2012, pp. 813–815. ACM (2012)
19. Barbot, B., Besançon, M., Lubart, T.: Assessing creativity in the classroom. Open Educ. J. **4** (Suppl1), 58–66 (2011)
20. Runco, M., Jaeger, G.: The standard definition of creativity. Creat. Res. J. **24**(1), 92–96 (2012)
21. Florida, R.: The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life. Basic Books, New York (2012)
22. Eysenck, H.J.: The measurement of creativity. In: Boden, M.A. (ed.) Dimensions of Creativity, pp. 199–242. The MIT Press, Cambridge (1994)

23. Lopes, A.: Design as Dialogue: Encouraging and Facilitating Interdisciplinary Collaboration. VDM Verlag Dr. Muller, USA, UK. Springer (2009). ISBN 978-3-639-11713-4
24. Sawyer, R.K.: Explaining Creativity: the Science of Human Innovation, 2nd edn. Oxford University Press, New York (2012)
25. Csikszentmihalyi, M.: Creativity: Flow and the Psychology of Discovery and Invention. Harper Press, New York (1996)
26. Franken, R.: Human Motivation, 6th edn. Cengage Learning, Boston (2006)
27. Burnette, C.: Conceptual Modeling, Imagination, and Analysis in Relational Thought (2014). https://independent.academia.edu/CharlesBurnette
28. Mahdi, R., Sukarman, S., Yok, M.: Fostering creativity through innovation engagement in science and technology education: case study of Universiti Teknologi MARA students, IOSTE BORNEO 2014. Procedia Soc. Behav. Sci. **167,** 256–260 (2015). Elsevier Ltd (2015)
29. Lemos, J., et al.: A systematic mapping study on creativity in requirements engineering. In: Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC 2012
30. Cherry, E., et al.: Quantifying the creativity support of digital tools trough the creativity support index. ACM Trans. Comput. Hum. Interact. **21**(4) (2014). Article No. 21
31. Latulipe, C.: The value of research in creativity and the arts. In: Proceedings of the 9th ACM Conference on Creativity and Cognition, Sydney, Australia, 17–20 Junes (2013)
32. Banaji, S., Burn, A., Buckingham, D.: The Rhetorics of Creativity: A Review of the Literature. Arts Council England, London (2007)

Development of an Automated Pressure Sensitive Thermesthesiometer and Its Application in Characterizing the Thermal Response of Human Tissue with Respect to Warm Surfaces

Paul Soldate(✉) and Jintu Fan

College of Human Ecology, Cornell University, Ithaca, NY 14850, USA
{pcs229, jf456}@Cornell.edu

Abstract. An advanced thermesthesiometer has been developed to measure the heat flux, touch-force, and internal temperature profile of a simulated body part coming into contact with a warm surface. A suitably chosen material with properties resembling those of human tissue is used for bare skin simulation and is thermally regulated using a programmable logic controller (PLC). An automated driving mechanism provides variable touch-occurrences and pressures over specially engineered test-surfaces with varying effusivities and surface roughness. Each test-surface is thermally regulated at a specific temperature to aid in quantifying prior research on human subject testing and pain threshold measurements. This study not only emphasizes the instrument design but also demonstrates its potential use in human-subjective feedback testing with respect to steady state and transient heat transfer processes.

Keywords: Thermesthesiometer · Human factors · Human-systems integration · Instrumentation heat transfer

1 Introduction

Computing devices and wearable electronics dissipate heat. Depending on the device, its functionality, placement, and overall power requirement, the total heat transferred from such a device to the end user may vary greatly depending on numerous factors. While it may be ideal to theoretically estimate the total energy transferred using fundamental heat flow theory or power consumption analysis, such calculations are not always possible. This is because the thermal and surface properties of many electronic platforms or devices are often unknown since they are usually composed of complex composite shells or housed in materials that vary in surface texture. Therefore, it is advantageous to develop a standard protocol to establish accurately the temperature response of user-device interactions and to gauge safety concerns as well as user-thermal comfort.

In consideration of 2^{nd} and 3^{rd} degree burns from bare skin contact, an extensive study has been reported [1]. Such hazards have been shown to be dependent on three

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_36

main variables: the relative temperatures of interfacing materials, their effusivities, and their contact resistance [2, 3]. Other research has established some safety guidelines for touch temperature limits and bare-hand pain thresholds [4]. Likewise, the International Standards Association has defined a method for assessing human response with respect to hot surfaces [5]. In addition, the American Standards for Testing and Materials has also developed a standard for determining skin contact temperature for heated surfaces [6]. However, most of these safety standards are based on temperatures at which bare skin will begin to burn and do not give information on temporal sensation of significant time intervals or relatively moderate temperatures. Moreover, no discussion is given on the effects of non-uniform temperature distributions within composite materials. However, all of these variables will significantly affect the heat transfer process.

In addition, prior studies provide little information regarding the duration of skin contact. This variable will also affect the thermal sensation because it takes time for heat to penetrate each layer of skin. Human tissue is composed of three primary layers: the epidermis, the dermis and the hypodermis [7]. While the epidermis sits on the outermost layer of the skin, the dermis is set 80 μm underneath, where the nerve endings reside. Therefore, in many cases, the outer layers of human tissue can burn quickly before thermal perception is even activated. With this in mind, being able to measure the transient effect during the heat transfer process will be of great interest.

Likewise, many seemingly smooth surfaces are actually composed of microscopic peaks and valleys with a number of air gaps which act as insulation [8]. However, prior research also has yet to quantify contact resistance with respect to touch perception and pain thresholds. The touch-force and surface area contribute heavily since they effectively determine the number of contact points over which the conductive heat transfer process occurs. Although Hatton et al. showed the importance of contact resistance theoretically, his method requires a material dependent heat transfer coefficient that is not always available [2]. Therefore, it will be important to be able to determine experimentally the skin contact force to assess accurately a material's thermal acceptability for various end-use conditions.

Another important aspect of thermal sensation is determined by the effusivity of a material:

$$\lambda = \sqrt{\kappa\rho c_p}. \tag{1}$$

where, κ is the thermal conductivity in watts per meter·kelvin $[\frac{W}{mK}]$, ρ is the density of the material in kilograms per cubic meter $[\frac{kg}{m^3}]$, and c_p is the specific heat capacity in joules per kilogram·kelvin $[\frac{J}{kgK}]$. Stoll et al. demonstrated the importance of a material's effusivity with respect to bare skin thermal contact by determining the amount of time before pain occurred in human subjects coming into contact with various warmed materials [3]. Hedge and Zhang gave a comprehensive literature review on similar subjective tests in their 2016 report to Intel [1].

Historically, while human subjective testing has been essential in investigating the thermal response in different parts of the human body, they are less ideal in characterizing materials for end-use conditions due to the inherent variability of real human body parts, habituation, and the subjectivity of human sensation. Hence, to aid in the analysis, an engineering approach is of primary interest. It is therefore, the present goal of the authors to develop an instrument to characterize accurately the heat flux and thermal response of various composite materials, each with different surface textures, over a range of temperatures, as well as various contact pressures.

2 Design Considerations

In 1972, Wu studied the properties of human skin and determined its thermal conductivity and thermal diffusivity [9]. From this, Marzetta implemented a thermesthesiometer design (Fig. 1), which was able to assess skin burn potential from various surfaces [10]. Such a device is usually composed of an insulating shell, an internal heating element and a thermocouple for feedback temperature control, and is filled with a silicone with a thermal inertia comparable to that of human tissue. However, while of great utility for its time, such a device cannot measure the force of thermal contact nor is it adaptable to high throughput test measurements. Therefore, to investigate platforms such as wearable electronics and quantify user-experience thresholds in regard to advanced materials, an automated thermesthesiometer with contact-force sensitivity, heat flux read out and high accuracy PID control has been developed.

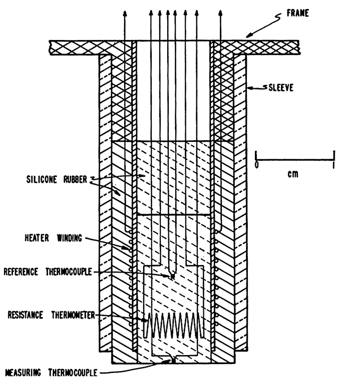


Fig. 1. Marzetta's depiction of the thermesthesiometer, including a cylindrical chamber and internal heating system for feedback temperature control.

Similar to Marzetta, a cylindrical chamber is implemented for symmetry. Likewise, a set of resistive heating elements, wrapped around a concentric, conductive path (CCP) is chosen. A thermocouple (PT1000 RTD) is secured at the CCP surface to monitor the internal temperature of the instrument, which is controlled by a data acquisition (DAQ) system using a proportional-integral-derivative (PID) algorithm. The CCP is then housed within a concentrically insulated capsule to reduce transient heat loss during tests. A heat flux sensor is attached at the surface of the CCP and a final thermocouple is installed at the tip of the instrument near the edge of the capsule to monitor the transient response within the simulated epidermal gap before and after thermal contact occurs. The Capsule is then filled with a room temperature, vulcanizing material (RTV) at 40% Polydimethylsiloxane (PDMS) and 60% Aluminum Oxide (Al_22O_3) to obtain a similar thermal inertia to that of human skin as described by Wu and Marzetta [9, 10].

Within the thermesthesiometer chamber, the heat flux and temperature sensors can be attached at any distance from the simulated skin surface, e.g. 0.080 mm deep, similar to the nerve endings under the epidermis. However, to aid in determining the best architecture for sensor placement, a 2D heat transfer simulation using the instrument, which includes the sensors, is created using ANSYS.

In bringing a simulated thermesthesiometer, initially set at 32.5 °C, into contact with an Aluminum test surface at 44 °C, the results showed a large instability when the internal sensors were placed at very short distances (Fig. 2). After optimizing the configuration, it was clear that the sensors could be placed at 10 mm from the contact surface and obtain a stable heat flux profile (Fig. 3).

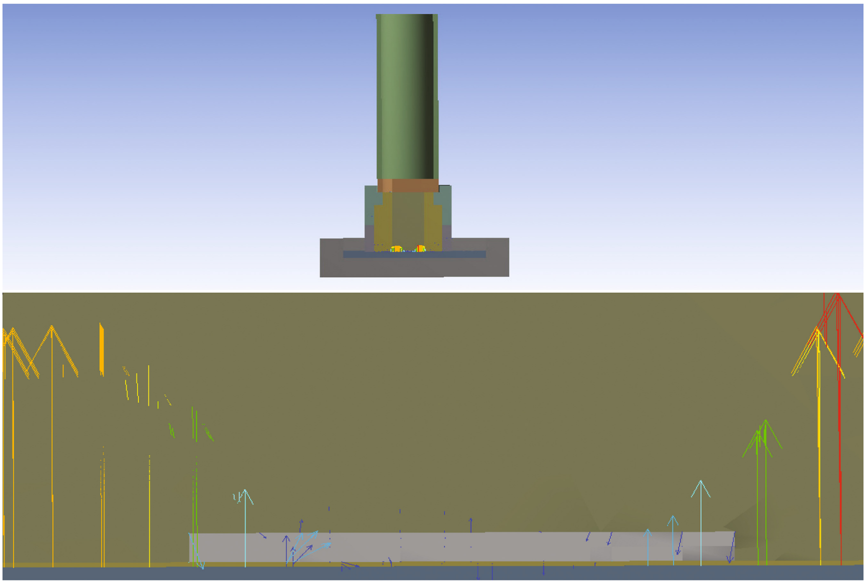


Fig. 2. For the extreme case, and in direct correlation with human skin tissue, the heat flux sensor is attached .080 mm from the simulated skin surface. The results are highly unstable.

A thermally insulated platform holding the heated material is driven by a Kapton (low thermal inertia) resistive heating element, which is then housed in an insulated square shell in-lay. Composite materials can be inserted into the in-lay and compressed tightly by a specially machined lid-press, which is secured to the insulated platform base. For multiple composite tests, a thin layer of thermal grease can be added between

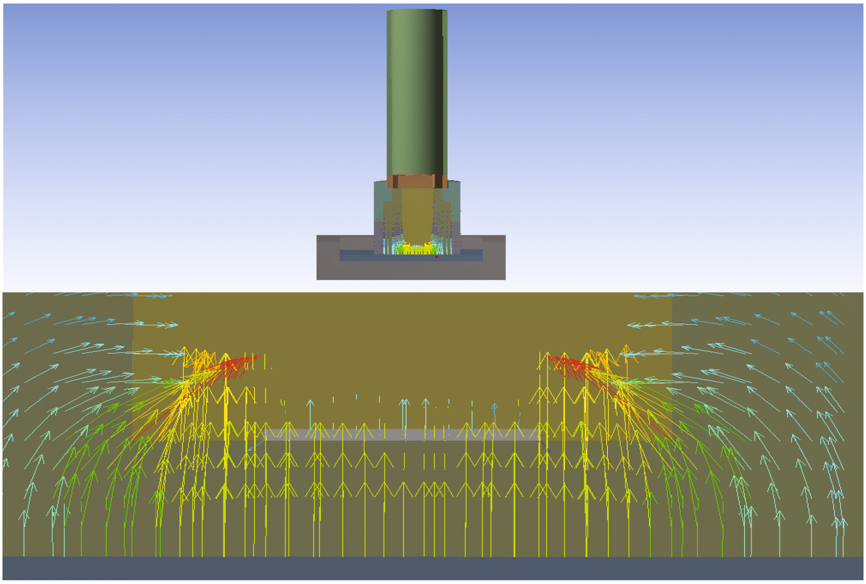


Fig. 3. The heat flux sensor is attached 10 mm from the simulated skin surface and shows much more stability.

layers. A PT1000 RTD sits at the surface to monitor and control the surface temperature. The heating element is then powered by a solid-state relay and a regulated power supply, which is then driven by the DAQ system and PID software.

The instrument is attached to a nonrotating slotted piston, which is threaded into a cylindrical driver connected to a stepper motor. The stepper motor rotates the cylindrical driver at various rates for touch velocity and acceleration, which creates the thermal contact between the thermesthesiometer and the heated test surface secured within the insulated base. A load cell is mounted to the insulated platform and secured to the device structure. As the thermesthesiometer makes contact with the heated material, the load cell registers a contact force. The stepper motor stops the motion when the pressure reaches a predefined value. For consistency, the overall process is automated for use when multiple tests are required under similar conditions.

The DAQ system separately stabilizes the temperature of both the thermesthesiometer and the test-surface before thermal contact occurs. During this time, a statistical algorithm is used to determine the parameters best suited for the environment in which the experiments are completed. After steady-state conditions are achieved, the parameters are utilized to fix the duty cycle and therefore the amount of power generated to heat each element in the system during contact. This is done to simulate metabolic functionality while minimizing PID overcompensation during heat loss or gain during contact. Figure 4 demonstrates the power input to both the instrument and the test surface while steady state conditions are achieved as well as during contact, when the duty cycle relayed to each of the heating elements remains constant.

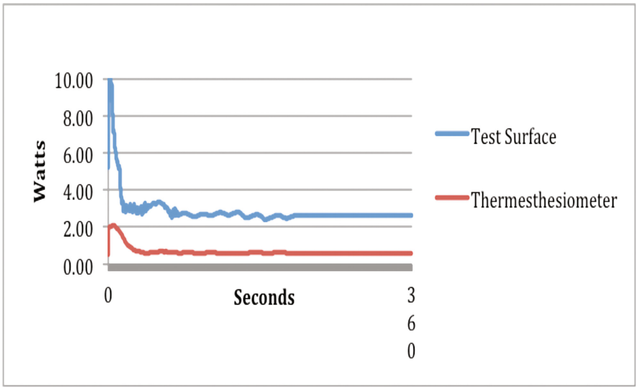


Fig. 4. The power input for each heating element while steady state conditions are achieved, as well as during contact, when the duty cycle relayed to the power supply remains fixed.

The instrument has three shelves (Fig. 5). The top shelf holds all electronics, e.g. the solid state relays for PID controls, an LCD to read the load cell easily, DAQ hardware, a stepper motor, a heat flux data logger, power supplies, a cooling fan and the mechanism for driving the instrument up and down. The middle shelf guides the piston while keeping the instrument from rotating with the driver. The bottom stage secures the load cell and the heated material to maintain a constant pressure during contact.

Fig. 5. This image shows the overall structure of the automated instrument.

3 Results

3.1 Thermal Inertia Comparison of Composites

To analyze the thermal response of human skin tissue coming into contact with various heated materials, we have devised several composite structures composed of a 1 mm 6061 Aluminum sheet with increasing thicknesses of ABS plastic layers secured at the Aluminum surface. Since the control variables of the instrument are automated, we may perform several tests periodically on each material to test for consistency. To see this, refer to Figs. 6 and 7, which show the temperature response and heat flux profile of four repeated tests, each under the same initial conditions, i.e. the thermesthesiometer held at 32.5 °C and the material held at 44 °C. Figures 6 and 7 give us confidence that the instrument is functioning as expected. The stability at longer time scales demonstrates the fixed duty cycle and power as previously discussed (Fig. 4). Notice, as expected, that increasing the plastic insulation in each heated material decreases the overall thermal response and transient intensity.

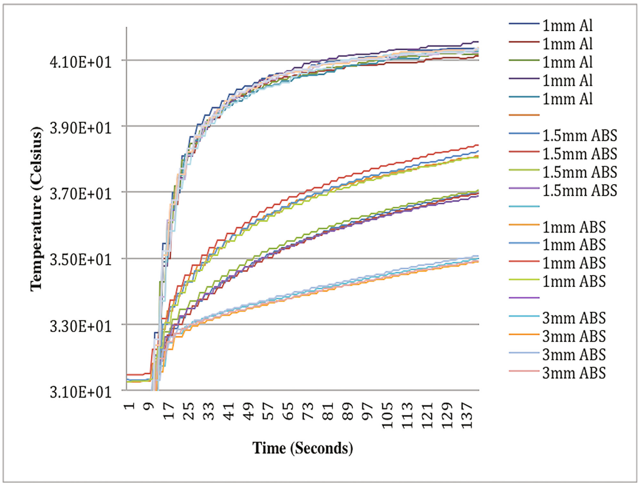


Fig. 6. The temperature response of several materials initially heated to 44 °C and coming into contact with the thermesthesiometer held initially at 32.5 °C. The legend gives the material thickness in addition to a 1 mm 6061 Aluminum sheet, e.g. "1.5 mm ABS" is a composite made of 1 mm Aluminum underneath 1.5 mm of ABS plastic, which is contacted by the thermesthesiometer.

3.2 Quantifying Human Subjective Studies

The results above show a strong correlation between the thermal effusivity of a heated surface and the thermal response of bare skin contact. To demonstrate the vast differences between two common materials and to begin taking steps toward quantifying the time-to-pain thresholds of various human subjective experiments (discussed above) we now turn our attention to two common materials: aluminum and Masonite.

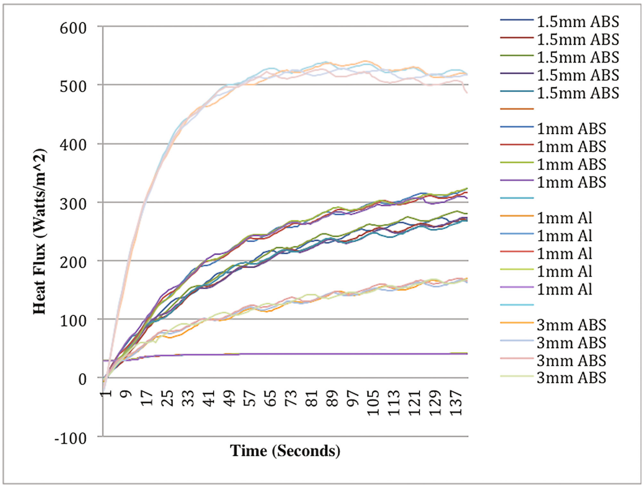


Fig. 7. The heat profile of several materials initially heated to 44 °C and coming into contact with the thermesthesiometer held initially at 32.5 °C. The legend gives the material thickness in addition to a 1 mm Aluminum sheet, e.g. “1.5 mm ABS” is a composite made of 1 mm Aluminum underneath 1.5 mm of ABS plastic, which is contacted by the thermesthesiometer.

Figures 8 and 9 show the touch response of a sheet of 6061 Aluminum as well as a sheet of Masonite held at 44 °C. Notice the difference in the transient response. Where Aluminum shows a rapid increase in skin temperature, i.e. toward 37 °C within the first three seconds, Masonite drives the simulated tissue to only 37 °C after 10 s. From experiments like this, it is evident that through correct calibration, human subjective studies, such as Stoll’s time to pain thresholds, may be quantified using the instrumentation.

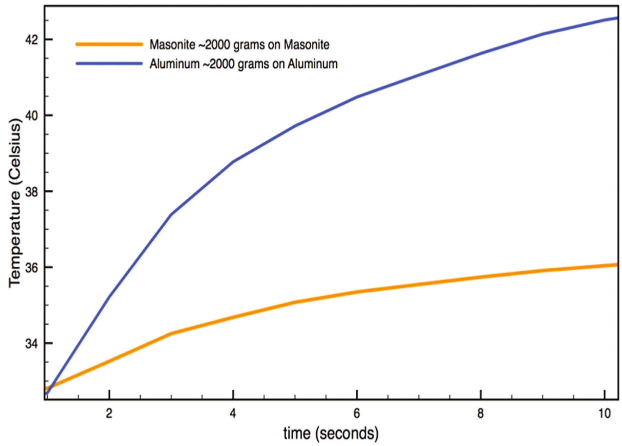


Fig. 8. This plot shows the temperature response of the thermesthesiometer when touching a 44 °C sheet of Aluminum and Masonite.

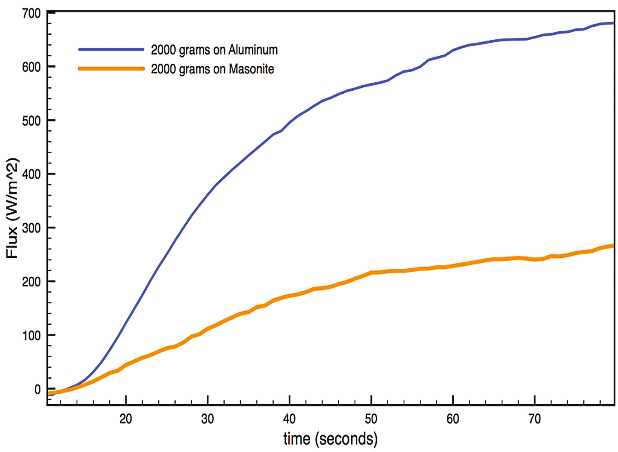


Fig. 9. This plot shows the heat flux profile of the thermesthesiometer when touching a 44 °C sheet of Aluminum and Masonite.

3.3 Touch-Force Sensitivity

While Figs. 8 and 9 provide insight into the effects of thermal inertia on touch sensation, it is not evident how the contact force will apply. To see this we have run several touch-sensitivity tests using the automated instrument, initially set at 32.5 °C, and brought into contact with a 1 mm sheet of 6061 Aluminum heated to 44 °C. Figures 10 and 11 show the thermal response and heat flux profile for load cell readouts of 10, 40, 200, 1250, and 2130 g. From this, it is evident that pressing a heated material harder increases the microscopic points of contact and will thus accelerate the heat transfer process. This is seen in the transient response curves for the entire duration of contact. It is also evident from Figs. 10 and 11 that there is a maximum saturation pressure to which the heat transfer process converges. From this, and for consistency, all of the results reported above were obtained using touch forces at the saturation point of contact, e.g. 2000–4000 g.

3.4 Surface Roughness Profiles

In examining composite materials in regard to their end-use applications, it is sometimes unimportant how much force is applied during thermal contact as the surface roughness of the material under investigation may be vastly pitted. From this, the touch force option on the instrument can be set to contact saturation ($\sim$4000 g), and several composites can be compared. Figure 12 shows the heat flux profile of a 1 mm rough ABS plastic sheet as well as a 2 mm smooth ABS plastic sheet. Likewise, Fig. 12 also compares a 2 mm rough ABS plastic sheet with a 4 mm smooth ABS plastic sheet. Each composite is set upon a 1 mm sheet of 6061 Aluminum and set at 44 °C, while the thermesthesiometer coming into contact is initially set at 32.5 °C. From the figure, it is easy to see that the rough surface architecture mitigates the heat transfer process and

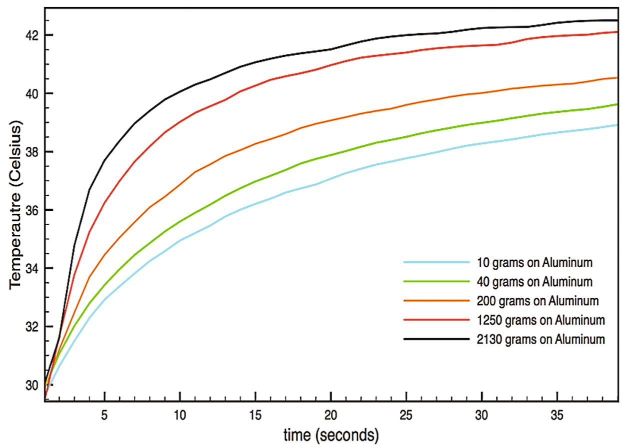


Fig. 10. The temperature response of the thermesthesiometer coming into contact with a 1 mm sheet of 6061 Aluminum under several different contact forces.

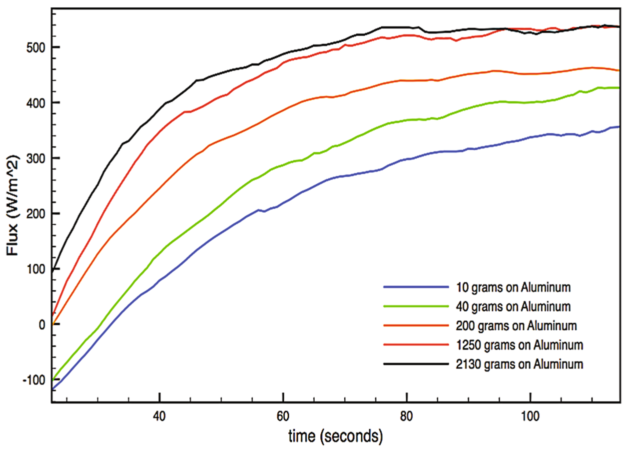


Fig. 11. The temperature response of the thermesthesiometer coming into contact with a 1 mm sheet of 6061 Aluminum under several different contact forces.

maintains nearly the same heat flux profile as that of a smoother surface that is twice as thick. This is the case for 1 mm rough vs. 2 mm smooth ABS, as well as 2 mm rough ABS vs. 4 mm smooth ABS.

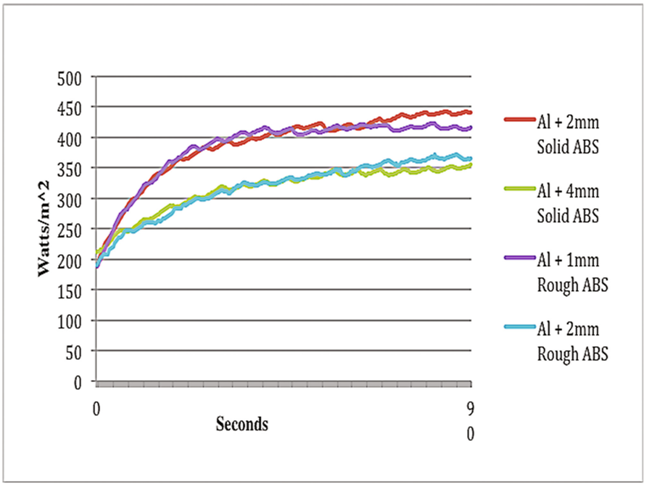


Fig. 12. Heat flux profiles for an initially set 32.5 °C thermesthesiometer coming into contact with various composites with smooth and rough surfaces initially set at 44 °C.

4 Conclusions

An advanced thermesthesiometer with touch-force sensitivity and feed back control has been developed to characterize materials for end-use conditions and to quantify thermal hazards in human subjective testing. The instrument provides a realistic simulation of a body part coming into contact with a heated test-surface while measuring the temperature response, heat flux profile and contact force. Results show that the instrument can sensitively differentiate between several materials in terms of effusivity, surface roughness and contact pressure using internal sensors and real-time monitoring.

While several materials are compared in this study, the characterization of other materials, e.g. glass Teflon, Steel, Titanium, Carbon, polycarbonate, etc. under a range of temperatures will be of interest for further comparisons in human subjective testing. Future work might also include calibrations for simultaneous human subjective testing to quantitatively analyze touch-force dependence.

Acknowledgements. The authors would like to thank Intel, Mark Macdonald, Mondira Pant, and Wu Yuen Shing for useful discussions. This work was supported by a grant from Intel and performed at Cornell University.

References

1. Hedge, A., Zhang, H.: Warm surfaces of electronic devices and human thermal responses, Final report for Intel (2016)
2. Hatton, A.P., Halfdanarson, H.: The role of contact resistance in skin burns. J. Biomed. Eng. **4**(2), 97–102 (1982)

3. Stoll, A.M., Chianta, M.A., Piergallini, J.R.: Thermal conduction effects in human skin. Aviat. Space Environ. Med. **50**(8), 778–787 (1979)
4. Ungar, E., Stroud K.A.: New Approach to Defining Human Touch Temperature Standards. NASA/Johnson Space Center (2010). http://naca.larc.nasa.gov/search.jsp?R=20090042589&qs
5. ISO 13732-1 Ergonomics of the Thermal Environment - Methods for the Assessment of Human Responses to Contact with Surfaces, Part 1: Hot Surfaces, ISO, Geneva, Switzerland (2006)
6. ASTM Designation: C1057 – 03 Standard Practice for Determination of Skin Contact Temperature from Heated Surfaces Using a Mathematical Model and Thermesthesiometer (2010)
7. Alberts, B., Wilson, J., Tim, H.: Molecular Biology of the Cell. Garland Science, New York (2008)
8. Cengel, Y.: Introduction to Thermodynamics and Heat Transfer, 2nd edn, p. 374. McGraw Hill Companies, New York (2008)
9. Wu, Y.C.: Materials property criteria for thermal safety. J. Mater. JMLSA **7**, 573–579 (1972)
10. Marzetta, L.A.: A Thermesthesiometer – An Instrument for Burn Hazard Measurement. In: IEEE Transactions on Biomedical Engineering, September 1974, pp. 425–427 (1974)

Apply Petri Nets to Human Performance and Workload Prediction Under Multitask

Peng Wang[1,2], Weining Fang[1(✉)], Beiyuan Guo[1], and Haifeng Bao[1,2]

[1] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Haidian District, Beijing 100044, China
c2t53y48@gmail.com, {wnfang, byguo, 12126341}@bjtu.edu.cn

[2] School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Haidian District, Beijing 100044, China

Abstract. Performance and workload prediction under multitasking are one of the most important aspects to consider during system design. This paper utilized a formal tool – namely petri nets, to predict individual's performance and workload when interacting with interfaces. The petri nets model is based on limited resources theory which indicates that certain attentional resources contribute to to several channels of perceptual, cognition and psychomotor. After interpreting the parameters and rules of a general model, a case study on Multi-Attribute Task Battery-II was investigated. Individual's performance was predicted by simulating the petri nets model, the instantaneous workload and total workload was calculated by identifying the attentional resources.

Keywords: Petri nets · Performance · Workload · Multi-Attribute Task Battery-II

1 Introduction

Human performance and workload evaluation have been studied for decades, turned into indicators of the design quality of human-machine interfaces, man-machine function allocation and tasks. As for the rise of complexity of system and tasks, operators need to deal with different tasks simultaneously, which leads to higher occurrence of human errors and safety issues. Various evaluation methods widely used like ACT-R [1] for predicting performance and workload.

Petri nets is a mathematical and graphical tool in modeling and performance evaluation of real systems with discrete events that are characterized as being concurrent, synchronous, asynchronous, distributed, parallel, nondeterministic, and stochastic, such as computer systems, communication networks and production systems [2]. It has been widely used in performance evaluation in many areas such as hybrid traffic, fuel cell systems and IT systems [3–5] for its ability in describing time parameters and allocation of resources. It may also use in predicting human performance for the inherent time and resource attributes of individuals.

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_37

In addition, it has great application potential, although not directly for workload assessment as a formal analysis tools. Firstly, petri nets have been applied to a variety of analyses and modeling of ergonomic tasks [6, 7], where task analysis and modeling are the essential part especially in some workload assessment approaches without human-in-the-loop. For a more accurate assessment, a clear definition of workload should be given. However, there is no unified definition of workload, three characteristics could be found in the definition of workload in previous studies: multi-dimension [8, 9], task demand [10] and resource (ability) of human [11]. We suggest that the definition of workload should be transformed from a literal expression to a symbolic expression, and finally to a formulate expression. Petri nets are suitable method of symbolic expressing workload. In addition, as an assistant tool, task analysis methods have influence on the accuracy of workload assessments. The different elements in task analysis methods will lead to different results. The elements in hierarchical task analysis [12] are "goal", "task", "plan" and "operation", while the GOMS (Goals, Operators, Methods, and Selectors) [13] consists of "goal", "operator", "method" and "selection rule". Petri nets have the ability that integrating related concepts and elements needed for workload description in one structure, and has a better unity than the other task analysis methods.

2 Methods

2.1 Petri Nets Based Model of Interaction Process

Given that the attention resources of individuals are limited, individuals cannot focus on many process, which are carried out simultaneously. There are also restrictions on the abilities of different attentional channels, such as: perception, cognition and psychomotor, which prohibits executing various complex behaviors, The perception/cognition/psychomotor processes demand time for resources and resources needed for relevant support channels. Moreover, the interaction process in a specific task is always caused by events of system, such as psychomotor behavior which will also change the state of system accordingly.

Identify the Components of Petri Nets. General petri nets are made up of places, transitions, directed arcs and tokens. This model contain four types of place, three of them are represent for states of resources of human. Attention resources place: used in placing the attention resources of individual, the capacity is equal to its initial value. Resources place for visual/auditory/cognition/psychomotor: indicate the ability of each channel of individual, their capacity is also equal to its initial value. Activation place: always connected with its corresponding transition, used in permitting fire. Another one type of place is used in expressing the states of system/machine.

Transitions in this model denote behaviors of human or events of interface, its firing lead to the consumption/transferring of resources or changing the state of system/machine. Tokens indicate the resources of human or signs of system, directed arcs mean the direction of resources flow of individual or the state flow of the system.

Time Parameters of Transitions. As mentioned in above, time was added as a parameter of transition in order to simulate real interaction processes. Durations were treated as the only exact form of transitions of human for that human behavior but did not have an accurate starting time. While starting times were always regarded as the concrete form of the time parameter of transitions of system, the results of human behaviors usually change the state of system, which will have a variety of durations. The starting times of events are easy to catch because they are set or can be found in log files of system. In this paper we proposed a simple way to determine the time of perception/cognition/psychomotor process, which includes the reaction times of task. We assume that there are two interaction processes A and B. Their perception/cognition/psychomotor behaviors are A1, A2, A3 and B1, B2, B3 respectively. For getting the precise time of each of them, eight combinations can be built as follows:

$$\begin{cases} A_1 + A_2 + A_3 = R_1, B_1 + B_2 + B_3 = R_2, A_1 + A_2 + B_3 = R_3, A_1 + B_2 + A_3 = R_4 \\ A_1 + B_2 + B_3 = R_5, B_1 + A_2 + A_3 = R_6, B_1 + A_2 + B_3 = R_7, B_1 + B_2 + A_3 = R_8 \end{cases}. \quad (1)$$

R_1 and R_2 are the reaction times of interaction A and B, and we have formed six new interaction process with reaction times R_3, R_4, R_5, R_6, R_7, R_8, which consist of the same interaction behaviors but with a different combinations. It can be found that the above eight equations just have six unknown parameters, any six of them could have a precise theory resolution or resolutions make a minimum error-sum. This equation set will obtain 48 different precise theory resolutions/resolutions make a minimum error-sum, the average of them can believe to be the precise time for A_1, A_2, A_3 and B_1, B_2, B_3.

Initial Position and Value of Tokens. Tokens in places are as quantitative representation of resources of human or states of machine. As mentioned in Sect. 2.1.1, attention resources place and channel resources places should have some tokens that reflect difference of individual, and the detail values could be defined by some theory like VACP rating scales [14, 15]. Tokens in places of machine should also be decided that reflect its initial states. The values of them are always equal to one.

Directed Arc's Weights. Two types of directed arc's weights can be seen in this petri nets model. One of them in representing the change of state of system. Their value is equal to one. Another type is used in interpreting the consumption/transfer/send back of tokens, and their values are related to the specific interaction behaviors.

2.2 General Description

The general description of petri nets based interaction model was shown in Fig. 1. Circles with yellow/white/green color represent the simplified states of machine, tokens and resources states of individual respectively. Arrow lines are the reflection of direction arcs, with a mark equal to their weights, or without mark for that their weights are equal to one. Rectangles with yellow color are the transitions of machine and white ones are as the transitions of human behaviors. Systems in reality may have a more complex internal structure while they can be concentrated to a two-state model when as

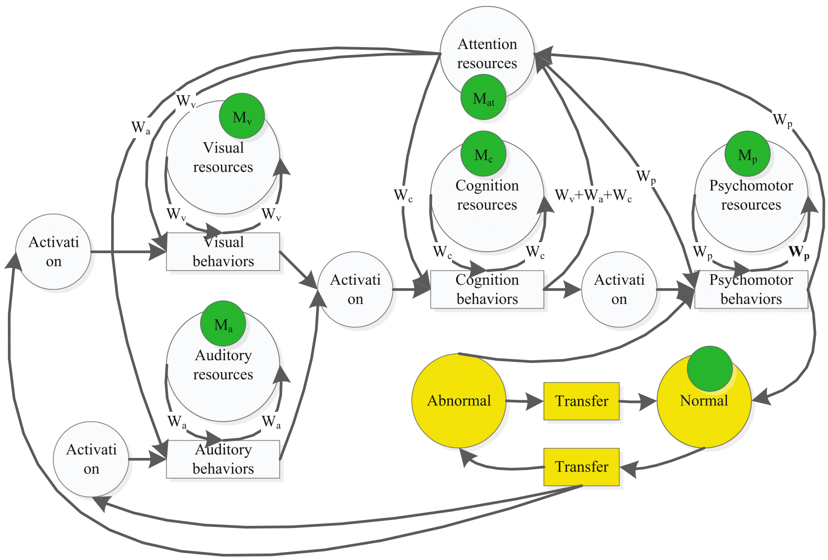


Fig. 1. General description of petri nets model.

an interface in interaction with human. As the yellow color in Fig. 1, when system change its state from normal to abnormal, the activation places tied with visual and auditory transitions gained a token, and the visual/auditory transitions are enabled now. The visual/auditory transitions can fire in case that tokens in attention resources/visual resources/auditory resources place are greater than or equal to the corresponding weights of directed arcs, that was explained in the formulas as follows:

$$M_{at} \geq W_a + W_v, \; M_v \geq W_v, \; M_a \geq W. \tag{2}$$

In some situations that the changing of states of system may only cause visual or auditory perceptions, thus the details of connection in the model should be also modified. The firing of visual/auditory transitions will cause the consumptions of tokens in attention resources place and a consumption-call back process of tokens of visual/auditory places, as well as add a token in the activation place tied with cognition behavior transition. The conditions of cognition behavior and the results caused by its firing are similar to the visual/auditory transitions, with the addition of the recycling of tokens to attention resource placement. This recycling of resources is triggered by cognition behavior and transition rather than visual/auditory transitions as per the multi-resource theory [9], where perceptual and cognition are in a same stage, thus resources will return when this step is complete.

3 Results on MATB-II

Multi-Attribute Task Battery-II (MATB-II) is the next generation of MATB [16], which has been extensively used in ergonomics experiments [17–19]. MATB-II contains four subtasks: System Monitoring (SYSMON), Tracking (TRK), Communication (COMMUN) and Resource Management (RESMAN). In SYSMON tasks, participants pressed the corresponding key, or clicked the icon when a green F5 icon turned off or red F6 icon turned on, or the indicator F1/F2/F3/F4 moved up and down independently on each value. For TRK, in MANUAL model, participants kept the target in the center of the inner box through a joystick, and in AUTOMATION model, participants did not need to do anything. COMMUN tasks require participants adjusted the corresponding frequency of the channel according to the voice when he listened to 'NASA 504'. The RESMAN task request participants adjusted eight pumps to keep the level of fuel in Tank A and Tank B within 200 ranges of 2500, and pumps could not be operated when they were in a fail state.

3.1 Integrate the Workload Parameters

Workload parameters described as directed arc's weights with tokens in place. This paper refers VACP rating scales [15] as a base to attribute the numerical values. In this scale, encoded specific behaviors belongs to different channels which have a value to measure its occupied resources of individual, and the maximum value of each channel are all equal to seven. The resources needed for channel resources transitions determine the directed arc's weights. For instance, if a behavior in visual channel was identified with V3, the weight of this arc should be equal to 40. Tokens in attention resources place has a value 105 as an initial assumption, and will be modified by subsequent performance evaluation of individual. Tokens in channel resources places are all equal to 70 for the max requirement of resources according to the scales.

3.2 COMMUN Task Modelling

This paper apply COMMUN task, the most complex of the four tasks in MATB-II, and includes all the four channel behaviors, as an example here to explain our model. As shown in Fig. 2, the task interface was described in a conflict structure. Through task design in MATB-II, the precise time of starting could be determined, also the start of each stage in a voice (name, port, frequency) are point out, that's to say, starting time could be regard as an parameter that attribute to the transitions with gray color. Two texts N_c and C_c are named in four places that denote name stage and port stage of a voice are complete respectively. The interaction process will begin when participants hear a voice, and the activation place tied to "Get name" transitions will be added a token, that makes it enable, and fire if there are enough tokens in attention resources place and auditory resources place. It can be seen that the results of the transition are costing 43 tokens from attention resources place, maintaining tokens invariably in auditory resources place and enabling the transitions belong to cognition behavior. Two transitions “Not 504” and “504” were represented for the cognitive process name judging, which have a different condition and results. The “Not 504” transition would be enable with the voice played without NASA504, not need a following interaction after name judging, but for the “504” transitions, it will activation next interactions that listening to the port and frequency provided by the voice, also, visual attention resources are needed for the preparing of next input. The input of this task contains two parts, choose the port and input the frequency, all of them are completed by mouse, and are expressed in transitions “port input” and “Frequency and OK input”. Four special transitions were named with green text, connected by arcs drawn by dash line there to describe a more realistic interaction process. The “See” transition here means that we can see things at any time without costing attention resources but just cost one channel resource, this is the same to other channels that “Hear”, “Reflex” and “Keep posture”.

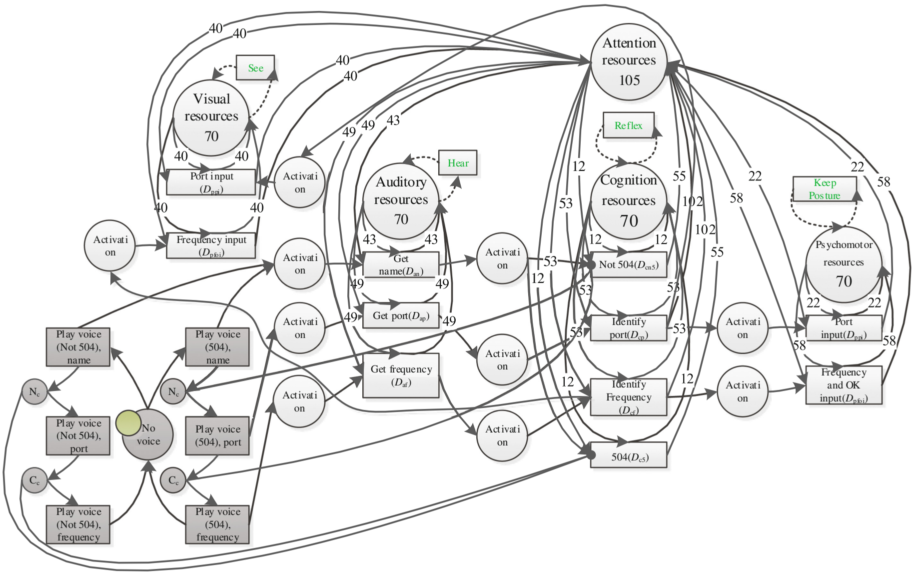


Fig. 2. Petri nets model of COMMUN task.

3.3 Behavior Times Calculation

In this section, we cover the duration parameters of transitions in Fig. 2. A timetable was proposed in Table 1 to give a detail description of different channel processes of COMMUN task. The records of operations of COMMUN tasks include several time nodes and numbers, they are moments of starting of a voice (T_{vs}), inputting the name and number of port (T_{nnp}), clicking the integer part and decimal part of frequency (T_{ip}, T_{fp}), inputting the OK button (T_{ok}); times of clicks of integer part and decimal part (C_{ip}, C_{dp}). Time parameters in Table 1 were combined with them to calculate the durations of behaviors. Equals are as follows:

Table 1. Detail describe of the interaction process of COMMUN task.

Stage of voice	Channels			
	Visual	Auditory	Cognition	Psychomotor
Playing "NASA 504" 1st time		Get name, duration is D_{an}		
Playing "NASA 504" 2nd time		(Get name, duration is D_{an})	Judge the name, duration is D_{c5}	
Playing port		Get port, duration is D_{ap}		
Playing frequency	Look and check, duration is D_{ppi} and D_{pfoi}	Get frequency, duration is D_{af}	1. Identify the name and number of the port, duration is D_{cp} 2. Identify the frequency, duration is D_{cf}	1. Input the name and number of the port, duration is D_{ppi} 2. Input the frequency of the port, and click the OK button, duration is D_{pfoi}
	Look and check, duration is D_{pfoi}		Identify the frequency, duration is D_{cf}	Input the frequency of the port, and click the OK button, duration is D_{pfoi}

The durations of auditory behaviors in Table 1 were carried out by analyzing the details of the voice file, such as D_{af} is equal to the playing time of frequency part. D_{c5} is regarded as a constant of 0.07s [14], D_{pfoi} is divided into D_{pfii}, D_{pfdi} and D_{poi} respectively, where are the durations of integer and decimal part of inputting of frequency, as well as inputting part of OK button. A behavior analysis was also conducted as the following equations:

$$5 * B_1 + C_{ip} * B_2 = D_{cf} + D_{pfii} = T_{ip} - T_{vs} - 2 * D_{an} - D_{ap} - D_{af} \tag{3}$$

$$2 * B_1 + B_2 = Dcp\ +\ D_{ppi} = T_{nnp} - T_{vs} - 2 * D_{an} - D_{ap} \tag{4}$$

B_1 means the cognitive behavior distinguishing one word, D_{cf} consists of five B_1 for that the frequency only needs to distinguish five numbers, such as 125.525, the first word is "1" all the while for all the voice. D_{cp} includes two B1 for that the port contains name (COM/NAV) and number (1/2). B_2 is the click time for one time, D_{pfii} includes C_{ip} times of clicks, while D_{ppi} only require one. D_{pfdi} can be calculated asas C_{dp} * B_2, and D_{poi} is expressed as $T_{ok} - T_{fp}$, Lastly, after carrying out D_{pfoi}, time parameters in petri nets model of COMMUN task are all endowed with.

4 Discussion

4.1 Workload and Performance Prediction of Multitask

Workload defined in this paper is following the definition that resources used will divide the total resources of each individual [20]. However, none of the previous studies explicitly indicate how individual's total resource is determined, and it is important for workload evaluation of different astronauts. In this paper, tokens in attention resources indicated as the total resources of individuals, initialized to 105. The steps to calculate workload are as follows:

1. Calculation of the durations of different behaviors for single task according to the method in Sect. 2.1.2, add them as a property of behavior transitions in multiply tasks petri nets model.
2. Record the reaction times of each single task in multiple task experiments.
3. Run the multiple tasks petri nets model, calculate the simulated reaction times of each single task.
4. Modifying the value of tokens in attention resources place constantly until the average deviation between simulated reaction times to actual of single tasks arrived minimum. At that time, the value can be seen as the total resources of the individual.
5. The instantaneous workload can be calculated as $R_{\mathrm{t}} - R_{\mathrm{r}}/R_{\mathrm{t}}$, where R_{t} which is the total tokens, and R_{r} is the remainder of the real-time tokens in attention resources place. The total workload was regarded as the time integral of instantaneous workload.

The performance index used, is reaction time and error rate, an error is recorded when reaction time is longer than system setting. Reaction time can be calculated by subtracting the appearing time of an event of system and the completion time of its corresponding psychomotor behavior. For example, in Fig. 2, the reaction time is obtained by subtracting the time transition "Play voice (504), name" firing and the time transition "Frequency and OK input" complete. When executing multitask, there may be more than one objects need to be focused on in each channel, this reflects in petri net model that more behavior transitions and corresponding activation places will be added in combined with each channel resources places. This adding will bring conflicts at many transitions are enabling when tokens in channel resources places and attention resources places are insufficient. At this time, some transitions will be preferentially firing and others must wait until resources is enough. Because that reaction time is calculated by adding visual/auditory, cognition and psychomotor behavior time, when a conflict appeared, one of the behaviors of the model will be put off, and will lead to much more time results in our model. For instance, when executing a triple tasks combination T_1, T_2, and T_3, attention resources of individual could be calculated by the above method that through the performance time of any two of them.

4.2 Individual's Characteristic

The proposed petri nets model varies from one individual to another due to the initial tokens in attention resources placement. Another two parameters with the same effect will be discussed as follows: First are the initial tokens in four channel placement, although the value in this paper are set to a constant 70, they are likely variables for \ different abilities of individuals. If a more accurate model is needed, these values should be adjusted in a modest variety. Another one is an appropriate duration for the corresponding resources of a behavior of individual. The time parameters in this paper were calculated by solving equations set up by a single task performance evaluation, also varies in different individuals. It can be seen that there is a maximum value of R_c/T_c, where R_c is resource consumed, T_c is the duration. Therefore, R_c/T_c can also be treated as a characteristic, and interpreted as the duration of transitions divided by the weight of its tied arc in model. In addition, R_c/T_c could also extend the ability of the petri nets model although it explains error as overtime now. As mentioned above, behaviors, which consume more resources, also need more time to complete. It will alter to behaviors needed fewer resources when under time pressure. This will lead to faults and could explain its appearance in complex tasks.

4.3 Further Work

Some further work on improving the model will be discussed in this section. First the aspect of the effort level, which has a share in workload, regarded as an index in some workload evaluation methods like NASA-TLX. The effort level can explain two parts, 1. spending more time, and 2. Additional focus, on tasks. In the task referred to in this paper, a higher effort level means more focus, it will lead to more interaction process without psychomotor. For example, a warning feedback task, when an operator sees a warning light, he needs to press a button. In our model, transitions fired of this process are mainly like “Get light”, “State judge (Warning)” and “Press button”, while, another operator with more efforts may spent more time on the warning light, and there will be more transitions like “State judge (Normal)” fired, this will lead to more workload. However, We can't evaluate how many times a transition “State judged (Normal)” will be fired, and it could be a modification after carrying out the workload result of the model if there are a prior setting of the effort level of the individual.

The execution sequence of transitions when meeting conflicts should also be concerned although they are difficult to appear absolute simultaneously in actuality. When interacting with multiple tasks, there will be conflicts both in inter channel and between different channels. In this paper, we adopt random selection when conflicts appeared in inter channel for the actual behavior (without any priority settings). While, for the conflicts between different channels, cognition behavior transitions are set as the highest priority, next is psychomotor for that some interaction object are both source of information input and object of psychomotor behaviors, and the last is perceptual behaviors. This also needs further studies.

Another issue here, is that the arousal level of individuals can bring some deviations when executing under multitasking situations. In the model proposed, the performance of multitasks should be decreased, since it may improve when the two tasks enhance arousal level exactly right. Although we can correct these deviations by resetting the behavior durations in multitasks. Lastly, if this model could be modified to be more precise, the method used to calibrate workload parameters like the weight of directed arcs will have lower importance. Researchers can select any other method to calibrate the resources needed for each behavior, and it will lead to the same workload result if these methods are considered with enough objectivity.

Acknowledgments. The National Science Foundation of China under Grant 51575037 supports this work.

References

1. Jo, S., Myung, R., Yoon, D.: Quantitative prediction of mental workload with the ACT-R cognitive architecture. Int. J. Ind. Ergon. **42**(4), 359–370 (2012)
2. Diallo, O., Rodrigues, J.J.P.C., Sene, M.: Chapter 11—performances evaluation and petri nets. In: Modeling and Simulation of Computer Networks and Systems, pp. 313–355. Elsevier, Amsterdam (2015)
3. Fecarotti, C., Andrews, J., Chen, R.: A petri net approach for performance modelling of polymer electrolyte membrane fuel cell systems. Int. J. Hydrog. Energy **41**(28), 12242–12260 (2016)
4. Li, L., Zeng, W., Hong, Z., Zhou, L.: Stochastic petri net-based performance evaluation of hybrid traffic for social networks system. Neurocomputing **204**, 3–7 (2016)
5. Müller, C., Rygielski, P., Spinner, S., Kounev, S.: Enabling fluid analysis for queueing petri nets via model transformation. Electron. Notes Theor. Comput. Sci. **327**, 71–91 (2016)
6. Kontogiannis, T., Leopoulos, V., Marmaras, N.: A comparison of accident analysis techniques for safety-critical man–machine systems. Int. J. Ind. Ergon. **25**(4), 327–347 (2000)
7. Wang, P., Fang, W., Guo, B.: A colored petri nets based workload evaluation model and its validation through Multi-Attribute Task Battery-II. Appl. Ergon. **60**, 260–274 (2017)
8. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
9. Wickens, C.D.: Multiple resources and performance prediction. Theor. Issues Ergon. Sci. **3** (2), 159–177 (2002)
10. DiDomenico, A., Nussbaum, M.A.: Interactive effects of mental and postural demands on subjective assessment of mental workload and postural stability. Saf. Sci. **43**(7), 485–495 (2005)
11. Brookhuis, K.A., van Driel, C.J., Hof, T., van Arem, B., Hoedemaeker, M.: Driving with a congestion assistant; mental workload and acceptance. Appl. Ergon. **40**(6), 1019–1025 (2009)
12. Al-Hakim, L., Maiping, T., Sevdalis, N.: Applying hierarchical task analysis to improving the patient positioning for direct lateral interbody fusion in spinal surgery. Appl. Ergon. **45** (4), 955–966 (2014)
13. Oyewole, S.A., Haight, J.M.: Determination of optimal paths to task goals using expert system based on GOMS model. Comput. Hum. Behav. **27**(2), 823–833 (2011)
14. Hamilton, W.I., Clarke, T.: Driver performance modelling and its practical application to railway safety. Appl. Ergon. **36**(6), 661–670 (2005)

15. Liang, S.F.M., Rau, C.L., Tsai, P.F., Chen., W.S.: Validation of a task demand measure for predicting mental workloads of physical therapists. Int. J. Ind. Ergon. **44**(5), 747–752 (2014)
16. Santiago-Espada, Y., Myer, R.R., Latorella, K.A., Comstock Jr. J.R.: The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide (2011)
17. Fairclough, Stephen.H., Venables, L., Tattersall, A.: The influence of task demand and learning on the psychophysiological response. Int. J. Psychophysiol. **56**(2), 171–184 (2005)
18. Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., Yamamura, T.: Physiological responses to workload change. A test/retest examination. Appl. Ergon. **40**(6), 987–996 (2009)
19. Phillips, C.A., Repperger, D.W., Kinsler, R., Bharwani, G., Kender, D.: A quantitative model of the human–machine interaction and multi-task performance: a strategy function and the unity model paradigm. Comput. Biol. Med. **37**(9), 1259–1271 (2007)
20. Boff, K.R., Kaufman, L., Thomas, J.P.: Handbook of perception and human performance, volume 2. Cognitive processes and performance. Harry G Armstrong Aerospace Medical Research Lab Wright-Patterson AFB OH (1994)

Systems Usability and Device Assessment

A Fuzzy Usability Assessment Methodology

Bruno Lemos[1] and Isabel L. Nunes[1,2(✉)]

[1] Faculdade de Ciencias e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
b.lemos@campus.fct.unl.pt, imn@fct.unl.pt
[2] UNIDEMI, Faculdade de Ciencias e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

Abstract. Usability is multidimensional and defining a model for the assessment of design solutions is challenging, since most features depend of user perceptions. This paper discusses the features of a new usability assessment model based on Fuzzy Logic and presents its application in a case study related with the development of a web portal based on a User Centered Design approach. The output of the proposed assessment model is the Usability Index.

Keywords: User Centered Design · Usability Index · Web portal

1 Introduction

Current fast-paced technological evolution leads to a significant increase in the number of software applications, web sites or portals that allow users to perform tasks and interact with information systems. One of the greatest challenges in software development is to ensure that users can interact with these systems with efficiency, effectiveness, and satisfaction, i.e. that such software presents high usability standards. Also, productivity and company economic success is affected by the lack of usability of its information systems.

In general terms usability expresses how easy an information system is to find, understand and use for its intended purpose without users becoming totally annoyed. The most commonly accepted definition for Usability is the one provided by ISO 9241-11 standard: "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [1].

Small and medium software development companies often lack the expertise to assess the usability of their products during the development process. In fact, this was the case of the company where the work here described was performed. The company didn't have a structured approach to perform usability assessments during the development of their products. These empirical findings agree with the conclusions of the study performed by Andreassona et al. [2]. These authors proposed the following set of recommendations to increase the focus on usability work in organizations:

- Increasing awareness of the benefits of usability work;
- Establishing usability work at management level;

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_38

- Supporting usability work among the personnel;
- Providing education in usability work;
- Mapping out the effects of deficient usability work;
- Following-up the results of the improved usability work process.

The existence of usability assessment methodologies can support the task of evaluating the quality of systems use considering specific users, context, and environment, therefore contributing to mitigate this issue.

The purpose of the work described in this paper was to develop a usability assessment model which produces as resulted a Usability Index (UI) indicating the quality of the evaluated software.

Developing a usability assessment methodology presents some challenges considering the number of attributes to consider and the uncertainty of their characterization. Fuzzy Set Theory, also referred as Fuzzy Logic [3] is a particularly adequate to deal with this type of assessment problems.

Fuzzy Set Theory (FST), which was formulated, in 1965, by Lotfi Zadeh, provides a mathematical framework for the systematic treatment of vagueness and imprecision. The subjective nature of human classification processes renders classical (Boolean/dichotomous) approaches almost useless to deal with human-centered systems. So FST facilitates the elicitation and encoding of knowledge affected by uncertainty and vagueness. It provides a representation mechanism that improves the flexibility for dealing with data associated to complex concepts. Using FST allows building more robust tools that perform better for a wider variety of conditions and users. From an encoding point of view, fuzzy sets support the representation of knowledge and its uncertainty as a unique entity. The resulting representation is very flexible and can be easily coupled with non-fuzzy forms of knowledge representation, and manipulated by a variety of evaluation methods [4].

A fuzzy set presents a boundary with a gradual contour, by contrast with classical sets, which present a discrete border. Formally, let U be the universe of discourse and u a generic element of U, a fuzzy subset A, defined in U, is one set of dual pairs:

$$A = \{(u, \mu_A(u)) | u \in U\} \quad (1)$$

where $\mu_A(u)$ is designated as membership function or membership grade u in A. The membership function associates to each element u, of U, a real number $\mu_A(u)$, in the interval [0, 1].

An important concept in FST is the one of linguistic variables [5–8]. A linguistic variable is a variable that admits as value words or sentences of a natural language, which can be represented as fuzzy sets. Linguistic variables (LV) serve as a means of approximate characterization of phenomena which are too complex or to ill-defined to be susceptible of description in precise terms allowing dealing directly with semantic concepts of imprecise nature, with a consistent mathematical formulation. LV are characterized by a set of terms to which are associated fuzzy sets that define their semantic. Taking in consideration the numerical efficiency related with computational systems LV terms can be assumed as discrete fuzzy sets. In the UI model two Linguistic Variables were considered: *agree* and *adequacy*.

Fuzzy logical operators are another important characteristic of FST. In fact, such operators play a key role in the aggregation of fuzzy sets and on the inference processes that use them as data. Application examples can be find in literature, e.g. [9, 10].

Typically, the processing of a fuzzy system is performed on 3 stages (Fig. 1). The first stage – fuzzification – transforms crisp input data (either numerical or linguistic) in fuzzy sets. The second stage – inference – where input data is combined by means of fuzzy logical operators and fuzzy IF-THEN rules producing fuzzy output results. The last stage – defuzzification – corresponds to the conversion of fuzzy results into crisp results (either numerical or linguistic).

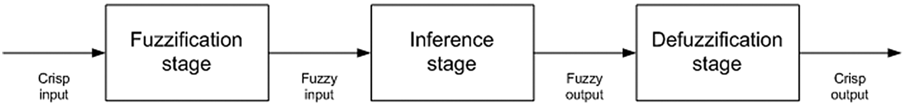


Fig. 1. Fuzzy logic system stages.

As mentioned before, the objective of this paper is to discuss the features of a new usability assessment model based on Fuzzy Logic and illustrate its application using a case study related with the development of a web portal following a User Centered Design approach. The output of the proposed assessment model is the Usability Index.

2 Usability Index Model

The UI model aims to support developers in the evaluation of web portals during the design of the solution and at the final acceptance stage. Thus, the UI model is coupled with the User Centered Design (UCD) framework [11], as depicted in Fig. 2.

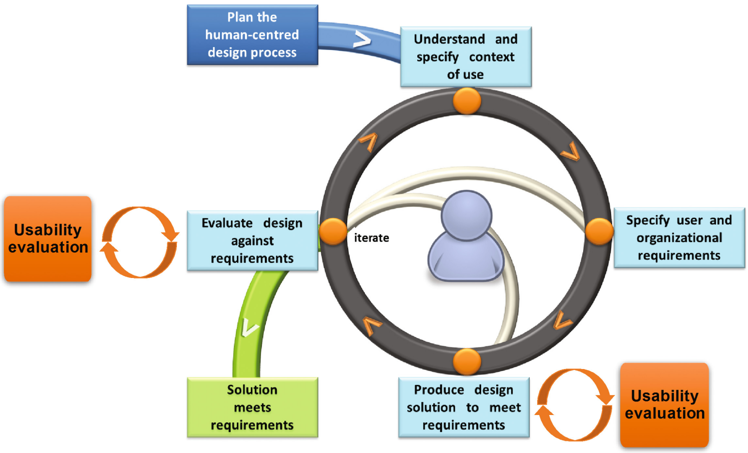


Fig. 2. Coupling of the UI model into the UCD framework

The UI model diagram is depicted in Fig. 3. The Usability Index computation involves collecting the perception of web portal users regarding the factors that affect usability (i.e., effectiveness, efficiency and satisfaction).

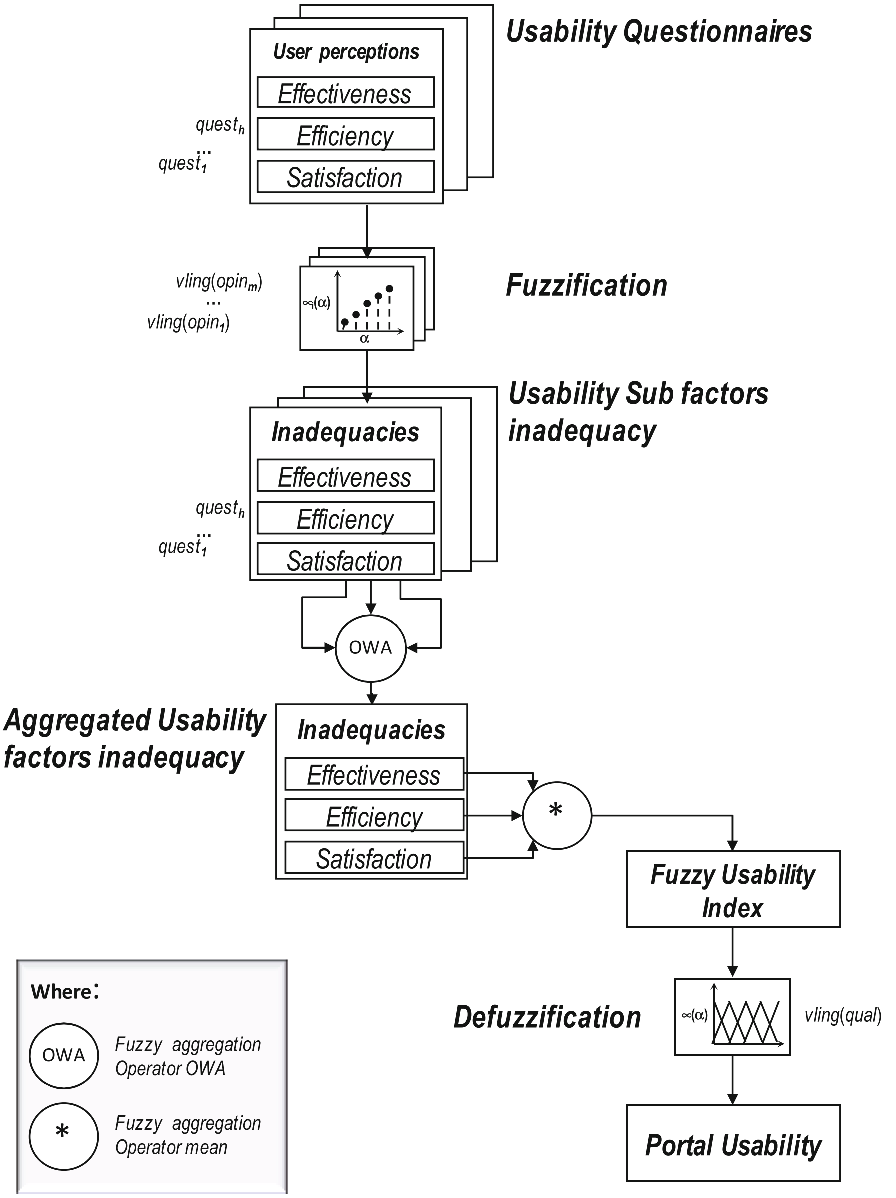


Fig. 3. Usability Index model.

Users perception is collected through the Usability Questionnaire that was developed for this purpose. The questionnaire has questions related with each of the 14 Sub factors listed on Table 1. For instance, for the Sub factor "Time spent on tasks" the question is "Is the time spent in the execution of the tasks adequate?"

Table 1. Model factors and sub factors.

Factors	Efficiency	Effectiveness	Satisfaction
Sub factors	(1) Time spent on tasks (2) Ease of navigation (3) Provision of information (4) Simplicity	(5) Adequate terminology (6) Number of errors (7) Understanding errors (8) Completion of tasks (9) System support material	(10) Visual attractiveness (11) Empathy with the system (12) User aesthetic overload (13) Frustration (14) Learning

The knowledge elicitation process for the definition of the mentioned Sub factors involved a group of 5 portals' users and 5 members of the development team. After a brainstorming session, the generated ideas where grouped using an Ishikawa diagram, in the 3 categories (Factors) that are considered in ISO 9241-11: Efficiency, Effectiveness and Satisfaction.

2.1 Linguistic Variables

The user's perception about the portal usability is collected using terms of the LV *agree*. This LV comprehends a set of 5 linguistic terms, to which correspond membership degrees uniformly distributed in the interval [0, 1]. Thus, the corresponding discrete fuzzy set is depicted in Fig. 4 and defined as:

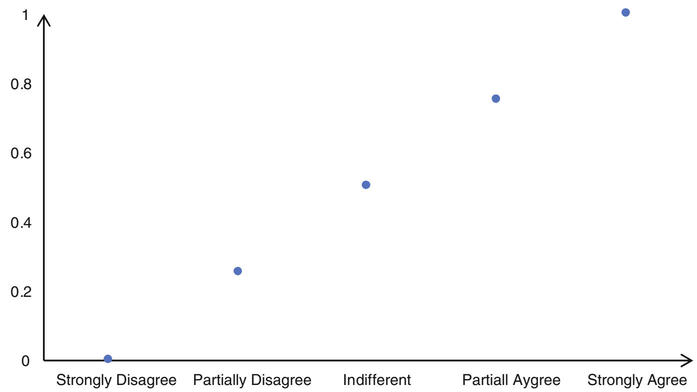


Fig. 4. Linguistic variable *agree*.

$$agree = \{0/\text{Strongly Disagree} + 0.25/\text{Partially Disagree} + 0.5/\text{Indifferent} + 0.75/\text{Partially Agree} + 1/\text{Strongly Agree}\} \quad (2)$$

The UI numerical evaluation results are converted to a natural language qualification using the continuous LV *adequacy*, through a process named defuzzification, which is presented on Fig. 5.

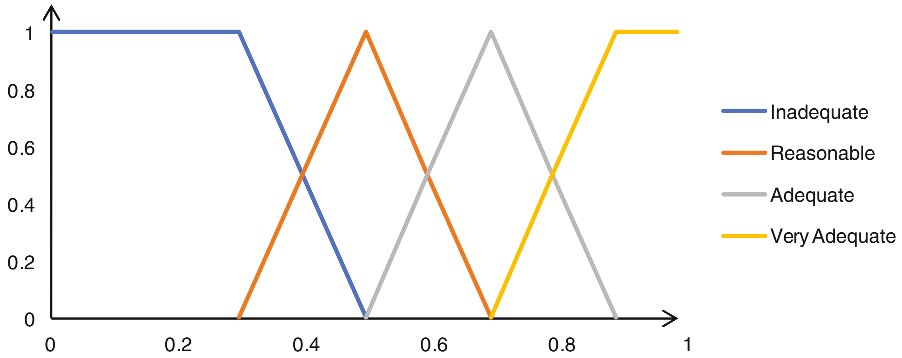


Fig. 5. Linguistic variable *adequacy*.

2.2 Fuzzy Operators

In the present model, Sub factors are aggregated using the ordered weighted averaging (OWA) operator, which was proposed by Yager [12]. This fuzzy operator has been increasingly used in a wide range of successful applications for aggregation of decision making problems [13].

Formally the OWA operator of dimension n is a mapping

$$F : R^n \rightarrow R \quad (3)$$

that has an associated collection of weights $w = (w_1, w_2, \ldots, w_n)$ in the unit interval and summing to one and with:

$$\mathrm{F}(a_1, \cdots, a_n) = \sum_{j=1}^{n} w_j b_j + \ldots + w_n b_n \quad (4)$$

where b_j is the jth largest of a_i.

The determination of OWA weights (w) is an important issue, since the result of the aggregation depends on the associated weights, hence in this case we have consider the following rational:

- usually in SME the usability evaluation is performed at maximum by 3 users;
- the UI is intended to consider all the users perceptions and homogenize their perceptions, i.e. intending to centralize the aggregation of the perceptions.

Therefore, the weights selected, translate this condition: a low weight, 0.25 is assigned to the extreme perceptions and 0.5 is assigned to the central value.

Therefore, the aggregation of Sub factors with the OWA operator, in an example where the portal usability evaluation is performed by 3 users (U_1, U_2 and U_2) is calculated using the following expression:

$$\text{Sub fator}\, X = \min[(U_1, U_2, U_3)] * 0.25 + \text{mean}[(U_1, U_2, U_3)] * 0.5 + \max[(U_1, U_2, U_3)] * 0.25. \quad (5)$$

The successive aggregation of the Sub factors to Factors, and then the aggregation of the three Factors leading to the calculation of the UI is performed with the fuzzy operator arithmetic mean:

$$\mu = \frac{1}{n}\sum_{i}^{n} \mu_i \quad (6)$$

where:
n – number of elements of a system;
μ_i – membership degree of element i.

3 Model Application: Results and Discussion

The Usability Questionnaire to collect user's perception was applied twice during the development of a web portal, according to the scheme presented on Fig. 2. Three users (U_1, U_2, U_3) participate in both usability evaluations. The data collected in the form of LV terms as well as the results are presented on Table 2, part (a) for the 1st evaluation and (b) for the 2nd evaluation.

The results of the 1st evaluation showed that the usability of the Portal was only *reasonable* (UI = 0.49).

To improve the portal the results were analyzed to identify the Sub factors associated with negative perception by the users (Sub factor results marked bold in Table 2). In collaboration with the development team several improvement measures were stabilised, and after their implementation, the results of the 2nd evaluation (UI = 0.63) correspond to a qualification of the portal usability as *adequate*, confirming a partially successful intervention. In fact, a comparison between the two evaluations of the usability Factors and the Usability Index (presented on Fig. 6), shows an improvement of 20% in Efficiency, 30% in Effectiveness and 40% in Satisfaction on the users' perception. The UI had an increment of 30% between the two evaluations.

The procedure of analyzing the usability evaluation results and implementing the measures identified to improve the Sub factors with a low evaluation can be repeated successively until the usability of the portal reaches a qualification of *very adequate*.

Table 2. Evaluation data. (a) 1st evaluation; (b) 2nd evaluation. Where SD - Strongly Disagree; PD - Partially Disagree; I - Indifferent; PA - Partially Agree; SA - Strongly Agree.

Factors	Sub factors	User perception			Results		
		U1	U2	U3	Sub factors	Factors	UI
(a) 1st evaluation							
Efficiency	1	PA	I	PD	0.50	0.53	0.49
	2	I	PD	PA	0.50		
	3	I	I	PD	0.44		
	4	I	PA	PA	0.69		
Effectiveness	5	PA	I	I	0.56	0.50	
	6	PA	I	PA	0.69		
	7	PA	PD	I	0.50		
	8	PA	PA	PA	0.75		
	9	SD	SD	SD	**0**		
Satisfaction	10	I	I	I	0.50	0.43	
	11	I	PD	PD	**0.31**		
	12	I	PD	PD	**0.31**		
	13	I	I	I	0.50		
	14	I	I	I	0.50		
Factors	Sub factors	User perception			Results		
		1	2	3	Sub factors	Factors	UI
(b) 2nd evaluation							
Efficiency	1	PA	SA	PD	0.69	0.63	0.63
	2	PD	PA	PA	0.63		
	3	I	PA	PD	0.50		
	4	PA	I	PA	0.69		
Effectiveness	5	I	PA	I	0.56	0.66	
	6	I	PA	PA	0.69		
	7	PD	PA	I	0.50		
	8	PA	PA	PA	0.75		
	9	SA	PA	PA	0.81		
Satisfaction	10	PA	I	I	0.56	0.60	
	11	I	PA	PA	0.69		
	12	I	I	PA	0.56		
	13	PA	PA	I	0.69		
	14	I	I	I	0.50		

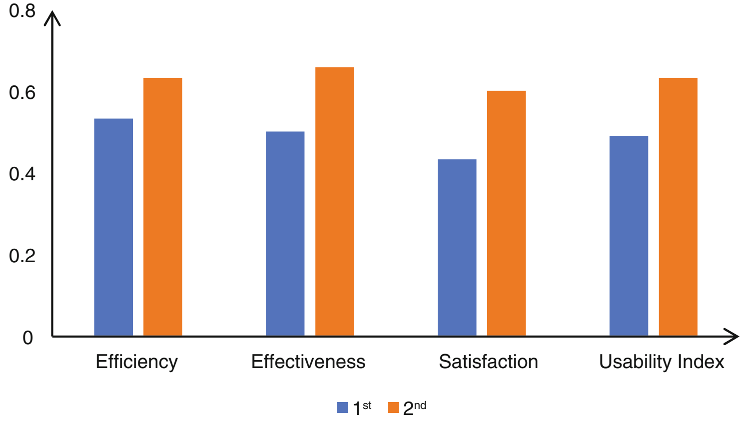


Fig. 6. Results of the 1st and 2nd web portal usability evaluation

4 Conclusions

The paper described the development of a Usability Index for practical use during web portal development, based on User Centered Design framework, targeting small and medium companies. The Usability Index model was developed using concepts of Fuzzy Logic. It aims to help the identification and evaluation of the factors that affect portal Usability, based on users' perception. This evaluation should be performed in two usability evaluation stages along the portal development process engaging potential users of the portal. The use of a FL approach presents some advantages over the classical methods commonly used, since it combines robust mathematical and logical methodologies with the capability of using LV to provide a natural language based means for collecting user perceptions and for presenting the evaluation result regarding the portal usability, using words instead of numbers.

A case study was presented to illustrate the application of the UI model. In this study improvement opportunities were identified through the application of this model and by analyzing the users' perceptions.

The study confirms that users should be a participating and active part throughout the development process, contributing with the identification of their needs and perceptions to the continuous improvement of the design of web portals.

References

1. ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) —Part 11: Guidance on Usability. International Organization for Standardization (1998)
2. Andreassona, R., Lindbloma, J., Thorvald, P.: Towards an increased degree of usability work in organizations. Procedia Manuf. **3**, 5739–5746 (2015)
3. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)

4. Nunes, I.L.: FAST ERGO_X—a tool for ergonomic auditing and work-related musculoskeletal disorders prevention. J. Prev. Assess. Rehabil. **34**(2), 133–148 (2009)
5. Zadeh, L.A.: Linguistic variables, approximate reasoning and dispositions. Med. Inform. **8** (3), 173–186 (1983)
6. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning—part I. Inf. Sci. **8**, 199–249 (1975)
7. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning—part II. Inf. Sci. **8**, 301–357 (1975)
8. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning—part III. Inf. Sci. **9**, 43–80 (1975)
9. Nunes, I.L.: Fuzzy systems to support industrial engineering management. J. Appl. Eng. Sci. **10**(3), 143–146 (2012)
10. Pinto, A., Nunes, I.L., Ribeiro, R.A.: Qualitative model for risk assessment in construction industry: a fuzzy logic approach. In: Camarinha-Matos, L.M., Pereira, P., Ribeiro, L. (eds.) Emerging Trends in Technological Innovation (First IFIP WG 5.5/SOCOLNET Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2010), 22–24 Feb, Costa de Caparica—Portugal, pp. 105–111. Springer, ISSN 1868-4238 (2010)
11. ISO 9241-210: Ergonomics of Human–System Interaction—Part 210: Human-Centred Design for Interactive Systems (2010)
12. Yager, R.R., Kacprzyk, J.: The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer, Norwell (1997)
13. Gholam, R., Amin, G.R., Emrouznejad, A.: Parametric aggregation in ordered weighted averaging. Int. J. Approx. Reason. **52**, 819–827 (2011)

Psychometric Analysis of Scales for Usability Evaluation of Pointing Devices

Denis A. Coelho[1(✉)], Miguel L. Lourenço[2], and Isabel L. Nunes[3]

[1] Human Technology Group, Department of Electromechanical Engineering and Centre for Mechanical and Aerospace Science and Technology, Universidade da Beira Interior, Covilhã, Portugal
dac@ubi.pt, denis.a.coelho@gmail.com

[2] Technical Scientific Unit of Engineering and Technology and Research Unit for Inland Development, School of Technology and Management, Instituto Politécnico da Guarda, Guarda, Portugal
mlopes@ipg.pt

[3] Faculty of Sciences and Technology, Universidade NOVA de Lisboa, Caparica, Portugal
imn@fct.unl.pt

Abstract. Usability evaluation involves measuring user performance (effectiveness and efficiency) and registering the user's perception of satisfaction related aspects. The paper reports on a method for assessing the satisfaction component, in computer handheld pointing devices, through a series of ratings related to three dimensions: comfort (discomfort), ease of use (per task performed) and effort (gripping, pointing, clicking). Cronbach's alpha is obtained as an estimate of the reliability of the assessment method made up of the multiple attribute evaluation of discomfort, ease of use and effort, based on Likert scales. Results revealed high values of internal consistency within each subjective dimension. Increased internal consistency would be obtained by removing the shoulder discomfort attribute from the discomfort evaluation questionnaire, and gripping and clicking attributes from the effort evaluation questionnaire. These divergences indicate that participants perceived these attributes as fundamentally different from the rest of the attributes within the evaluation dimensions.

Keywords: Ergonomics · Internal consistency · Ease of use · Effort · Discomfort · Human-systems integration

1 Introduction

ISO 9242-11 is an international standard and is part of "ISO 9241 – Ergonomic requirements for office work with visual display terminals (VDTs)". The ISO 9241-11 standard was finalized in 1998. ISO defines usability as: "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [1]. The three key expressions in this definition are: effectiveness – this key expression describes to what extent a goal or mission has been accomplished; efficiency – this describes, as opposed to effectiveness,

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_39

the degree of effort that was exerted in order to reach a goal or mission (the less effort is exerted, the better); satisfaction – this key expression refers to the degree of satisfaction and positive feelings that the product brings forth when used [2]. The two consensually accepted pillars of usability evaluation are hence objective evaluation done by measuring user performance (effectiveness and efficiency) and subjective evaluation taken from registering the user's perception of satisfaction and satisfaction related aspects when using the product or system. Psychometrics refers to psychological measurement and is concerned with the construction and validation of assessment instruments such as raters' judgements, as well as statistical research bearing on measurements.

This article reports a method for collection of ratings of the ease of use satisfaction component for computer pointing devices, through a series of assessments associated with comfort (discomfort), ease of use (task related) and effort (gripping, pointing, clicking). Cronbach's alpha is obtained as an estimate of the reliability of the instrument made up of the multiple item Likert scales of discomfort, ease of use and effort. The internal reliability measures are based on data collected from an empirical study by Lourenço [3] concerning the usability evaluation of a conventional PC mouse (Microsoft Optical 200), used in standardized graphical user interface tasks. The instrument design presented was developed for Computer Aided Design (CAD) mouse operations, while the attribute selection included in the instrument was developed further from information set in standard ISO 9241-9 [4]. Likert scales complete the instrument used to rate discomfort (laid out in 3 attributes), effort (spread out over 5 attributes) and ease of use (composed of 7 attributes).

2 Method

2.1 Research Approach Framing the Method

Framing of research in a methodological perspective supports the planning activity which guides the whole research process throughout its carrying out. The design of the research process defines a plan where the research strategies aim the collection of data and the respective methods of analysis, in order to provide answers to the research questions. The research process conducted in the experimental and design work conducive to the doctoral dissertation by Lourenço [3] was orientated towards the development and ergonomic analysis of new PC mice. This research process was framed within positivism, associated to exact sciences, adopting a structured research methodology, enabling its replicability.

The methodology of the research carried out by Lourenço [3], involved intertwined research methodologies and associated research questions employed in the development and ergonomic evaluation of PC mice. The taxonomy of research approaches proposed by Järvinen [5] was the basis for the classification of the research methodology implemented.

In theory testing studies, several methods may be used, such as laboratory experiments and survey research. In this kind of research approach (theory testing) the theory is acquired from literature review or developed and refined for the study. Here the research question is “do observations confirm or falsify the theory?”. In building a new artifact the research question underlying it is “is it possible to build a certain artifact?”, while in an artifact evaluation approach the researcher may pose the question “how efficient is this artifact?”. The experimental work that the methods presented in this paper support combined the approach for artifact building and artifact evaluation, comprised within a theory testing approach. The former approaches are the starting point for product development and comparative evaluation, with the effect of expanding existing theory as a result of their implementation.

In pursuing the eliciting of answers to the research questions, an elaborate research design was implemented making use of several research methods, including literature review, laboratory experiments, field experiments, product development and data collection instrument development. In implementing these research methods several research techniques were put to use, including the use of measuring devices as well as of assessment scales (featured in this paper), observation of the use of artifacts and prototyping.

2.2 PC Mouse Subjective Usability Evaluation Attributes

Standard ISO 9241 sets the ergonomic requirements for work with visual display equipment and defines usability as the extent to which a product may be used by specific users to attain specific goals with effectiveness, efficiency and satisfaction [1]. Usability is commonly seen as an attribute of the quality of a system or product that enables easy learning and easy use by the person or user. Usability evaluation may be based on a set of factors such as the performance of the operator (completing a task with ease while incurring in a reduced error rate) and satisfaction or ease of learning. The paper reports on a method that proposes assessing the satisfaction component of usability, in the domain of application of computer handheld pointing devices, through a series of ratings related to the dimensions: comfort (discomfort), ease of use (per task performed) and effort (gripping, pointing, clicking). Twenty participants evaluated the pointing devices according to three dimensions considered on the method for the subjective evaluation of usability. The evaluation was done after using the PC mice in the completion of standardized tasks representative of Computer Aided Design activity (pointing, dragging, steering, with different sized targets and using the three common mouse buttons alternatively). The design of the questionnaire considered the information set in standard ISO 9241-9 [4], which presents in its appendices examples of

Participant Identification: ______________________________________

Device use start time____:____

Device (select the PC mouse that you are assessing)

☐ ☐ ☐

Rate from 1 to 6 (concerning the PC mouse you selected above)

Discomfort of the hand/arm/shoulder system

Hand	Extreme discomfort	1	2	3	4	5	6	No discomfort at all
Forearm		1	2	3	4	5	6	
Shoulder		1	2	3	4	5	6	

Fig. 1. First part (testing details and discomfort rating of three attributes – hand, forearm and shoulder using 6 point Likert scales) of the three-part questionnaire used for the subjective evaluation of the usability of computer pointing devices.

means of evaluating usability. That information was taken as a point of departure and adapted to the study reported in this paper. The questionnaire is shown in three parts in Figs. 1, 2 and 3.

Ratings were provided in 6-point Likert scales filled by the participating subjects. Participants rated each attribute of the three dimensions using a scale from 1 to 6, where 1 stands for worst (high discomfort, very difficult, extreme effort) and 6 stands for best (no discomfort, very easy, no effort). The subjective evaluation of discomfort was parsed into three body regions: hand, forearm and shoulder. Ease of use was parsed into 7 ratings: one per each one of 6 tasks and one evaluation in terms of global ease of use, or lack thereof. The effort assessment is composed of 5 attributes. Subjects were given a three-part questionnaire addressing the several attributes of the three usability analysis subjective dimensions: discomfort (Fig. 1), ease of use (Fig. 2) and effort (Fig. 3).

Difficulty (ease) of use

Overall Difficulty / Ease in handling the device during task execution		Very difficult	1	2	3	4	5	6	Very easy
Pointing (pointing and clicking)			1	2	3	4	5	6	
Dragging	Dragging – **pointing finger** (left button)		1	2	3	4	5	6	
	Dragging - **middle button** (pressing scroll)		1	2	3	4	5	6	
	Dragging – **middle finger** (right button)		1	2	3	4	5	6	
Steering (clicking and dropping, moving and clicking at destination)			1	2	3	4	5	6	
Using Scroll (rotating scroll wheel in both directions)			1	2	3	4	5	6	

Fig. 2. Second part (difficulty (ease) of use rating of seven attributes – overall, pointing, dragging with left button, dragging with middle button, dragging with right button, steering and scrolling using 6 point Likert scales) of the three-part questionnaire used for the subjective evaluation of the usability of computer pointing devices.

Effort

Overall effort exerted to operate the device	Extreme effort	1	2	3	4	5	6	No effort
Effort in **gripping** the device		1	2	3	4	5	6	
Effort in **controlling** the movement of the device		1	2	3	4	5	6	
Effort in **clicking** on the buttons		1	2	3	4	5	6	
Effort in **pointing with accuracy**		1	2	3	4	5	6	

Fig. 3. Third part (effort rating of five attributes – overall, gripping using 6 point Likert scales) of the three-part questionnaire used for the subjective evaluation of the usability of computer pointing devices.

2.3 Experimental Tests of the Instrument

The subjective evaluation was carried out with 20 subjects (10 female and 10 male), while performing pointing, dragging and steering tasks adapted from previous studies [6, 7]. All subjects performed the tasks in the following order: pointing at large targets (pointing large), pointing at medium sized targets (pointing medium), pointing at small targets (pointing small), dragging with the left button (dragging left), dragging with the middle button (dragging middle), dragging with the right button (dragging right), and, finally, steering. A comparative overview of the graphical setup of the tasks is shown in Fig. 4. Tasks were implemented by a purpose built graphical software application. Sessions lasted between 10 and 12 min, depending on the duration of pauses the subject choose to undertake between tasks.

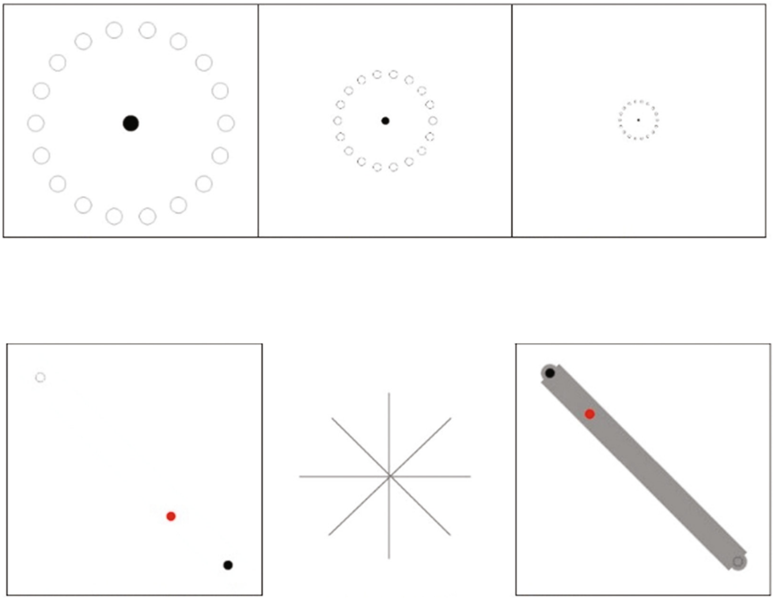


Fig. 4. Computer generated (purpose built software) graphical setup for the tasks in the individual sessions of the participants' usability testing of a standard PC mouse (clockwise from top left corner: pointing large, pointing medium, pointing small, steering, dragging and steering directions and dragging).

3 Psychometric Analysis of the Scales

Psychometrics refers to psychological measurement and is concerned with the construction and validation of assessment instruments such as raters' judgements, as well as statistical research bearing on measurements. Cronbach's alpha is obtained as an estimate of the reliability of the instrument made up of the multiple attributes of discomfort, ease of use and effort using a Likert scale. The internal reliability measures are based on data that was previously reported [8–10] concerning the comparative usability evaluation of a set of PC mice, used in standardized graphical user interface tasks (presented in the previous section). Statistical analysis was carried out using IBM SPSS. The results (Table 1) revealed high values of internal consistency among the three subjective dimensions of usability. Increased internal consistency would be obtained by removing the shoulder discomfort attribute from the discomfort assessment and gripping and clicking attributes form the effort assessment. These divergences indicate the participants perceived these attributes as fundamentally different to the rest of the attributes within the evaluation dimensions.

Table 1. Internal consistency analysis results of the scales presented in this paper - instruments developed to assess discomfort, effort and ease of use (Cronbach's alpha).

Usability dimension	Discomfort (Disc.) (3 attributes)	Ease of use (Eu) (7 attributes)	Effort (5 attributes)
Attributes (summarized)	Disc. of hand Disc. of the forearm Disc. of the shoulder	Global Eu Eu - pointing Eu - dragging w. left b. Eu - dragging w. middle b. Eu - dragging w. right b. Eu - steering Eu - scrolling	Global effort Effort to grip Effort to control Effort to click Effort to point acc.
Cronbach's alpha	Alpha = 0.768	Alpha = 0.903	Alpha = 0.877
Comments	Alpha = 0.784 without the attribute of shoulder discomfort	Removing any of the attributes from the Eu dimension would not lead to an increase in the value of Cronbach's alpha	Alpha = 0.885 without the effort to click attribute; Alpha = 0.893 without the effort to grip and effort to click attributes

4 Discussion

The goal in designing a reliable instrument is for scores on similar items to be related (internally consistent), but for each to contribute some unique information as well. The results revealed high values of internal consistency among the three subjective scales. Increased internal consistency would be obtained by removing the shoulder discomfort item from the discomfort scale and grasping and clicking items form the effort scale. These divergences indicate the participants perceived these items as fundamentally different to the rest of the attributes within the scales.

Acknowledgements. The authors would like to thank all the subjects that participated in the experimental trials leading to the psychometric analysis of the newly developed scales. The authors extend their acknowledgement to Rui A. Pitarma, Ph.D. for supportive action and to Noel Lopes, Ph.D. for collaboration in graphical software development. The current study was funded in part by Fundação para a Ciência e a Tecnologia project UID/EMS/00151/2013 C-MAST.

References

1. ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11 Guidance on Usability. International Organization for Standardization (ISO), Geneva, Switzerland (1996)
2. Roula, A.: Usability in IT-Systems of Financial Firms – A Distant Dream or Reality? Royal Institute of Technology, School of Computer Science and Communication, Stockholm, Sweden (2011). http://www.csc.kth.se/utbildning/kandidatexjobb/teknikmanagement/2011/roula_amira_K11102.pdf. Accessed 4 Mar 2017
3. Lourenço, L.M.L.: Desenvolvimento e Análise Ergonómica de Dispositivos Manuais Apontadores para Computador [Development and Ergonomic Analysis of Computer Handheld Pointing Devices]. Doctoral Dissertation. Industrial Engineering and Management. School of Engineering. Universidade da Beira Interior, Covilhã, Portugal (2016)
4. ISO 9241-9: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Non-keyboard Input Device Requirements International Organization for Standardization (ISO), Geneva, Switzerland (2000)
5. Järvinen, P.: On a variety of research output types. In: Proceedings of the IRIS, vol. 23, pp. 251–265 (2000)
6. Odell, D.L., Johnson, P.W.: Evaluation of a mouse designed to improve posture and comfort. In: Proceedings of the 2007 Work with Computing Systems Conference-International Ergonomics Association (2007)
7. Houwink, A., Hengel, K.M.O., Odell, D., Dennerlein, J.T.: Providing training enhances the biomechanical improvements of an alternative computer mouse design. Hum. Factors J. Hum. Factors Ergon. Soc. **51**(1), 46–55 (2009)
8. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Association of hand size with usability assessment parameters of a standard handheld computer pointing device. In: Proceedings of International Symposium on Occupational Safety and Hygiene—SHO 2016. Balkena and CRC Press, Boca Raton (2016)
9. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Ergonomic development of a computer pointing device – a departure from the conventional PC mouse towards CAD users. In: Proceedings of the 19th Triennial Congress of the International Ergonomics Association, Melbourne (2015)
10. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Horizontal and Vertical Handheld Pointing Devices Comparison for Increasing Human Systems Integration at the Design Stage. In: Nunes, I. (ed.) Advances in Human Factors and System Interactions. Advances in Intelligent Systems and Computing, vol. 497, pp. 15–24. Springer, Cham (2017)

Enhancement of User Experience by Hierarchical Task Analysis for Interaction System

Khadija Fyiaz, Shabana Tabassum(✉), and Ali Hasnain

Department of Computing and Information Technology,
University of Lahore Gujrat Campus, Gujrat, Pakistan
khadijafayyaz010@gmail.com,
shabana.tabassum@lbs.uol.edu.pk, alinain43@gmail.com

Abstract. This research Investigates how Task Analysis allows user to accomplish their task and obtain the information they require in an efficient and effective manner. Usability is the major factor for the use of any product or system and can be enhanced through many ways. Usability can be defined as necessity for web. If a website does not fulfill the needs of the user or it is difficult for a user to use, the user will leave the page. The research deals with the enhancement of user experience by task analysis. In this research issues related to usability of SME's (Small and Medium Enterprise) websites would be highlighted and enlighten how Task Analysis would be helpful to remove the shortcomings or flaws of existing websites. This research will use survey method and card sorting technique along task analysis for redesigning. The proposed design would be according to the user's need and easy for user to fulfill its goal.

Keywords: Usability · Task analysis · User experience · SME

1 Introduction

Usability has become a competitive factor for any system or product. User-Centered Design (UCD) is a technique to get a better usability from a system. This is also called as usability engineering or human centered design [1]. In this era when usability has become major component for the use of any IT related product or any system or web, the need of enhancement of usability for any product, system or web. Usability of any product or system can be increased through many ways. First thing to do is check the existing or running system, product or web by testing its usability. Usability inspection can be done by using Task Analysis technique (TAT). To resolve problem for any web or system TAT would be helpful to gain the user's desired system or web with high usability.

Today, the expectation with real utilization feature from the software used in SMEs is less than technical and functional feature. But, usability is taken as important as technical and functional features [2]. In this era when the need of website has highly increased, with the demand rapid growth of Small and medium Enterprises in Pakistan

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_40

the need of websites has increased. With the development of websites, it is also important to increase the usability of websites and to facilitate user as per its needs.

The focus of this research is to enhance the usability of SME's websites. Different websites related to Small and Medium Enterprise would be evaluated through Task analysis. Survey technique would be used for gathering data from user. Usability can be enhanced by using different methods. With the help of usability techniques and user involvement, will enhance the user experience. And also gives the user desired result and ease to achieve its goals.

2 Literature Review

Usability can be defined as: interaction between the user and the product or with the system. Usability is measurement of a system's capacity to achieve the desired result. However, mankind systems that are easy to use like, a tool or device that also referred as usability. The degree of a system to get desired results with efficiency and satisfaction used by selected persons in a precise circumstance of use [3]. In a research a researcher has adopted ISO's wide definition of usability which is consist of above mentioned three distinct features (Effectiveness, efficiency and satisfaction [4, 5].

2.1 Usability and Its Importance

On the web, usability is the key to survive. Usability can be used as necessity for web. If a website does not fulfill the needs of user or it is difficult for a user to use, the user will leave the page [6]. Same like this if your homepage doesn't deliver the purpose of your website then it would be useless to develop it.

Usability plays a vital role for any web. According to Jakob Nielsen, if the user task or requirement is being fulfilled and the user is spending enough time on website to achieve its goal and getting desired result it would be by providing maximum usability [7].

The result of a research shown that usability is vital to develop favorable perception of the user. Globally, the researcher has discovered four usability features from five calculated to be sufficiently connected to the reliability of the vendor. Hence, the researcher come up with a result that usability of the system has an influence on building trust [8].

2.2 SMEs Websites of Study Area

The need of fully efficient and effective website is also high in the electronic era. A user always search for an easier and effective way to find its desired information within minimum time. "The role of small and medium enterprises (SME), especially small businesses, has significantly increased in the developing of interactive applications. Given the relevance of these companies in the market, the adoption of usability and user experience (UX) practices in these organizations is fundamental for the improvement of interactive systems" [9].

Following are the few SME websites of the study area. The study area of my research is Gujranwala, following are the websites which are developed and in running process. There are also many of the SMEs who don’t have their running websites. The main focus in this research is the SME websites of Gujranwala region. As Gujranwala is an industrial area so enhancement of SME websites is needed here. In the following list we can see that among 11 SME only 5 enterprises have running websites. And the running websites have a lot of problems due to which a user is not getting the desired information.

2.3 SMEs Websites and Usability

Usability is often mentioned as an important determinant of harmonious and productive online system. Globally, the researcher has discovered four usability features from five calculated to be sufficiently connected to the reliability of the vendor [10].

After years of work on this project of research and development, it is still hard to find great quality, mash ups on this system. In contrast, we can say that still there are no stable development exercise and devices [11]. There are many websites and the trend of developing websites is getting higher but the focus of high quality or higher usable websites is still missing.

GOMS Model. It is good to examine the information to perform a work in conditions of goals, operators, methods and selection principles. A technique to explain the information of method that a user should have to control a product. This technique is explained by Card, Moron and Newell [12, 13].

Goals - using this system what kind of aims can be completed?
Operators - what kind of work can be done?
Methods - what order of persons can be used to achieve the goal?
Selection Rules - what technique can be used to achieve the goal?

2.4 Task Analysis

Task analysis means learning about users’ goals and users’ ways of working. Figuring out what more specific tasks users must do to meet those goals and what steps they must take to accomplish those tasks also meant to Task Analysis. Along with user and task analysis, a third analysis is often used: understanding users’ environments (physical, social, cultural, and technological environments). TA methods can be used to eliminate the preconditions that give rise to errors before they occur.

They can be used as an aid in the design stage of a new system, or the modification of an existing system. They can also be used as part of an audit of an existing system. The technique should be used during the analysis phase of design to ensure proper description of user activities. It can be used to analyses interactions with an existing system or as a means to structure discussions about a hypothetical product [14].

Task Analysis and Websites. By using testing technique “Task Analysis” any system or website can be evaluated. Task Analysis can help to check following things:

- Goal: Things they want to do.
- Task: What things they use.
- Action: What they must know.

To check usability of website Task Analysis technique applied on different websites following are the few examples of Task Analysis and usability issues:

Example: Master Sanitary Fittings Website.

Hierarchical Task Analysis. In the below figure, it represent the Hierarchical Task analysis of Home page of Master Sanitary Fittings website:

Textual Task Analysis. Following is the Textual Task Analysis of Master Sanitary Fittings' website's Home Page:

0. Master Sanitary Fittings
1. Home
2. About Us
3. Products
4. Media Center
5. Customer Care
6. Contact Us
7. Wish list

Sub-task Analysis. In the sub-Task Analysis a major task is divided in to sub-tasks for example: there is a task to search oasis overhead (Product of Master Sanitary). How a task would be divided in to sub-task, following are the hierarchical and textual division of a task (Fig. 1).

Hierarchical Task Analysis.

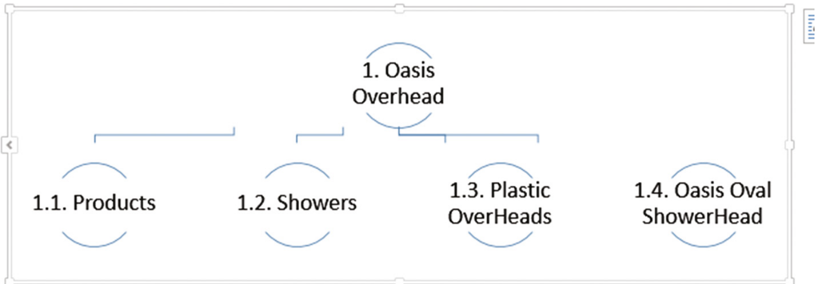


Fig. 1. Simple Task performed on SME's website

Textual Task Analysis.

- Oasis Overhead

1.1 Products
1.2 Showers
1.3 Plastic Overheads
1.4 Oasis Oval Shower Head

Now above mentioned Fig. 1 is the simple Task performed on SME's website. In above mentioned hierarchy the goal of a user is to search Oasis overhead, next is the task user performed to achieve its goal and the actions user know before performing the task. Now following are the screenshots of website on which a task is performed (Fig. 2):

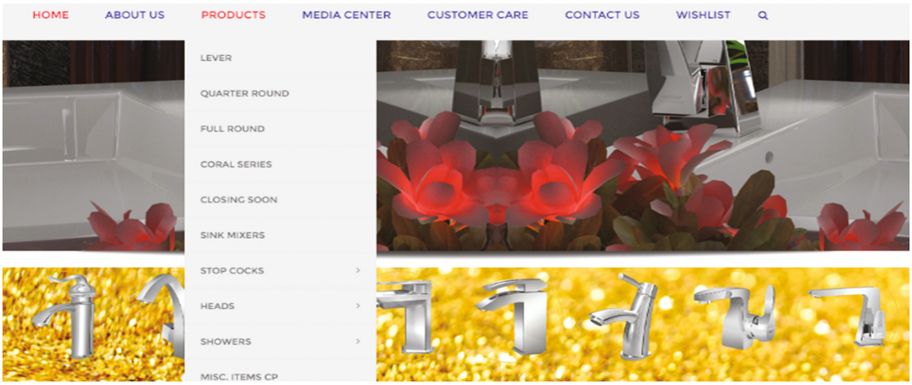


Fig. 2. Products list

On the website user need to search for a specific product of a sanitary. It would go for products list on the homepage and will go for the sub-tab of products which is of shower.

Fig. 3. Subtab of Products

In Fig. 3 users has find its desire product is the sub-tab of products, which is shower. This is an example of a task performed on the website.

As above are the simple task performed on the website same like that to identify the problems in the existing system. Following are the few problems in the existing system.

2.5 Card Sorting

The term card Sorting applies to a wide variety of activities involving the grouping and/or naming of objects or concepts. These may be represented on physical cards; virtual cards on computer screens; or photos in either physical or computer form [8].

Following are the methods of card sorting:

Open Card Sorting. Participants are given cards with no pre-established groupings. They are asked to sort cards into groups that they feel are appropriate and then describe each group [15].

Closed Card Sorting. Participants are given cards showing site content with an established initial set of primary groups. Participants are asked to place cards into these pre-established primary groups. Closed card sorting is useful when adding new content to an existing structure, or for gaining additional feedback after an open card sort [15].

There are few advantages of card sorting which are above mentioned:

Simple – Card sorts are very easy to implement by organizers and participants as well.
Cheap – Cost of a stack of 3 × 5 index cards, sticky notes, a pen or printing labels, and your time.
Quick to execute – Can perform several sorts in short time that gives significant amount of data.
Established – The technique has been used for over 10 years, by many designers.
Involves users – Information structure suggested by a card sort involves real users and hence it should be easier to use.

2.6 Problem Statement

With the needs of any common searcher or a user of website, a website also fulfils the needs of its organization. In this era where the users are being facilitate by electronic information, users are also facing many problems. The major issue of interfaces or of websites is "Usability" missing of cultural essences [16–22]. Users are encountering usability issues on the SME's interfaces and websites, also unable to find their desired information on the websites. This is the major problem of having worst "User Experience".

2.7 Research Questions

This research investigated how task analysis allows user to accomplish their task and obtain the information they require in an efficient and effective manner. Therefore, the research question is as follow:

- How can Task analysis enhance the user experience?
- The main research question has been split into the following sub-questions:

- How does Task Analysis influence user?
- How can Task Analysis identity areas for design improvements?
- How can these areas be used to design?

In order to investigate the abovementioned research question, the study aimed to achieve the following specific aims and objectives.

Aims. The research addressed the above research questions by investigating:

- How to understand User-Centered design techniques
- How these techniques can be leveraged for user interface design
- What impact these techniques have on usability adoption

Objective. The objectives of the research are:

- To understand how Task Analysis can increase the effectiveness and efficiency of SME's interface design.
- To develop a process model of user interface design.
- To develop and evaluate experimental user interface design.

3 Methods

3.1 Participants

In order to gather the information from different users and people from different groups related to Small and Medium enterprise domain the survey method would be used through the method of questioner, the information would be gather from different people of different levels. Data would be also collected from specific group of people who are directly involved in SME's domain.

3.2 Methodology

Task Analysis technique would be used to check the usability of existing Small and Medium Enterprise's websites. Through the Task Analysis the websites would be evaluated. The comparison of different websites in Pakistan would be done.

In next phase card sorting technique, would be used to have a view of user that which kind of website or platform a user needs to have.

In the last phase, the modification in the SME's website would be done according to the need of user and demand of the era. Through task analysis the redesigned website would be evaluated.

3.3 Pilot Study

This section details the pilot study that was conducted for research project of title "Enhancement of user experience by Task Analysis" In the user experience of Small and medium enterprise website users. While using the SME websites what type of problems a user faced and how much it is difficult to interact with the interface? SME has vital importance in Pakistan. In Pakistan SME is not the minor sharer till yet but the reality is that the Pakistan's whole economy is highly dependent on the pace and productivity of SMEs. With the rapid growth in SMEs the need of E-business has also increased. The best way of E-business is to have a website or a platform where to meet the needs of user. To provide maximum productivity and to enhance the user experience.

The aims of the pilot study were to:

- To investigate the usage of SME websites
- For which purpose respondents use SME websites
- To investigate on which a SME website should focus on?
- Find out how easily a user gets its desired information?

Method

Participants. Twenty SME related people have participated in pilot study as volunteers who were directly linked with any small and medium enterprise. They were randomly selected from different age group. They had varying experience and knowledge.

Questionnaire. A questionnaire was designed to get the view from the people who are directly linked with small and medium enterprises. The questionnaire was distributed among the 20 people of a small and medium enterprise in Gujranwala. The SME in which questionnaire was distributed is Master Sanitary Fittings Ltd. In the questionnaire few basic questions were asked related to the usage of SME websites and for which purpose people mostly use the SME websites. And on which aspects or features do a SME websites should focus on?

Results

Participant's Response against the Information of SME. There were 20 respondents of questionnaire and all the respondents have knowledge about Small and medium Enterprise as they are directly linked with SME.

Usage of SME websites. Figure below shown how many people use the SME websites. There were still few people who are directly linked with small and medium enterprise but don't use SME website (Fig. 4).

Which Purpose Respondents Use SME Websites? During the survey, through questionnaire it was asked for which purpose does the peoples use SME websites? Below Fig. 5 has the results on the purpose of the SME website use.

On which a SME Website Should Focus on? Through a survey done in a small and medium enterprise, respondents were asked to tell us on which there enterprise website should focus on? Following is the response over the questionnaire (Fig. 6):

Find the Desired Information on the Web? Through the questionnaire it was also asked that the user of SME websites get all the desired information on the web? Different response was received from the respondents most of the users answer was no and there were very few people who have found their desired information on the web. In the following chart it is clearly shown how much percent of respondents get the desired information on the SME website (Fig. 7).

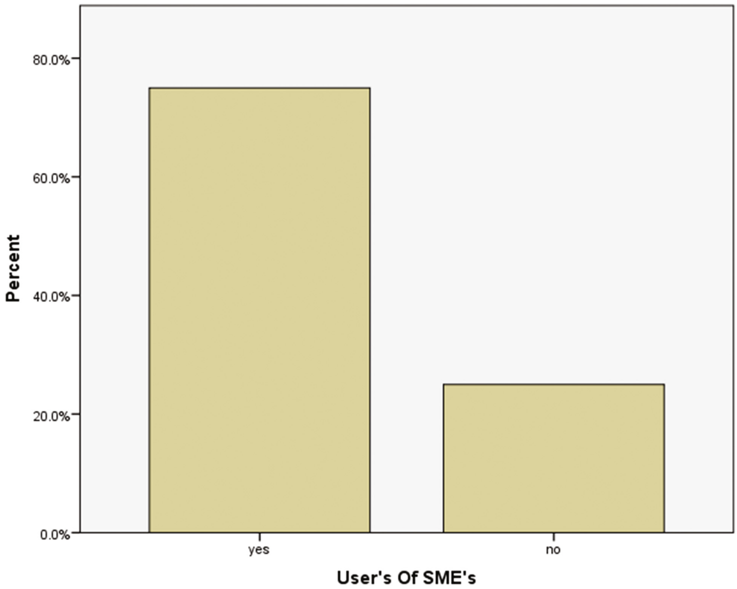


Fig. 4. How many people use the SME websites

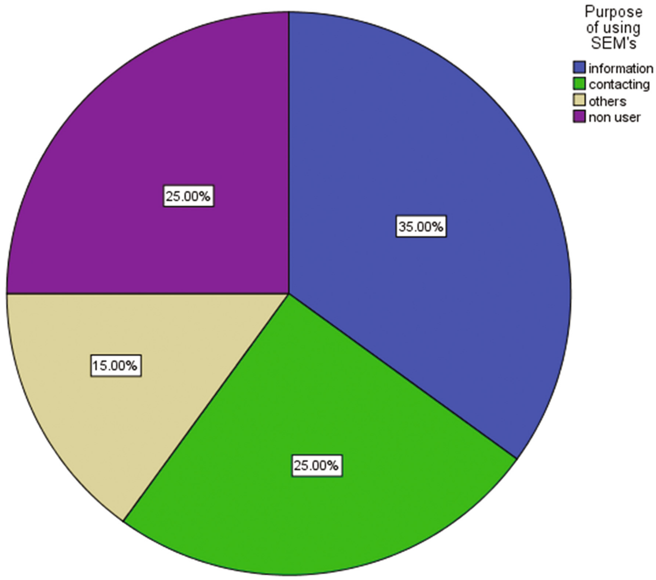


Fig. 5. Purpose of SME website use

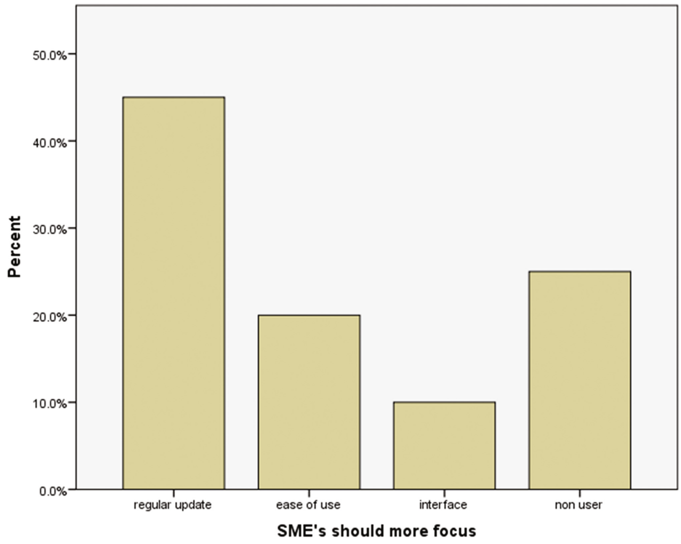


Fig. 6. Which their enterprise website should focus on

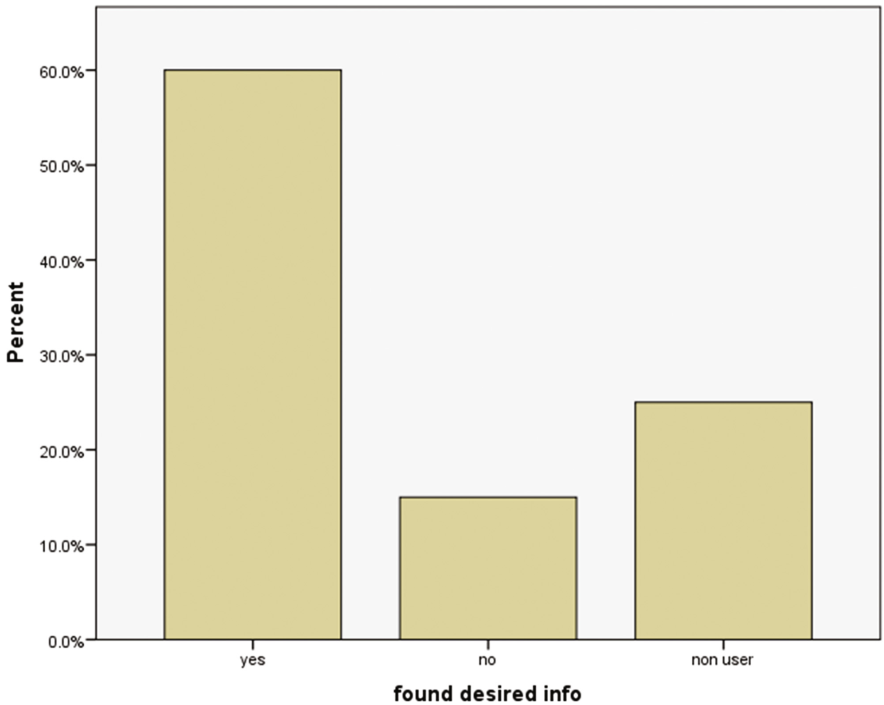


Fig. 7. How much get the desired information on the SME website

4 Expected Results

By evaluating existing Small and medium enterprise website through Task Analysis it would be helpful to overcome the shortcomings or flaws of existing websites. By using survey method, information will be gathered that what is the need of user. It would be helpful for redesigning a website for a Small and Medium Enterprise, with the help of Task analysis redesigned website would be reevaluate to enhance the usability of the website.

5 Significance

With the deployment of proposed method a new design would be developed to facilitate user who are directly or indirectly involved with Small and Medium Enterprise and also to enhance the usability. The proposed design would be according to the need of user so it can be easy for user to fulfill its goal by performing same tasks which was not fulfilling the need of user in existing website.

References

1. Speicher, M.: What is Usability? A Characterization Based on ISO 9241-11 and ISO/IEC 25010 (2015). arXiv preprint arXiv:1502.06792
2. Scheiber, F., et al.: Software Usability in Small and Medium Sized Enterprises in Germany: An Empirical Study. In: Software for people, pp. 39–52. Springer, Berlin (2012)
3. Lapin, K.: Task analysis. http://web.vu.lt/mif/k.lapin/files/2016/03/5_task-analysis_mockups-2016.pdf
4. Brooke, J.: Usability, change, adaptable systems and community computing. In: Proceedings Fourth International Conference on Human–Computer Interaction, vol. 2 (1991)
5. Frøkjær, E., Morten H., Kasper, H.: Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM (2000)
6. Lew, P., Luis O., Li, Z.: Quality, quality in use, actual usability and user experience as key drivers for web application evaluation. In: International Conference on Web Engineering. Springer, Berlin (2010)
7. Nielsen, J.: Usability 101: Introduction to Usability. (2003)
8. Speicher, M., Andreas, B., Martin, G.: Inuit: the interface usability instrument. In: International Conference of Design, User Experience, and Usability. Springer International Publishing (2015)
9. de Lima Salgado, A., et al.: Usability and UX practices in small enterprises: lessons from a survey of the Brazilian context. In: Proceedings of the 34th ACM International Conference on the Design of Communication. ACM (2016)
10. Aubert, B.A.: The impact of interface quality on trust in web retailers. Cahier du GReSI no 1, 05 (2001)
11. Orehovački, T., Cinzia C., Maristella, M.: Identifying relevant dimensions for the quality of web mashups: an empirical study. In: 18th International Conference on Human–Computer Interaction (HCII), Toronto. Springer, Heidelberg (2016)

12. John, B.E., Kieras, D.E.: Using GOMS for user interface design and evaluation: which technique? ACM Trans. Comput. Hum. Interact. (TOCHI) **3**(4), 287–319 (1996)
13. David E.K.: GOMS Models- Simplified Cognitive Architectures (2016)
14. Embrey, D.: Task Analysis Techniques. Human Reliability Associates Ltd 1 (2000)
15. Hudson, W.: Card sorting. In: The Encyclopedia of Human–Computer Interaction, 2nd edn. (2013)
16. Duncker, E., Sheikh, J.A., Fields, B.: From global terminology to local terminology: a review on cross-cultural interface design solutions. In: HCI (15). Lecture Notes in Computer Science, vol. 8023, pp. 197–207. Springer (2013)
17. Sheikh, J.A., Fields, B., Duncker, E.: The cultural integration of knowledge management into interactive design. In: HCI (14). Lecture Notes in Computer Science, vol. 6771, pp. 48–57. Springer (2011)
18. Sheikh, J.A., Fields, B., Duncker, E.: Cultural representation by card sorting. In: Ergonomics for All: Celebrating PPCOE's 20 Years of Excellence, pp. 215–220. Selected Papers of the Pan-Pacific Conference on Ergonomics, 7–10 November 2010, Kaohsiung, Taiwan. CRC Press (2011)
19. Sheikh, J.A., Fields, B., Duncker, E.: Multi-culture interaction design. In: Advances in Cross-Cultural Decision Making, pp. 406–415. CRC Press (2010)
20. Sheikh, J.A., Fields, B., Duncker, E.: Cultural based e-Health information system. In: Presentation at the Health Libraries Group Conference 2010, 19–20 July. CILIP, Salford Quays, UK (2010)
21. Sheikh, J.A., Fields, B., Duncker, E.: Cultural representation for interactive information system. In: Proceedings of the 2009 International Conference on the Current Trends in Information Technology, Dubai (2009)
22. Sheikh, J.A., Fields, B., Duncker, E.: Cultural representation for multi-culture interaction design. In: HCI (13). Lecture Notes in Computer Science, vol. 5623, pp. 99–107. Springer (2009)

S-EMG of Forearm Muscles Activity in Conventional PC Mouse Use

Miguel L. Lourenço[1] and Denis A. Coelho[2(✉)]

[1] Technical Scientific Unit of Engineering and Technology and Research Unit for Inland Development, Instituto Politécnico da Guarda, Guarda, Portugal
mlopes@ipg.pt

[2] Human Technology Group, Department of Electromechanical Engineering and Centre for Mechanical and Aerospace Science and Technology, Universidade da Beira Interior, Covilhã, Portugal
dac@ubi.pt, denis.a.coelho@gmail.com

Abstract. Prevalence of static muscular activity is believed to represent a higher risk for musculoskeletal health than dynamic muscular activity. PC mouse usage is a kind of activity where static muscle activation levels, while low in general, can indicate potential for musculoskeletal disorders. The study reported in this paper proposes a forearm four muscle approach to evaluation of muscular activity during mouse use, describing the rationale for selection of the muscles monitored. Additionally, signal processing is also characterized in the paper. The paper also presents results of surface electromyography assessment of 20 participants interacting with a conventional type of PC mouse. Analysis of the electromyographic signals was performed via a procedure of Amplitude Probability Distribution Function (APDF) calculation for the 90th, 50th and 10th percentiles. Average APDF90 values of muscular activity as a percentage of Maximum Voluntary Contraction found were between 7% and 21%.

Keywords: Surface electromyography · PC mouse · Muscle selection

1 Introduction

Prevalence of static muscular activity of muscles activated in the use of the conventional PC mouse is believed to represent a higher risk for the musculoskeletal health of the user than dynamic muscular activity. Static muscular activities concern physical effort caused by a posture that is maintained over a long period of time. This type of effort promotes the increase in the load of the contracted muscles, which contributes to rapidly achieving muscular fatigue. The absence of dynamic activation of a contracted muscle hinders the blood flow that is needed to transport nutrients to the muscle and to remove residual muscular metabolic products [1]. This notwithstanding, higher levels of activation are potentially more harmful than lower values of activation, due to more fatigue being caused. According to Sjøgaard and Jensen [2], activities that imply repeated movements are typically accepted as implying dynamic muscular efforts without taking into account the eventual occurrence of repeated muscular static efforts. The same study refers that the risks related to static efforts probably arise from the fact

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_41

that muscular contraction is continuously maintained, listing a set of activities where static ‘low level’ and long duration contractions occur. Among these activities the authors name computer work, especially the work with computer aided designing (CAD) [2]. Therefore, Sjøgaard and Jensen [2] define the ‘low level’ static effort as a condition associated to muscular work in which the muscle is activated at such a low order level that it allows the work to be maintained over a long period of time. The use of computer handheld pointing devices that causes the pronation of the forearm, such as the conventional PC mouse, may reveal the occurrence of this type of static muscular activity, which may eventually indicate potential risk factors related to musculoskeletal disorders. The computer mouse’s shape (its geometry) can be assessed through the use of surface electromyography (S-EMG) by measuring the activity levels of certain muscles during the use of the device [3]. Wrist extension is associated to musculoskeletal disorders of the hand, namely through compression of the median nerve that crosses the carpal tunnel. Moreover, ulnar deviation is associated with injury of the ulnar nerve [3]. It is also known that the patterns of muscle activation of the fingers extensor muscles (needed to lift and maintain the fingers above the buttons) combined with wrist extension postures may contribute to the occurrence of pain in the upper extremities during intensive use of the computer mouse [4–7]. Thus, considering the aim of this study which consists in characterizing the forearm muscular activity involved in the use of a conventional computer mouse during the realization of standard graphic tasks [8], the monitoring of the activity of four muscles of the forearm of 20 participants has been performed [9]. The monitoring of muscle activity was implemented through S-EMG and the referred muscles were judiciously selected.

2 Methods

2.1 Muscle Selection

Wrist extension can be monitored through electromyographic activity of the *Extensor Digitorum Communis* (EDC or just ED), *Extensor Carpi Ulnaris* (ECU) and *Extensor Carpi Radialis* (ECR) muscles, whereas the extension of the medial fingersoccurs through EDC muscular activity, with slight or no activity of the other two said muscles [13]. Furthermore, registered electromyographic activity of the ECR and ECU muscles enables monitoring radial and ulnar deviations of the wrist [13]. Several studies related to the problem of the use of the computer mouse and involving monitoring of muscle activity through S-EMG have been published. In this context, some of these studies stand out because they commonly include the monitoring of the electromyographic activity of the EDC, ECU and ECR forearm muscles [4, 10–12]. In the current study, muscles whose activity was monitored assume special relevance in the generation of postures considered inadequate and potentially harmful and that are likely to be adopted during the use of the computer mouse. These postures involve wrist extension and, or, extension of the medial fingers, as well as radial deviation and ulnar deviation of the wrist or hand [14]. Additionally, the selection process of the muscles whose activity was monitored also received the useful and personal counselling given by Professor David M. Rempel [15]. After weighing all the information from literature and expert advice, the monitoring of the ECR, ECU, ED and the *Abductor Pollicis Longus* (APL) muscles was performed. Despite the fact that the *Abductor Pollicis Brevis* (APB) muscle had been initially indicated (applied on thenar protuberance of the hand); knowing that this muscle activity is related to thumb motion, the APB muscle was replaced by the APL muscle, since it was found after a few pilot trials that the sensor applied on the hand was being dragged on the mouse pad and on the surface of the mouse. Replacement of the APB muscle by the APL muscle leads to undervaluing monitoring of the pinch grasp posture the hand might eventually perform when grasping the mouse. However, abduction of the thumb is also recorded by either of these muscles.

2.2 Electrode Placement

The surface electrodes used to collect the electromyographic signals from ECR, ECU and ED muscles were placed following Perotto's guidelines [16], as shown in Figs. 1, 2 and 3 respectively. Despite the fact that the guidelines refer to insertional electromyography [16], the area marked with x in Figs. 1, 2 and 3 does not present any other muscle in an outer layer. This condition allows the use of the referred spots (precisely defined) not only in insertional electromyography but also in surface electromyography. For the ECR muscle (Fig. 1), the x spot distance might be slightly increased to about 5 cm [13]. In Figs. 2 and 3, ulna assumes the reference length to marking the x spot.

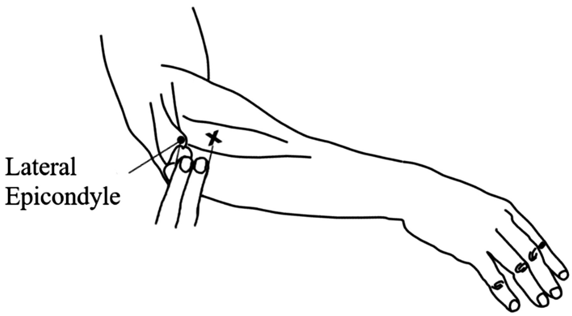


Fig. 1. *Extensor Carpi Radialis* (ECR): x - electrode placement, right forearm (adapted from Perotto [16]).

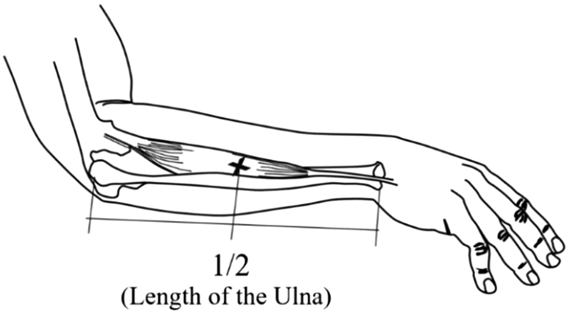


Fig. 2. *Extensor Carpi Ulnaris* (ECU): x - electrode placement, right forearm (adapted from Perotto [16]).

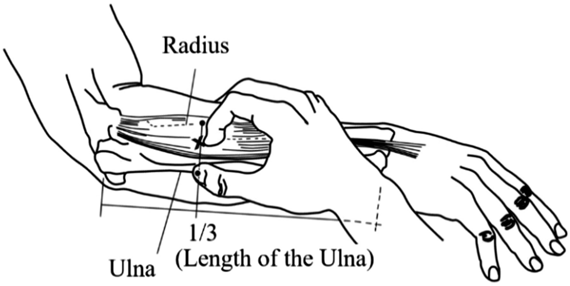


Fig. 3. *Extensor Digitorum* (ED): x - electrode placement, right forearm (adapted from Perotto [16]).

Concerning the APL muscle the situation is different, since the author's suggested position [16] for the insertional electromyography corresponds to the forearm area where this muscle resides, which is under the ED muscle. In this case, the best spot to apply the sensor was chosen according to Criswell' guidelines [13] considering the monitoring of the APL muscle activity using S-EMG (Fig. 4). Due to the proximity of the APL muscle to the *Extensor Pollicis Brevis* (EPB) muscle which, once activated, is related to thumb extension, the applied sensor records the muscular activity of both APL and EPB muscles [13] through S-EMG, as presented in Fig. 4. However, this condition does not abrogate the evaluation of muscular effort in abducting the thumb. The surface electrodes were placed using a double adhesive tape that kept them in permanent contact with the skin. As a complement, a soft elastic forearm sleeve was used to prevent eventual removal during the movements of the upper limb while the trials with a computer mouse were being performed. During all the placement process the x area was always being touched and checked while the participant activated the chosen muscle, according to Table 1 (see also Fig. 5). This procedure combined with real-time monitoring of the EMG signal helped optimizing the placement of the electrodes and repositioning them whenever necessary.

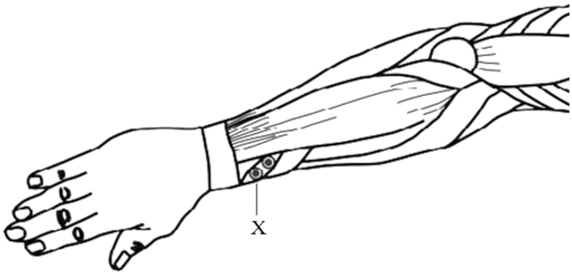


Fig. 4. *Abdutor Pollicis Longus* (APL): x - electrode placement, right forearm. (adapted from Criswell [13]).

Table 1. Muscle activation

	Muscles			
	ECR	ECU	ED	APL
Behavioral Test	Wrist extension and radial deviation	Ulnar deviation of the wrist	Finger extension	Abduction of the thumb (thumb up)

2.3 Electromyographic Recording and Muscle Activity Evaluation

Electromyographic recording was performed with a Bagnoli 4 channel system, DE-2.1 Single Differential Electrode. The single differential electrodes have two parallel cylindrical silver rods, with a 1 cm gap, which were applied according to the direction of the muscle fibers of the chosen muscle, while the circular reference electrode (adhesive disc) was placed on the iliac crest of the participant. The EMG signals were normalized from a set of three maximal voluntary contractions of each muscle and for each one of the 20 participants, each voluntary contraction lasting 3 s, followed by 60 s of rest in between trials [17]. The maximum voluntary contraction value (MVC), used to normalize the data, resulted from the maximum observed value of the three maximum contraction attempts. During the test, the researcher manually performed retention of the movements resulting from the joints attempt to move by way of contracting the focused muscle. The real-time viewing of the electrical signal, at first from only one channel and then from 4 synchronous channels turned out to be extremely useful for the pursuit of the best gain adjustment as well as in the search for any electrical noise. Therefore, whenever any noise effects such as *clipping, crosstalk, heart-rate*, among others, were detected, all the measures were taken in order to minimize or eliminate these effects. During actual data collection, a sample rate of 1000/s per channel and an amplification corresponding to a factor of $\times 1000$ and $\times 10000$ were used. The acquisition system was implemented through EMGworks 4.1.1 Acquisition software. The instantaneous EMG signals were applied to the RMS function through a moving window applied every 0.125 ms, and the monitoring of each muscle lasted 60 s. Then, data pretreatment was performed using EMGworks 4.1.1 Analysis software. The workstation and instrumentation (left side) implemented and used to acquire S-EMG signals and the actual computer mouse tested (right side) are shown in Fig. 6. Each subject performed standard graphical computer interface tasks [8] in the following order: pointing at large targets, pointing at medium targets and pointing at small targets at first. Then, dragging targets with the left button, dragging with the middle button, and steering targets inside a tunnel, finally the subjects performed a scroll up task.

Fig. 5. Electrode placement procedures (behavioral test with palpation of the active muscle).

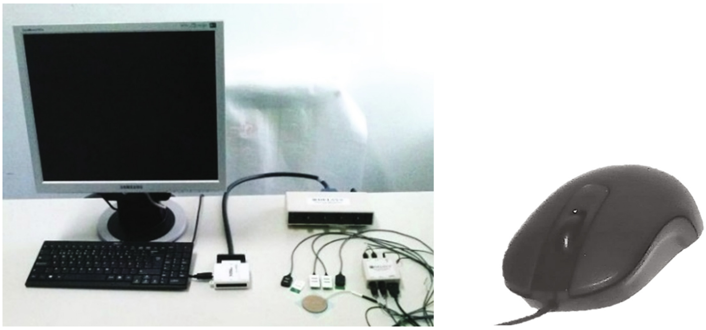


Fig. 6. Workstation (left side) and computer mouse tested (right side), Microsoft® optical 200 (dimensions: 106 × 56 × 30 mm, total mass: 78 g).

The experiments took place during the warm season with regular ambient temperature around 30 °C, whereby an air conditioning system was used in the small laboratory where the tests were conducted. Hence, indoor temperature ranged from 22 °C to 24 °C. This procedure avoided electrical conductivity problems due to skin sweating that could affect the measurements. EMG data normalized with MVC values for each muscle (ECR, ECU, ED and APL) and for each subject were subsequently transformed in order to obtain the Amplitude Probability Distribution Function (APDF) of the electromyographic signal [18]. The probability of amplitude at a certain level of muscle contraction is the probability of myoelectric activity being less than or equal to that level of contraction and may be expressed as the fraction of the total duration at which the signal is less than or equal to that level. If this fraction is estimated from a sufficiently large number of levels, a good estimate of the Amplitude Probability Distribution Function (APDF) is obtained. In the present study the APDF values were calculated by using 960 points per 60 s recording of EMG signal sampled at intervals of 0.0625 s. Analysis of the S-EMG signals was performed by normalizing the signals using a Root Mean Square function, followed by a procedure of Amplitude Probability Distribution Function calculation for the 10th (APDF10), 50th (APDF50) and 90th (APDF90) percentiles for each muscle monitored. APDF10 has been recognized as related with baseline activity, the APDF50 as median activity level and APDF90 as peak activity [19].

3 Results

Table 2 shows the mean values of APDF10, 50 and 90, as well the standard deviation, for a set of (seven) standardized graphical tasks performed by the subjects using the conventional PC mouse. The APDF values vary, approximately, between 7% MVC and 21% MVC. The highest values refer to the ECU and ED muscles for all three evaluation parameters for reveled, suggesting muscular load related to ulnar deviation performing the graphical tasks of pointing, dragging, steering and scrolling activities, as well as muscular load related to radial deviation of the wrist. The peak muscular activity indicator (APDF90) exceeds 20% of the maximum voluntary contraction for the ED muscle, suggesting this somewhat high load is related to the extension of the medial fingers, most probably due to the extension of the fingers that operate the mouse buttons. This condition may occur during the movement of the computer mouse when it is grasped and the fingers (index and middle finger) are stressed and lifted up above the device as observed by Lee [20].

Table 2. Average APDF values of muscular activity as a percentage of Maximum Voluntary Contraction (and Standard Deviation) - ECR, ECU, ED and APL muscles.

	Muscles			
	ECR	ECU	ED	APL
APDF10	9.83% (SD = 5.85)	13.75% (SD = 6.41)	13.18% (SD = 5.34)	6.96% (SD = 5.33)
APDF50	10.49% (SD = 5.81)	15.37% (SD = 6.81)	16.19% (SD = 5.38)	8.80% (SD = 6.14)
APDF90	11.44% (SD = 5.82)	17.99% (SD = 7.59)	20.62% (SD = 6.61)	11.33% (SD = 7.47)

4 Discussion

The results of muscle selection and S-EMG signal treatment were the basis for an experimental study yielding quantitative evaluations of muscular activity across a range of standardized tasks. The resulting values, while generally low, are higher for the ED and ECU muscles than for APL and ECR, suggesting a yet unexplored potential for reduction of peak muscular activity levels. This reduction might be achieved by changing the shape, promoting anchorage for the hand and, or, changing the size of the PC mouse, towards promoting less extension of the medial fingers and of the wrist in ulnar deviation.

A judicious process supported by expert advice led to the selection of the adequate instruments and the selection of four forearm muscles for monitoring muscular activity. 20 participants performed a set of graphical tasks described by Lourenço [8] while monitoring of muscular activity of the selected muscles using surface electromyography (S-EMG) ensued. Selection of the muscles for monitoring of activity resulted from a multi-pronged process. Literature review of previous studies concerning forearm muscle activity while using the PC mouse was one of the dimensions of the process. In a concurrent approach, muscles involved in sustaining inadequate and potentially harmful postures of the forearm (during PC mouse use) were also identified with a focus on wrist extension, medial finger extension, as well as radial and ulnar deviations.

Acknowledgements. The Authors which to thank Professor Rui Pitarma from Instituto Politécnico da Guarda and Doctor Cristina Roque Ferreira from Centro Hospitalar e Universitário de Coimbra for their support. In addition to the results reported in this paper, previous publications in English are available [8, 21, 22] reporting on results from the doctoral dissertation by Lourenço [9]. The current study was funded in part by Fundação para a Ciência e a Tecnologia project UID/EMS/00151/2013 C-MAST.

References

1. Nwaigwe, A.F.: Ergonomics Considerations in IT-Enabled Computer-Aided Design for Discrete Manufactured Products. Ph.D. Thesis. University of Pittsburgh, EUA (2015)
2. Søgaard, K., Sjogaard, G., Finsen, L., Olsen, H.B., Christensen, H.: Motor unit activity during stereotyped finger tasks and computer mouse work. J. Electromyogr. Kinesiol. **11**, 197–206 (2006)
3. Agarabi, M., Bonato P., De Luca C.J.: A sEMG-based method for assessing the design of computer mice. In: Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 4, pp. 2450–2453 (2004)
4. Lee, D., Fleisher, J., McLoone, H., Kotani, K., Dennerlein, J.: Alternative computer mouse design to reduce static finger extensor muscle activity. Hum. Factors **49**(4), 573–584 (2007)
5. Jensen, C., Borg, V., Finsen, L., Hansen, K., Juul-Kristensen, B., Christensen, H.: Job demands, muscle activity and musculoskeletal symptoms in relation to work with the computer mouse. Scand. J. Work Environ. Health **24**, 418–424 (1998)
6. Karlqvist, L., Hagberg, M., Selin, K.: Variation in upper limb posture and movement during word processing with and without mouse use. Ergonomics **37**, 1261–1267 (1994)

7. Keir, P., Bach, J., Rempel, D.: Effects of computer mouse design and task on carpal tunnel pressure. Ergonomics **42**, 1350–1360 (1999)
8. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Horizontal and vertical handheld pointing devices comparison for increasing human systems integration at the design stage. In: Nunes, I. (ed.) Advances in Human Factors and System Interactions. Advances in Intelligent Systems and Computing, vol. 497, pp. 15–24. Springer, Cham (2017)
9. Lourenço, M.L.: Desenvolvimento e Análise Ergonómica de Dispositivos Manuais Apontadores para Computador [Development and Ergonomic Analysis of Computer Handheld Pointing Devices]. Doctoral Dissertation. Industrial Engineering and Management. School of Engineering. Universidade da Beira Interior, Covilhã, Portugal (2016)
10. Houwink, A., Hengel, K.M.O., Odell, D., Dennerlein, J.T.: Providing training enhances the biomechanical improvements of an alternative computer mouse design. Hum. Factors J. Hum. Factors Ergon. Soc. **51**(1), 46–55 (2009)
11. Chen, H.M., Lee, C.S., Cheng, C.H.: The weight of computer mouse affects the wrist motion and forearm muscle activity during fast operation speed task. Eur. J. Appl. Physiol. **112**(6), 2205–2212 (2012)
12. Lin, M.Y.C., Young, J.G., Dennerlein, J.T.: Evaluating the effect of four different pointing device designs on upper extremity posture and muscle activity during mousing tasks. Appl. Ergon. **47**, 259–264 (2015)
13. Criswell, E.: Cram's Introduction to Surface Electromyography, 2nd edn. Jones and Bartlett Publishers, Burlington, MA (2011)
14. ISO 9241-400: Ergonomics of Human–System Interaction. Part 400: Principles and Requirements for Physical Input Devices International Organization for Standardization (ISO), Geneva, Switzerland (2007)
15. Rempel, D.M.: Personal Communication Regarding Appropriate Surface EMG Equipment and Selection of Muscles for Monitoring. Professor of Occupational Medicine, University of California, San Francisco CA (2013–2015)
16. Perotto, A.: Anatomical Guide for the Electromyographer: The Limbs and Trunk. Charles C. Thomas, Springfield, IL (1994)
17. Young, J.G., Trudeau, M.B., Odell, D., Marinelli, K., Dennerlein, J.T.: Wrist and shoulder posture and muscle activity during touch-screen tablet use: effects of usage configuration, tablet type, and interacting hand. Work **45**(1), 59–71 (2013)
18. Hagberg, M.: The amplitude distribution of surface EMG in static and intermittent static muscular performance. Eur. J. Appl. Physiol. Occup. Physiol. **40**(4), 265–272 (1979)
19. Jonsson, B.: The static load component in muscle work. Eur. J. Appl. Physiol. **57**, 305–310 (1988)
20. Lee, D.L., McLoone, H., Dennerlein, J.T.: Observed finger behaviour during computer mouse use. Appl. Ergon. **39**, 107–113 (2008)
21. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Association of hand size with usability assessment parameters of a standard handheld computer pointing device. In: Arezes, P.M., Baptista, J.S., Barroso, M.P., Carneiro, P., Cordeiro, P., Costa, N., Melo, R.B., Miguel, A.S., Perestrelo, G. (eds.) Occupational Safety and Hygiene IV, pp. 339–343. Balkena and CRC Press, Boca Raton (2016)
22. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Ergonomic development of a computer pointing device – a departure from the conventional PC mouse towards CAD users. In: Proceedings of the 19th Triennial Congress of the International Ergonomics Association, Melbourne, Australia (2015)

Research Methodologies in the Ergonomic Development and Evaluation of PC Mice

Miguel L. Lourenço[1] and Denis A. Coelho[2(✉)]

[1] Technical Scientific Unit of Engineering and Technology, Research Unit for Inland Development, School of Technology and Management, Instituto Politécnico da Guarda, Guarda, Portugal
mlopes@ipg.pt

[2] Human Technology Group, Department of Electromechanical Engineering, Centre for Mechanical and Aerospace Science and Technology, Universidade da Beira Interior, Covilhã, Portugal
dac@ubi.pt, denis.a.coelho@gmail.com

Abstract. Framing research within a methodology supports the planning which guides the research process. The process design defines a plan where research strategies aim collection of data and respective methods of analysis, in order to provide answers to research questions. This paper describes the research process conducted in the experimental and design work conducive to the doctoral dissertation by Lourenço [1] focusing on the development and ergonomic analysis of new PC mice. This research was framed within positivism, associated to exact sciences, adopting a structured research methodology, enabling enhanced replicability. The aim of this paper is report on research methodology, in particular in what concerns the intertwined research methodologies and associated research questions involved in the development and ergonomic evaluation of PC mice. The taxonomy of research approaches proposed by Järvinen [2] is the basis for the classification of the research methodology presented.

Keywords: Research questions · Human-systems integration · Ergonomics · Product development · Usability testing · S-EMG

1 Introduction

The design of the research process defines a plan where the research strategies aim the collection of data and the respective methods of analysis, in order to provide answers to the research questions. The research process conducted in the experimental and design work conducive to the doctoral dissertation by Lourenço [1] was orientated towards the development and ergonomic analysis of new PC mice. This research process was framed within positivism, associated to exact sciences, adopting a structured research methodology, enabling its replicability.

The methodology of the research carried out by Lourenço [1], involved inter-twined research methodologies and associated research questions employed in the development and ergonomic evaluation of PC mice. The taxonomy of re-search

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_42

approaches proposed by Järvinen [2] was the basis for the classification of the research methodology implemented.

In theory testing studies, several methods may be used, such as laboratory experiments and survey research. In this kind of research approach (theory testing) the theory is acquired from literature review or developed and refined for the study. Here the research question is "do observations confirm or falsify the theory?". In building a new artifact the research question underlying it is "is it possible to build a certain artifact?", while in an artifact evaluation approach the re-searcher may pose the question "how efficient is this artifact?". The experimental work that the methods presented in this paper support combined the approach for artifact building and artifact evaluation, comprised within a theory testing approach. The former approaches are the starting point for product development and comparative evaluation, with the effect of expanding existing theory as a result of their implementation.

In pursuing the eliciting of answers to the research questions, an elaborate re-search design was implemented making use of several research methods, including literature review, laboratory experiments, field experiments, product development and data collection instrument development. In implementing these re-search methods several research techniques were put to use, including the use of measuring devices as well as of assessment scales (featured in this paper), observation of the use of artifacts and prototyping.

2 Aims of the Research and Subsequent Research Plan Development

This research aims comparing the horizontal (traditional) pointing device with vertical and slanted, in terms of usability and muscular activity of the forearm. The recommended research process regarding the evaluation of these devices is based on the operationalization of instruments that are searchable and detectable by literature review. However, it is also expected that the proposed new processes or methods of analysis will be used to advance the state of the art regarding evaluation of usability and/or muscle activity in the use of computer handheld pointing devices.

The process of developing new geometries of pointing devices was conducted through a systematic and user-centered approach incorporating recommendations obtained from specialized literature. As a goal for one of the new geometries authors attempted to promote adequate palmar support, a neutral posture to be adopted by the individuals in static support of the right hand as well as to enhance the manual control of accurate and effortless grip. Increasing efficiency was also sought, by approximation of the thumb, index and middle fingers during the use of the device with the said characteristics. For this new model, a roughly conical geometry and lateral inclination at around 45° (half-pronation of the forearm) were envisaged. Another new geometry is also intended to facilitate palmar support, adoption of a neutral posture for the static support of the hand and accurate manual control grip, while adopting a geometry archetype less radical than the previous one. A geometry akin to the form of a shell, with support for the thumb, and a lateral slope at around 30° is proposed. These new models are intended to be tested along with other commercial devices and in particular with devices sold with the 'ergonomic' seal. The comparative tests are based on standardized tasks in order to achieve replicability in usability evaluation and specific muscle activity assessment.

2.1 Methodological Framework

Saunders, Lewis and Thornhill [3] used the onion metaphor to describe the general steps of any research process. The process begins in the more peripheral layer and it takes place in successive steps towards the core, and it is recommended not to proceed to a more inner layer before the process step relative to the previous layer has occurred. Of the six layers, three relate to the overall planning of the research process, strategies, choices, and time horizons, while the more peripheral layers refer to the research philosophies and approaches (deductive or inductive methodology) adopted. According to Saunders, Lewis and Thornhill [3], the first step in any research process is to define the appropriate research philosophy and, secondly, to position the research methodology on the most appropriate approach. In the third layer there are strategies, which include experimentation, research by questionnaires or interviews, case studies, action research, grounded theory, ethnography and research based of data available on file. In the fourth step, the choice of method typology (use of quantitative and, or, qualitative techniques) should be followed, followed by time horizons and, finally, the techniques and procedures used for data collection and analysis.

Thus, the research process conducted within the doctoral thesis by Lourenço [1] assumes the following framework: as far as philosophy of research it is based on the concept of positivism, it is associated with the exact sciences, and the research methodology in use is structured, thus facilitating replication. Positivism is also associated with the observation of facts in the form of quantifiable measures, in addition to which observations and conclusions are based on statistics, with the researcher assuming a role an observer who does not influence observations. As regards research approaches, a deductive approach (hypothesis and theories formulation that are then tested by the research strategy), of the quantitative type, is used in the product evaluation process since data collection and analysis yield numerical results. However, the inductive approach is applied in the development of a new evaluation tool for computer handheld pointing devices. The research strategy is based on experimentation, involving the choice of multiple methods, since more than one method is used, always obtaining quantitative data from experimental studies. Regarding temporal horizons, the research includes a cross-sectional component (usability assessment and evaluation of muscle activity), but also includes a longitudinal component in usability evaluation (Fig. 1).

Moreover, given the nature of the research reported in this paper which contemplates an artifact creation approach (product development) and an artifact evaluation approach (regarding usability and muscular activity during artifact use), the methodology used in the combined process of design and research can be appropriately classified through the taxonomy of research methods developed by Järvinen [4]. Although the proposed taxonomy is oriented toward the development of information systems (IS), its broad application to technical systems and technology is assumed. The artifact can be

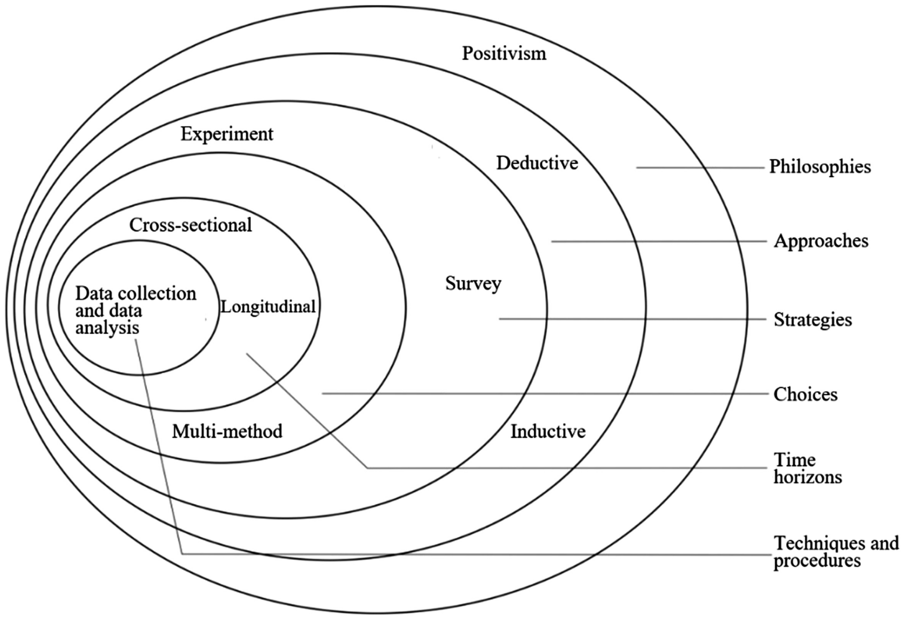


Fig. 1. Research onion [3] as applicable to the research carried out by Lourenço [1].

classified as something intangible or tangible. This involves the operationalization of a scientific research approach focusing in particular on the usefulness of artifacts and combining the ‘artifacts building’ approach with the ‘artifact evaluation’ approach. Järvinen [4] defines a research approach as a set of research methods that can be applied to a research object and to similar research questions. It therefore considers different approaches to research and indicates various research methods that can be applied in the context of a particular approach. The taxonomy of research methods proposed by Järvinen [4] presents 6 categories, namely, mathematical, conceptual-analytical, theory-testing, theory-creation, artifact-creation and artifact-evaluation approaches.

In theory testing studies a number of methods (or strategies) are used such as laboratory experimentation, questionnaire research, field study, etc. In a study guided by a theory testing approach, the theory, model or framework is acquired from literature studies research or it is developed or refined for the study at hand. The research question may be: do the observations confirm or disprove that theory? In turn, the following methods (or strategies) are often used in studies aimed at theory creation: case study, ethnographic method, grounded theory, contextualism, discursive analysis, longitudinal study, phenomenological study, etc. With regard to approaches that emphasize the usefulness of artifacts, in the construction of a new artifact the research question could be: Is it possible to construct a certain artifact? As far as the evaluation of an artifact is concerned, the researcher may ask: How efficient is this artifact? When searching for answers to research questions, criteria are used and measurements are taken. Regarding the evaluation of artifacts, Sweeney et al. [5], cited by Järvinen [4], suggest an investigative methodology regarding the evaluation of human-computer interaction for usability evaluation, composed of three dimensions: the evaluation strategy, the type of evaluation and the evaluation time in the context of the product life cycle. Such strategies are listed as user-based, theory-based, and expert-based. Cunningham [6], quoted by Järvinen [4], indicates that there are at least nine types of case studies, suggesting that comparative cases are aimed at developing concepts based on case comparisons, stating that the comparative cases belong to theory testing approaches.

The research methodology based on the development of systems, according to Nunamaker et al. [7], cited by Järvinen [8], fits comfortably in the category of applied science, belonging to the formulative research type in engineering development. This methodology was considered both general and fulcral, representing a super-methodology containing sub-methodologies. The researchers outlined the following steps for the research methodology they presented: 1 - construct a framework or conceptual framework; 2 - develop a system architecture; 3 - analyze and design the system; 4 - build the prototype and 5 - observe and evaluate the system. Steps 2, 3 and 4 pertain to the development of the system itself. Step 1 is associated with the sub-methodology of constructing the theory from a utility point of view, that is, to support the artifact creation process. The said authors also present several methods that integrate the research process. With particular reference to the focused research process, reference should be made to the methods relating to point 5, to observe and evaluate the system, to observe the use of the system through case studies and field studies, and to evaluate the system through laboratory experimentation or field experimentation.

In order to satisfy the general objective of the focused research: "To perform a comparative ergonomic study between different geometries of computer pointing devices and to develop new geometries, integrating them in the comparative study, seeking to satisfy and validate assumptions of ergonomic character; develop new indicators and evaluation tools for the classification of devices" Järvinen's [4] taxonomy is applied to the research process. It is an empirical study that sits on the approach of theory testing and expansion, also encompassing the approach to the creation of artifacts and the approach for evaluating the artifacts (Fig. 2).

The more traditional research processes, according to Bhattacherjee [9], tend to be deductive and functionalist. Such process is divided into three phases: exploration, research project and research execution.

The research design phase includes the selection of a research methodology, the operationalization of constructs under interest, and the development of an adequate sampling strategy. The operationalization consists of the operational design that defines the construct under interest. In this context, we search for existing, pre-validated instruments that can be used directly or modified to measure the construct. If the instruments do not exist or the results obtained from them are poor or reflect a different conception than the one the researcher prefers or defends, new instruments can be designed to measure the construct. At the same time, researchers must decide which research methods are most appropriate for collecting data in order to provide answers to research questions. These methods may include quantitative methods such as experimentation or survey research, or qualitative methods such as case studies or action research, or the combination of both. Another important aspect is the choice of the target population and the strategy for selection of the samples of this population

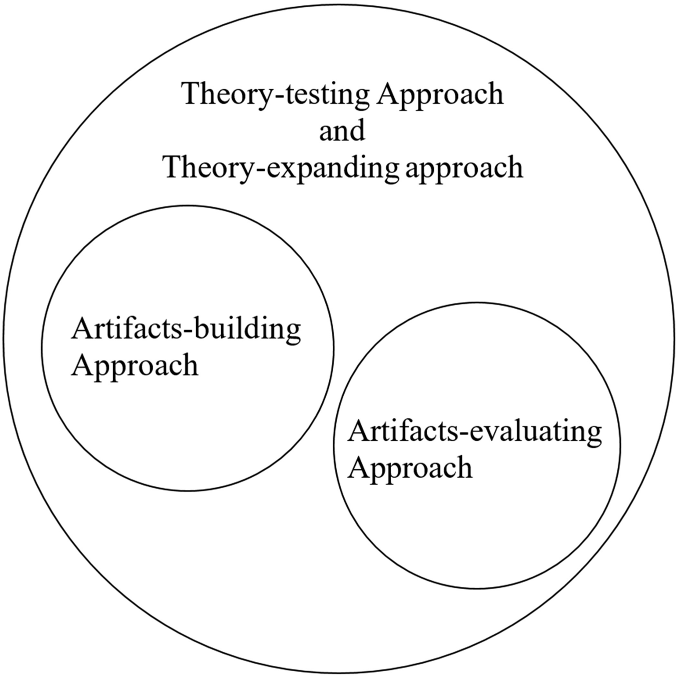


Fig. 2. Research approaches selected from Järvinen's taxonomy [4] and intertwined that are applicable to the research carried out by Lourenço [1]

(random or convenience sampling technique). Bhattacherjee [9] states that care should be taken in the selection of a sample in order to avoid biased observations, which is not always possible when researchers only have the possibility of using less generalized and random sampling procedures, such as the use of convenience samples.

Between the research design and the research execution stages, it is useful to write a detailed proposal of the intended research. The research proposal should describe the procedures inherent in the research process with reasons given for the decisions taken on the basis of the literature review. Moreover, it should address the research questions, prior art, theories and hypotheses to be tested, methods for measuring the constructs and the sampling strategy to be used. Upon deciding what to study (concepts), who to study (subjects), and how to obtain the data (research methods and techniques), the research process may proceed with implementation, i.e. pilot testing, data collection and analysis of the data. In the collection of data stage, quantitative research techniques such as measuring instruments and scales of evaluation and, or, qualitative research techniques, such as ethnographical methods, may be used. In data analysis, quantitative methods include statistical techniques (e.g. regression, parametric, non-parametric, etc.). Qualitative techniques include content analysis and coding, for example. The process ends with the final research report, which should be drafted in sufficient detail, and no more than that, in order to allow the replication of the research study by other researchers, testing the results, or assessing whether the conclusions are scientifically acceptable.

2.2 Objectives and Research Methods and Techniques

The research methodology reported in this paper is based on a multi-objective research study. The following objectives were set for that study: to present the multiple domains of knowledge relevant to the research process; to present a comprehensive methodological framework linking the specific objectives and the multidisciplinary components of the research process; to carry out the methodological presentation of product design and specific experimental design with respect to usability assessment and assessment of muscle activity; to develop innovative geometries supported by literature review and adequate design methodology; to establish the state of the art of usability evaluation using a focused literature review; to develop and characterize the experimental design for evaluating the usability of the devices under consideration; to propose tools to advance the state of the art of evaluating the usability of computer mice and contribute to inform consumer choice, to establish the state of the art of the assessment of muscular activity through surface electromyography (S-EMG) using a focused literature review; to develop and characterize the experimental design for evaluation of muscular activity by S-EMG during the use of the devices under consideration; to propose processes or methods of analysis to advance the state of the art of the evaluation of muscular activity with respect to the use of PC mice; and, to present the data in graphical form and carry out its statistical analysis.

The research methodology of the research focused in this paper, is supported by the taxonomy of Järvinen [4], combining the approach to artifact creation with the approach to artifact evaluation, which are both encompassed in the theory testing approach to research. The two approaches contained in the latter are the starting point for product development and a comparative evaluation, with the expansion of the theory (creation of new indicators) as a result. Research methods and techniques, included in the research methodology, are associated with the objectives that underlie their application, in the following paragraphs.

The research methods encompassed in the research design are literature review, laboratory experimentation (a quantitative method), field experimentation (a quantitative method), the process of designing research per se, the product design (development) process, the process of designing (developing) instruments and the expansion of knowledge (based on research results). The quantitative research techniques employed for data collection are based on the use of measuring instruments, evaluation scales, the observation of the use of artifacts and prototyping (product development for experimentation support). The sampling techniques or strategies deployed consisted primarily of convenience sampling (with partial randomness). Data analysis resorted to quantitative techniques, notably statistical techniques were used for the analysis of quantitative data.

2.3 Research Questions

An empirical study based on the theory that the most appropriate way to classify computer handheld pointing devices should be to value their usability while simultaneously reducing muscle activity as well as promoting the adoption of more neutral postures by the user, follows a theory testing approach. In this context, the research question is: Do the observations confirm the theory or not? In the process of creating new parameters and indicators, such as an efficiency indicator, characterized from field experimentation results, an approach of creation or expansion of the theory is adopted, according to Järvinen's research methodology [4, 8]. Here the research question takes the following formulation: What kind of construct or model would explain the observations made?

The process involving the development of new hand-held computer pointing device geometries follows an artifact building research approach to. In this way, the following research question is associated with it: Is it possible to construct a particular artifact according to the requirements derived from the theory? For the evaluation process the devices have been subjected to, with regard to usability assessment more specifically, the research question is: How effective are the designed and prototyped artifacts?

3 Discussion

The content of this article defines the methodological approach of research carried out within the scope of a focused research process, presenting the research methodology that supports it and the chosen research methods and techniques that are oriented toward the fulfillment of the objectives outlined. Thus, the article serves as a case example of a methodological plan that guides a significant research process, which has had part of its results reported elsewhere [10–12].

Acknowledgements. The current study was funded in part by Fundação para a Ciência e a Tecnologia project UID/EMS/00151/2013 C-MAST.

References

1. Lourenço, L.M.L.: Desenvolvimento e Análise Ergonómica de Dispositivos Manuais Apontadores para Computador [Development and ergonomic analysis of computer handheld pointing devices]. Doctoral Dissertation, Industrial Engineering and Manage-ment, School of Engineering, Universidade da Beira Interior, Covilhã, Portugal (2016)
2. Järvinen, P.: On a variety of research output types. In: Proceedings of the IRIS, vol. 23, pp. 251–265 (2000)
3. Saunders, M., Lewis, P., Thornhill, A.: Research Methods for Business Students, 6th ed., Edited by Pearson Education Limited (2012)
4. Järvinen, P.: On a variety of research output types. Department of Computer and Information Sciences working paper, University of Tampere D-204-6 (2004)
5. Sweeney, M., Maguire, M., Schackel, B.: Evaluating user computer-interaction: a framework. Int. J. Man Mach. Stud. **38**, 689–711 (1993)
6. Cunningham, J.B.: Case study principles for different types of cases. Qual. Quant. **31**, 401–423 (1997)
7. Nunamaker, J.F., Chen, M., Purdin, T.D.M.: Systems development in information systems research. J. Manag. Inf. Syst. **7**(3), 89–106 (1991)

8. Järvinen, P.: Research questions guiding selection of an appropriate research method. Department of Computer Sciences, University of Tampere D-204-5 (2004)
9. Bhattacherjee, A.: Social Science Research: Principles, Methods, and Practices. Textbooks Collection, Book 3 (2012)
10. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Association of hand size with usability assessment parameters of a standard handheld computer pointing device. In: Arezes, P.M., Baptista, J.S., Barroso, M.P., Carneiro, P., Cordeiro, P., Costa, N., Melo, R.B., Miguel, A.S., Perestrelo, G. (eds.) Occupational Safety and Hygiene IV, pp. 339–343. Balkena and CRC Press, Leiden (2016)
11. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Ergonomic development of a computer pointing device – a departure from the conventional PC mouse towards CAD users. In: Proceedings of the 19th Triennial Congress of the International Ergonomics Association, Melbourne, Australia (2015)
12. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Horizontal and vertical handheld pointing devices comparison for increasing human systems integration at the design stage. In: Nunes, I. (ed.) Advances in Human Factors and System Interactions. Advances in Intelligent Systems and Computing, vol. 497, pp. 15–24. Springer, New York (2017)

Development of a New Ergonomic Computer Mouse

Miguel L. Lourenço[1], Rui A. Pitarma[1], and Denis A. Coelho[2(✉)]

[1] Technical Scientific Unit of Engineering and Technology, Research Unit for Inland Development, Instituto Politécnico da Guarda, Guarda, Portugal
{mlopes, rpitarma}@ipg.pt

[2] Human Technology Group, Department of Electromechanical Engineering, Centre for Mechanical and Aerospace Science and Technology, Universidade da Beira Interior, Covilhã, Portugal
dac@ubi.pt, denis.a.coelho@gmail.com

Abstract. Developing a new computer pointing device abiding to the requirements set out in ergonomics literature necessitates joining contributions from several areas including the systematization recommended for product design projects, ergonomic recommendations and principles and guidelines applicable to hand tools as well as anthropometric considerations. The development of new geometries for computer pointing devices constitutes a complex process as these are simultaneously hand tools, enabling the interaction with the computer, and at the same time, meant for use by people from both sexes and practically all ages, as its use today is practically ubiquitous. The paper reports on a study aimed at developing an innovative PC mouse geometry supported by literature review and an adequate design methodology. In particular, reducing forearm pronation was set as a goal for the new design.

Keywords: Ergonomic product development · Ergonomic computer mouse · Human-systems integration · Usability

1 Introduction

Computer usage can be associated with the development of upper extremity pain, especially hand and forearm musculoskeletal pain induced by intensive mouse use [1]. In this regard, it is commonly accepted that the use of PC mice which promote the reduction of forearm pronation is less susceptible to cause musculoskeletal injuries when compared to more conventional designs. Notwithstanding the increasing use of touch screens, extended use of the PC mouse is bound to endure in present and future time, because in computer tasks like pointing, dragging and steering graphical targets, continuously needed in some computer applications, like in 3D-CAD, touch screens have so far not been able to replace the mouse [2]. Alternative geometries have been evaluated, Lourenço *et al.* [3] in an experimental set up with 20 participants, performed usability evaluation of two PC mice geometrically quite distinct, one was a standard device (*Microsoft optical 200*) and the other device was an alternative vertical PC

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7_43

mouse, supporting the adoption of a neutral pronation forearm posture (*Evoluent VerticalMouse 4 Right*). The results of the comparison reported suggest designing hybrid configurations of handheld pointing devices, in order to achieve a compromise between the expected long term effects on health and the usability parameters. Another study suggests that a proper mouse weight could hence benefit users in terms of increasing movement efficiency; its dimensions and geometry should be based on anthropometry, hand gestures and comfortable hand postures [4]. Hand size of the subjects seems to make a difference during mouse usage, affecting grasp position and the level of muscle activity, suggesting that a computer mouse ought to be chosen according to the size of the hand of the subject [5]. Moreover, previous tests performed on a standard PC mouse revealed statistically significant association between hand width and effectiveness of dragging targets with the middle (scroll) button of the mouse [6]. Hence, the present study was conducted by compyling a set of requirements and recommendations, as well as guidelines, towards improvements in respect to characteristics previously identified in literature as requiring melioration, focusing on ergonomic and usability considerations.

2 Methods

2.1 Product Development Operational Model

A bespoke design method was created and implemented to support attaining the goals set forth for the present study. A task clarification for the product under development was accomplished through the activities of product goals definition, analysis and definition of product requirements and setting up of the product specification with support from literature review. Concept generation was attained by means of sketches and study of variants, leading to a structured presentation of alternative concepts. The evaluation and refinement stage involved the definition of an evaluation matrix derived from the specification, selection of the best concept, execution of mockups and improvement and refinement of mockups. The detailed design stage was conducted in view of prototyping as well as ergonomic assessment and usability evaluation.

The generic operational model for product design proposed by Hales [7], reviewed by Lewis and Bonolo [8] and cited by Coelho [9] was adopted as a reference. In an industrial design context, the subordinate processes are divided into five phases: 1-task clarification; 2-concept generation; 3-evaluation and refinement (of design concepts); 4-detailed design (of the preferred concept) and 5-communication of results (of the design project). However, for the present study, having an appropriate physical prototype for comparative ergonomic evaluation (validation) was considered the main goal throughout this particular design process. Hence, some adaptations from reference operational model were undertaken (Fig. 1).

In this context, subordinate processes 3 (evaluation and refinement) and 4 (detail design) were, in a certain way, grouped. The study on the production processes is not reported in this paper and subordinate process 5 (communication of results) is assumed to be out of context within the scope of the present paper that focuses on the design with the goal of achieving the stage of evaluation with special emphasis on the detailed

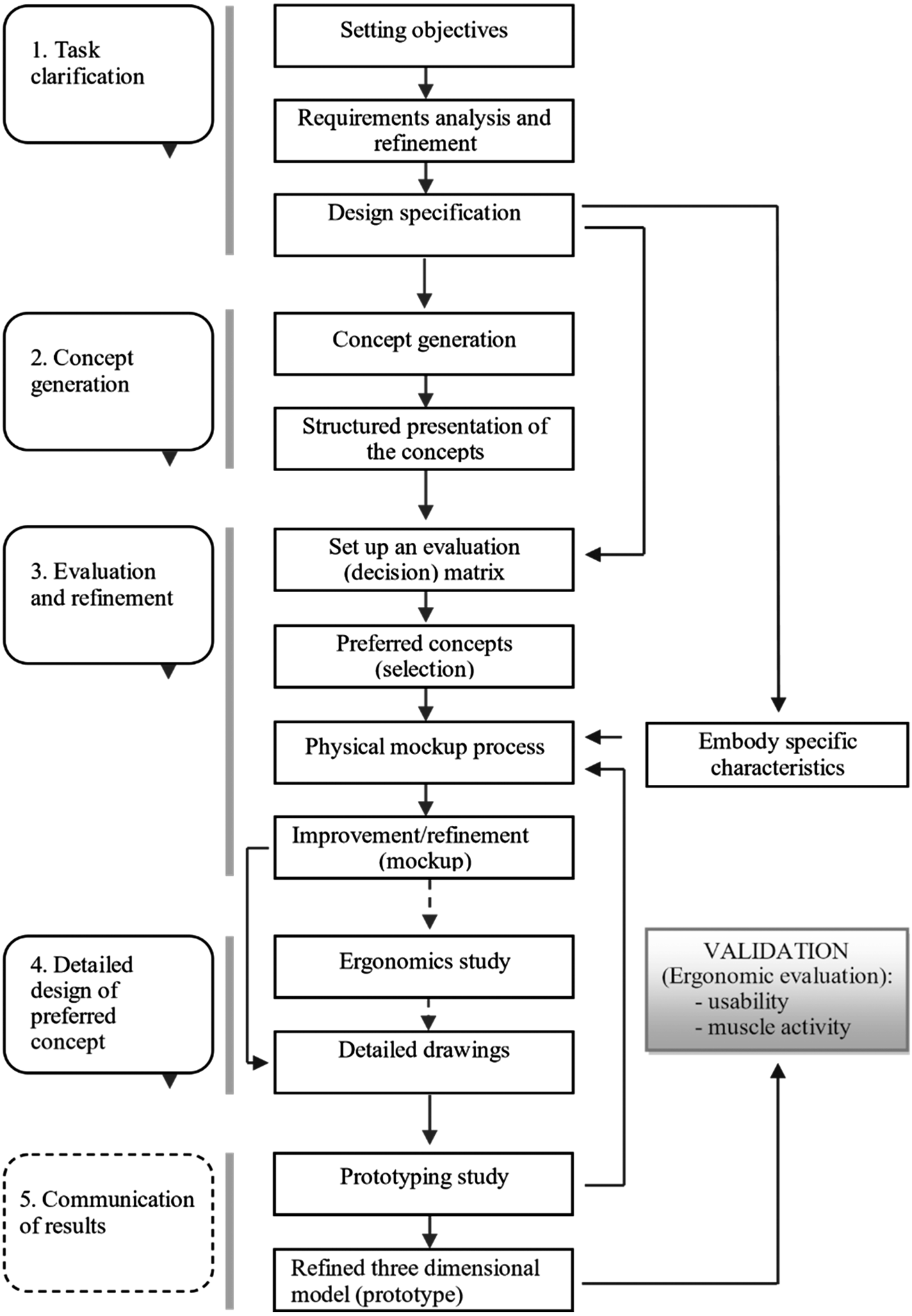


Fig. 1. Flowchart of the operational model adopted in product development (new PC mouse geometry) (adapted from Lewis and Bonollo [7], Hales [8].

ergonomic study. Thus, development of functional prototypes occurred upstream of communication of results since these were indispensable to the ergonomic study, which included studies of usability and muscle activity (surface electromyography – (S-EMG)) of the participants. The evaluative studies involved several dozen users carefully selected according to ISO 9241-9: 2000 [10].

Figure 1 shows the flow of information that took place between the various stages. A set of dashed lines shows the stage of communication of results and other sequences of operations that were not formally carried out according to the operational model of the design process, but which are being implemented in dissemination efforts, such as this paper. Detailed drawings were needed as early as in the process of developing the models, which were materialized by 3D printing of a thermoplastic. Subsequently, after needed refinement of the models, it was again necessary to use detailed designs, which are fundamental in the process of realizing functional physical prototypes. The shape, dimensions and location of the PC mouse buttons are among the features that were tested during upgrade and refinement of the mockups. Thus, said detailed drawings integrated the three-dimensional computer aided design necessary for the 3D printing process, a method that was adopted in the production of both the mockups and the functional prototype. The prototype was developed for comparative tests with commercial models. As shown in Fig. 1 the flow of information can occur between the several stages with the possibility of returning to previous steps from downstream phases. This iterative process allows improving and refining the models in order to incorporate features springing from the specification. However, the implementation of certain requirements is only possible in the execution phase of the functional prototype, including those dependent on the mechanisms and electronic circuits housed inside the prototype. The measurement of the force required to activate the buttons is a parameter, and, or, characteristic only measurable in the fully functional prototype; the same applied to the resistance to inadvertent button activation during the intended use of the computer mouse.

2.2 Task Clarification

The new geometry should embody features fitting the requirements of the applicable standards, the principles of ergonomics and the requirements resulting from the analysis of the scientific literature of the specialty. The prototype should use the same hardware as a reference device in order to study (validate) the geometry compared to other geometries of computer mice available in the market. A characteristic inclination between 90° and 0° is desired, respectively between the neutral forearm posture (0° pronation) and 90° pronation of the forearm of the user. The hardware of the *Microsoft optical mouse 200* commercial model was selected for use in the new geometry, whose implementation occurred in the execution phase of the physical prototype.

Requirements and recommendations for PC mice design were collected from standard ISO 9241 (Ergonomics of human-system interaction) [10–13] as well from scientific publications [14–23]. Hence, the specification listed below was developed based on all of these requirements and recommendations. Design specification:

- input devices should be designed so that they can operate without requiring undue deviation of the hand, fingers, arm, shoulder and head from their respective neutral positions;
- the input device should be operated by the user without undue deviation from a neutral posture and without excessive effort, hence, the biomechanical load shall be minimized and the design of the device shall take into consideration the minimizing of static muscle load;
- the input device should minimize the need for extreme positions such as wrist extension, radial or ulnar deviation, and forearm pronation;
- for common tasks or commands, the wrists and forearms should be near their neutral postures and wrist and finger extension are not recommended, in general the most comfortable hand gestures are those where the wrists are kept straight and the fingers slightly flexed (gently curved) or in a loose fist;
- input devices should be designed to accommodate the hand size of the intended user population;
- the weight, and hence inertia, of the input device should not degrade the accuracy of the device during use;
- the input device should be designed to be resistant to inadvertent button activation during its intended use and it should be possible to press the buttons on the mouse without reducing control of the device;
- the device shape and the buttons location should minimize finger extension or other movement or positioning that could cause finger strain or static load of the extensor muscles of any fingers;
- buttons should be shaped to assist finger positioning and button actuation;
- buttons should have a displacement force within the range of 0.5 N to 1.5 N until actuation and should have a minimum displacement of 0.5 mm and maximum of 6 mm;
- the motion sensing point should be located under the fingers (precision grip posture) rather than under the palm of the hand;
- grip surfaces should be of sufficient size, shape and texture to prevent slipping;
- the input device shall enable anchoring some part of the fingers, hand, wrist, or arm on either the input device or the worksurface to create a stable relationship between the hand and the point of action;
- the input device should promote an intuitive interface, adapting to skills already acquired to minimize the learning threshold and optimizing for perceived comfort;
- the intended use of an appropriately designed input device for a primitive task (such as pointing, selecting and dragging) is either obvious or easily discovered;
- the input device should be effective, efficient and satisfactory for the task being performed and the intended work environment;
- the new geometry should be innovative.

2.3 Concept Generation

In this stage, concepts are generated based on the specification presented in the previous list. Two of the generated concepts are presented in Fig. 2 as viewed from two alternative perspectives. In this conceptual phase, an attempt was made to incorporate the characteristics necessary to fulfill the specification, establishing compromises in the satisfaction of conflicting requirements, through form solutions. Some of the requirements involved in more conflicts were: offer of support to the palm of the hand, finger anchoring to ease the movement of the pointing device with less effort of the fingers (trying to minimize muscular activity), precision grip by approaching the distal ends of the index finger and thumb, promotion of postures in which the adjacent fingers (middle finger, annular and pinky) do not assume positions different from each other, innovative aesthetics, among others. The geometries were created using manual modeling techniques [24] using clay paste. This method enables the accomplishment of preliminary tests, either of form or of functional movement of the device, also testing several aspects related to the postures of the hand fingers. It was thus possible to generate new geometries, adapted to the anatomy of the hand, despite the difficulties that this task imposed [13].

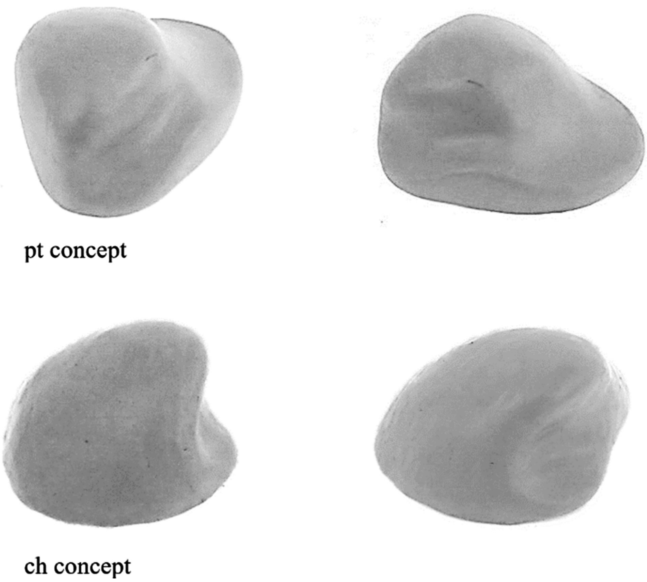


Fig. 2. Concept generated.

2.4 Evaluation, Refinement and Detailed Design of Preferred Concept

From the specification (requirements) and the subsequent generation of concepts, an evaluation matrix composed of 16 criteria was established (Table 1). Each assessment criterion was previously assigned a weight ranging from 1 to 3. The value 3 was attributed to characteristics considered more important and the value 1 was attributed to characteristics considered less important in the present context. For each concept, a score ranging from 1 to 4 was attributed to the satisfaction of each of the criteria, the value 1 corresponding to the worst classification. The evaluation matrix was then completed by multiplying the score given to each concept by the weight previously assigned to the respective criterion under consideration. The total score obtained was 83 points for the pt concept and 102 points for the ch concept. The ch concept was then selected and advanced to the model making process. In the process of selection of the preferred concept, the manually produced clay models proved to be decisive for a thorough evaluation.

Table 1. Evaluation matrix (weights for each factor ranging from 1 (less important) to 3 (most important); the rating for each factor range from 1 (worse) to 4 (better)).

Rating criteria (factor)	Weight	Concept			
		pt		ch	
		Rating	Score	Rating	Score
The input device shall enable anchoring some part of the fingers and, or, the hand	1	2	2	2	2
The input device should minimize ulnar and radial deviation of the hand during motion	3	1	3	2	6
The input device should minimize wrist extension and wrist flexion during motion	3	2	6	2	6
The input device should minimize forearm pronation and forearm supination	3	2	6	3	9
The shape and location of the buttons should minimize finger extension and finger strain	3	4	12	4	12
The input device should be designed to accommodate the hand size of the intended user population	2	1	2	2	4
The hand (fingers) should keep slightly flexed (gently curved) or in a loose fist when grasping the device	2	3	6	3	6
The input device shape should avoid discordant adjacent fingers postures (middle finger, ring finger and pinky)	2	4	8	4	8
The input device shape and the buttons location should avoid finger extension when clicking or static load of the extensor muscles of any fingers	3	2	6	3	9
The input device shape should facilitate the implementation of the most suitable buttons	3	3	9	4	12

(*continued*)

Table 1. (*continued*)

Rating criteria (factor)	Weight	Concept			
		pt		ch	
		Rating	Score	Rating	Score
The motion sensing point should be located under the fingers (precision grip posture)	2	4	8	4	8
The input device center of gravity should be situated on the grasp axis with regard to handle grasp	2	2	4	3	6
The input device should adapting to skills already acquired	1	3	3	4	4
The physical characteristics of the input devices should conform to the established stereotypes	1	3	3	4	4
The input device should promote an intuitive interface	1	3	3	4	4
The new input device geometry should be innovative	1	2	2	2	2
Total weighted score			83		102

The improvement and refinement of the models correspond to an iterative process in an attempt to incorporate the various desired characteristics. Figure 3 illustrates the final model of the ch concept. This geometry, in contrast to other more disseminated models, presents a greater volumetry (Fig. 5) allowing a more effective curved support for the hand and the fingers and also the reduction of pronation of the forearm (offering support for the hand with 30° of inclination). The ch geometry (Fig. 3) also has a pronounced cavity for the support of the thumb, allowing anchoring and stabilization during guiding of the displacement of the device.

Fig. 3. ch concept mockup (male adult hand, length 190 mm, width 88 mm, corresponding respectively to 40th percentile and 50th percentile, according to Gordon *et al.* [25] – United States Army Personnel sample).

The model enabled testing the ease of movement of the device, the deviation of the fingers, the influence of the design of the buttons on the positioning of the fingers and the ease of operation and also the suitability of the shape and size of the surfaces in contact with the hand. These tests enabled refining the model thus leading to the development of the functional prototype. After refinement of the model, the technical drawings needed to develop the functional prototype using 3D-CAD techniques (Fig. 4) were improved.

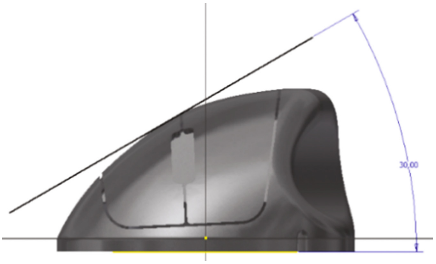


Fig. 4. ch concept frontal view obtained from 3D-CAD model (slanted angle of about 30°).

Figure 4 shows the digital model during its three-dimensional parametric modeling, based on a reverse engineering (using a 3D scanner) process complemented with the transposition of the physical dimensions of the models by manual collection of the dimensions. This process enables generating alternative sizes of the same geometry or varying the proportions of its main dimensions. Figure 5 shows the functional prototype of the new model (on the right), with all the mechanisms and electronic circuits of the reference model (shown on the left) being implemented inside it. Figure 5 also shows, comparatively, the differences in shape and size between the reference model and the new model. The prototype was materialized in two pieces (carcass and base) using the Fused Deposition Modeling technique with ABS thermoplastic. The iterative process, illustrated in Fig. 1 led to the final prototype and was then followed by the remaining tests, such as the measurement of the force required to push the buttons and the tests for ergonomic validation of the new geometry according to the operating model adopted (Fig. 1).

Fig. 5. Left: Microsoft optical mouse 200 (dimensions: 106 × 56 × 30 mm, total mass: 78 g); right: ch prototype (dimensions: 120 × 90 × 49 mm, total mass: 128 g).

3 Results

Requirements, guidelines and recommendation for the development of PC mice were surveyed and compiled both from ISO 9241 (Ergonomics of human-system interaction) standard series and from multiple applicable previously published scientific experimental studies. A product requirements specification consisting of both qualitative as well as quantitative product requirements was elicited. Two conceptual sketches were assessed against a 16 criteria evaluation matrix, and scored using a system of weighted criteria yielding the preferred concept for further development. A mockup was generated and subject to preliminary testing (shape and fitness to hand anthropometric dimensions). As an outcome of the development process depicted, a fully functional prototype was unveiled and more quantitative physical testing ensured (e.g. measuring the force required to activate buttons).

4 Discussion

Demonstrating an approach to principle driven design, following a systematic product development methodology, a new ergonomic mouse was developed and prototyped to a fully functional state, to enable usability evaluation and ergonomic assessment. The multiple requirements imposed on the computer pointing devices, necessitated establishing trade-offs between seemingly conflicting requirements which implies that a large variety of alternative designs may be created from the same product specification. Usability evaluation and ergonomic assessment is hence needed to sort out the complexity and identify the best alternatives. The study is targeted towards satisfying a set of requirements and recommendations, as well as guidelines, towards improvements in respect to characteristics previously identified in literature as requiring improvement, with a fundamental focus on ergonomic and usability considerations. Validation from a usability and a muscular activity standpoint were carried out and are to be reported elsewhere.

Acknowledgements. The current study was funded in part by Fundação para a Ciência e a Tecnologia project UID/EMS/00151/2013 C-MAST

References

1. Conlon, C.F., Krause, N., Rempel, D.M.: A randomized controlled trial evaluating an alternative mouse or forearm support on change in median and ulnar nerve motor latency at the wrist. Am. J. Ind. Med. **52**(4), 304–310 (2009)
2. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Ergonomic development of a computer pointing device – a departure from the conventional PC mouse towards CAD users. In Proceedings of the 19th Triennial Congress of the International Ergonomics Association, Melbourne (2015)

3. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Horizontal and vertical handheld pointing devices comparison for increasing human systems integration at the design stage. In: Nunes, I. (ed.) Advances in Human Factors and System Interactions. Advances in Intelligent Systems and Computing, vol. 497, pp. 15–24. Springer, Cham (2017)
4. Rempel, D.M., Camilleri, M.J., Lee, D.L.: The design of hand gestures for human-computer interaction: lessons from sign language interpreters. Int. J. Hum. Comput. Stud. **72**(10–11), 728–735 (2014)
5. Agarabi, M., Bonato P., De Luca C.J.: A sEMG-based method for assessing the design of computer mice. In: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society, vol. 1, pp. 2450–2453, San Francisco (2004)
6. Lourenço, L.M., Pitarma, R.A., Coelho, D.A.: Association of hand size with usability assessment parameters of a standard handheld computer pointing device. In: Arezes, P.M., Baptista, J.S., Barroso, M.P., Cordeiro, P., Costa, N., Melo, R.B., Miguel, A.S., Perestrelo, G. (eds.) Occupational Safety and Hygiene IV, pp. 339–343. Balkena and CRC Press, Leiden (2016)
7. Lewis, W.P., Bonollo, E.: An analysis of professional skills in design: implications for education and Research. Des. Stud. **23**, 385–406 (2002)
8. Hales, C.: Analysis of the Engineering Design Process in an Industrial Context. Gants Hill Publications, Eastleigh (1991)
9. Coelho, D.A.: A method for user centering systematic product development aimed at industrial design students. Des. Technol. Educ. Int. J. **15**(2), 44–54 (2010)
10. ISO 9241-9:2000: Ergonomic requirements for office work with visual display terminals – part 9: requirements for non-keyboard input devices (2000)
11. ISO 9241-11:1998.: Ergonomic requirements for office work with visual display terminals (VDTs) – part 11: guidance on usability (1998)
12. ISO 9241-400:2007.: Ergonomics of human-system interaction – part 400: principles and requirements for physical input devices (2007)
13. ISO 9241-410:2008+A1:2012.: Ergonomics of human-system interaction – part 410: design criteria for physical input devices (2012)
14. Aaras, A., Dainoff, M., Ro, O., Thoresen, M.: Can a more neutral position of the forearm when operating a computer mouse reduce the pain level for visual display unit operators? A prospective epidemiological intervention study: part II. Int. J. Hum. Comput. Interact. **13**, 13–40 (2001)
15. Aptel, M., Claudon, L., Marsot, J.: Principles of handtool design. In: International Encyclopedia of Ergonomics and Human Factors, pp. 947–950 (2001)
16. Ullman, J., Kangas, N., Ullman, P., Wartenberg, F., Ericson, M.: A new approach to the mouse arm syndrome. Int. J. Occup. Saf. Ergon. **9**, 463–477 (2003)
17. Agarabi, M. Bonato P., De Luca C.J.: A sEMG-based method for assessing the design of computer mice. In: Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society, vol. 4, pp. 2450–2453, San Francisco (2004)
18. Odell, D.L., Johnson, P.W.: Evaluation of a mouse designed to improve posture and comfort. In: Proceedings of the 2007 Work with Computing Systems Conference-International Ergonomics Association (2007)
19. Chen, H.M., Lee, C.S., Cheng, C.H.: The weight of computer mouse affects the wrist motion and forearm muscle activity during fast operation speed task. Eur. J. Appl. Physiol. **112**(6), 2205–2212 (2012)

20. Rempel, D.M., Camilleri, M.J., Lee, D.L.: The design of hand gestures for human-computer interaction: lessons from sign language interpreters. Int. J. Hum Comput Stud. **72**(10–11), 728–735 (2014)
21. Jordan, P.: An Introduction to Usability. Taylor & Francis, London (1998)
22. Story, M.F.: The principles of universal design. In: Preiser, W.F.E., Smith, K.H. (eds.) Universal Design Handbook, 2nd edn, pp. 4.3–4.12. McGraw-Hill Inc., New York (2011)
23. Mueller, J.L.: Universal design for products. In: Null, R., Cherry, K. (eds.) Universal Design, pp. 104–113. Professional Publications, Belmont (1996)
24. Shimizu, Y., Kojima, T., Tano, M., Matsuda, S.: Models and Prototypes. Graphic Sha Pub. Co., Tokyo (1991)
25. Gordon, C.C., Blackwell, C.L., Bradtmiller, B., Parham, J.L., Barrientos, P., Paquette, S.P., Corner, B.D., Carson, J.M., Venezia, J.C., Rockwell, B.M., Mucher, M., Krietensen, S.: 2012 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics. Technical Report Natick/TR-15/007, United States Army Natick Soldier Research, Development and Engineering Center, Natick, Massachusetts 01760-2642, December (2014)

Author Index

I.L. Nunes (ed.), *Advances in Human Factors and Systems Interaction*,
Advances in Intelligent Systems and Computing 592, DOI 10.1007/978-3-319-60366-7