Advances in Intelligent Systems and Computing 959

Isabel L. Nunes *Editor*

Advances in Human Factors and Systems Interaction

Proceedings of the AHFE 2019 International Conference on Human Factors and Systems Interaction, July 24–28, 2019, Washington D.C., USA

Springer

Advances in Intelligent Systems and Computing

Volume 959

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 such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily 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.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

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 2019 International Conference on Human Factors and Systems Interaction, July 24–28, 2019, Washington D.C., USA

Springer

Editor
Isabel L. Nunes
Faculty of Science and Technology
Universidade NOVA de Lisboa
Caparica, Portugal

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-030-20039-8 ISBN 978-3-030-20040-4 (eBook)
https://doi.org/10.1007/978-3-030-20040-4

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

Advances in Human Factors and Ergonomics 2019

AHFE 2019 Series Editors

Tareq Ahram, Florida, USA
Waldemar Karwowski, Florida, USA

10th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences

Proceedings of the AHFE 2019 International Conference on Human Factors and Systems Interaction, held on July 24–28, 2019, in Washington D.C., USA

Advances in Affective and Pleasurable Design	Shuichi Fukuda
Advances in Neuroergonomics and Cognitive Engineering	Hasan Ayaz
Advances in Design for Inclusion	Giuseppe Di Bucchianico
Advances in Ergonomics in Design	Francisco Rebelo and Marcelo M. Soares
Advances in Human Error, Reliability, Resilience, and Performance	Ronald L. Boring
Advances in Human Factors and Ergonomics in Healthcare and Medical Devices	Nancy J. Lightner and Jay Kalra
Advances in Human Factors and Simulation	Daniel N. Cassenti
Advances in Human Factors and Systems Interaction	Isabel L. Nunes
Advances in Human Factors in Cybersecurity	Tareq Ahram and Waldemar Karwowski
Advances in Human Factors, Business Management and Leadership	Jussi Ilari Kantola and Salman Nazir
Advances in Human Factors in Robots and Unmanned Systems	Jessie Chen
Advances in Human Factors in Training, Education, and Learning Sciences	Waldemar Karwowski, Tareq Ahram and Salman Nazir
Advances in Human Factors of Transportation	Neville Stanton

(continued)

(continued)

Advances in Artificial Intelligence, Software and Systems Engineering	Tareq Ahram
Advances in Human Factors in Architecture, Sustainable Urban Planning and Infrastructure	Jerzy Charytonowicz and Christianne Falcão
Advances in Physical Ergonomics and Human Factors	Ravindra S. Goonetilleke and Waldemar Karwowski
Advances in Interdisciplinary Practice in Industrial Design	Cliff Sungsoo Shin
Advances in Safety Management and Human Factors	Pedro M. Arezes
Advances in Social and Occupational Ergonomics	Richard H. M. Goossens and Atsuo Murata
Advances in Manufacturing, Production Management and Process Control	Waldemar Karwowski, Stefan Trzcielinski and Beata Mrugalska
Advances in Usability and User Experience	Tareq Ahram and Christianne Falcão
Advances in Human Factors in Wearable Technologies and Game Design	Tareq Ahram
Advances in Human Factors in Communication of Design	Amic G. Ho
Advances in Additive Manufacturing, Modeling Systems and 3D Prototyping	Massimo Di Nicolantonio, Emilio Rossi and Thomas Alexander

Preface

Human Factors and Systems Interaction aims to address the main issues of concern within system interface with a particular emphasis on the system life-cycle development and implementation of interfaces and the general implications of 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 system 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 nine sections presented in this book are as follows:

Section 1 Management of Productivity by Using Assistance Systems
Section 2 Management of Productivity in Smart Manufacturing/Industry 4.0
Section 3 Human–Machine Interaction Applications
Section 4 Systems Interaction Design and Automation
Section 5 System Accessibility Design
Section 6 Security and Crisis Management
Section 7 Innovative Methods and Approaches for Systems Interaction Design
Section 8 Human–Computer Interaction in e-Health and Environmental Applications
Section 9 Advances in User Experience, Affordance and Technology

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

Francesco Biondi, UK
Nina Berry, USA
James P. Bliss, USA
Filipa Carvalho, Portugal
Denis Coelho, Portugal
Shuchisnigdha Deb, USA
Frank Flemisch, Germany
Kazuo Hatakeyama, Brazil
Sven Hinrichsen, Germany
Tim Jeske, Germany
Christopher Lowe, UK
Ravi Mahamuni, India
Pamela McCauley, USA
Jorge Luis Pérez Médina, Ecuador
Bonnie Novak, USA
Maria Papanikou, UK
Alexandra B. Proaps, USA
William Prugh, USA
Carla Quintao, Portugal
Adriano Bernardo Renzi, Brazil
Yves Rybarczyk, Ecuador
Michael W. Sawyer, USA
Mario Simoes-Marques, Portugal
Vesna Spasojevic Brkic, Serbia
Paolo Trucco, Italy

July 2019 Isabel L. Nunes

Contents

Systems Interaction Design and Automation

System Accessibility Design

Management of Productivity by Using Assistance Systems

Method for Measuring the Application Potential of Assembly Assistance Systems

Sven Bendzioch[1](✉), Sven Hinrichsen[1], Benjamin Adrian[1], and Manfred Bornewasser[2]

[1] Ostwestfalen-Lippe University of Applied Sciences and Arts, Campusallee 12, 32657 Lemgo, Germany
{Sven.Bendzioch, Sven.Hinrichsen, Benjamin.Adrian}@th-owl.de

[2] University of Greifswald, Franz-Mehring-Straße 47, 17489 Greifswald, Germany
Bornewas@uni-greifswald.de

Abstract. Because of the trend towards smaller batch sizes and customer-specific products, employees in assembly have to absorb and process more and more information. With the help of assistance systems, the process of information absorption and processing can be designed in a more economical and human-oriented way. Despite the potentials of such systems, companies pay too little attention to the informational design of assembly systems, resulting in a number of significant deficits in information management. To identify such deficits in operational practice, a questionnaire has been developed. This allows the potential use of informational assistance systems for an assembly system to be estimated. In this article, results from an exploratory factor analysis of the developed instrument will be presented and interpreted as a basis for the further development of the questionnaire.

Keywords: Complexity evaluation · Manual assembly · Questionnaire · Work analysis method · Assistance systems · Exploratory factor analysis

1 Introduction and Problem Statement

The development trend toward an ever-increasing number of parts and part variants, together with an increasing number of products, is leading to increasing complexity and thus to new challenges in assembly. In this context, automation solutions are often not economically implementable [1], resulting in a reliance on the motor and cognitive abilities of employees [2]. Smaller batches and a higher number of product variants demand a high degree of employee versatility [1] because continually different product variants have to be assembled. Assembly work is becoming more complex in this respect, since more and more information has to be processed, and a great number of decisions must be made by the employee. For employees in manual assembly, this development – more variants, more action alternatives, and more choice decision – is leading to greater cognitive demands and thus to a subjectively perceived increase in complexity [3]. These demands are accentuated in particular by deficits in operational

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 3–11, 2020.
https://doi.org/10.1007/978-3-030-20040-4_1

information management. Five deficit categories can be distinguished [4]: (1) Required information is missing in the assembly system. (2) Unnecessary information is displayed. (3) Information is provided at the wrong time and in the wrong amount. (4) Information is not current and/or (5) not prepared in such as way that the employee can absorb and process it easily.

In this context, innovative informational assistance systems are required that guide, support, and train employees in a situation-based way [5–7]. According to studies carried out by Kasselmann and Willecke [8] as well as Koch et al. [9], suitable strategies to introduce such systems are lacking. In addition, Fürntratt et al. [10] perceive a need for research to identify work systems suitable for supporting employees via assistance systems. In summary, it can be concluded that companies are increasingly facing the question of how assistance systems can be designed to conform with requirements and how the uses of such systems can be assessed. The approaches of Hold et al. [5] or Herder and Aurich [11], which are based on MTM studies, are already attempting to identify suitable work systems for the implementation of an assistance system. However, the application of methods can be considered costly due to the required MTM studies.

2 Objective and Procedure

As part of the "Montexas4.0" project sponsored by the German Federal Ministry of Education and Research, a method is being developed, with which existing manual assembly systems can be evaluated in a short time with respect to the application potentials of informational assistance systems. This will ultimately contribute to easing the burden on employees and increasing productivity. The method is based, on the one hand, on a questionnaire, with which the informational complexity of an assembly system as perceived by employees will be recorded. On the other hand, the method will objectively determine the economic potentials that can be exploited with selected informational assistance systems. This article discusses only the first part of the inventory of methods.

Initial findings regarding the developed instrument have already been presented in publications by Bendzioch et al. [7, 12, 13]. On this foundation, this article explores the results of an exploratory factor analysis in more detail [13]. With the help of factor analysis, the theoretical construct of the questionnaire will be reviewed with empirical data. The article is subdivided into three sections. The first section presents the theoretical principles of informational complexity. The second section discusses the results of an exploratory factor analysis. This is followed by an interpretation of the results in view of further developing the questionnaire. The final section looks ahead to further approaches in the development of the questionnaire.

3 Informational Complexity in the Context of the Questionnaire

Klabunde describes complexity using the characteristics of variety, connectivity, and dynamics [14]. Accordingly, a complex system can be defined by the type and number of elements (variety) and the relations existing between the elements (connectivity). The concept of dynamics represents the unpredictability and indeterminacy of system conditions [14]. Every increase in variety, connectivity, and dynamics or a combination of two or all three factors is considered an increase in complexity. One example of this is the current development towards smaller batch sizes and customer-specific products, which is leading to a general increase in complexity within the company. For employees in manual assembly, this development means mastering a large amount of action alternatives and uncertainty. The employee must absorb order information, select components and tools, and carry out assembly processes in accordance with specified assembly guidelines. Information must therefore be processed continuously. The sequence of work steps, and thus the implementation of information, can vary. For the most part, orders differ by batch sizes or product variants, thus calling for a varying number of parts to be installed or tools to be employed. Complexity cannot be increased arbitrarily without leading to deficiencies in product quality or productivity. These deficiencies can then be primarily attributed to increased mental strain on employees.

To reduce mental workloads and enhance the performance of employees, the informational complexity of the activity to be performed must be reduced. Following Klabunde [14], informational complexity can be defined by the type and number of pieces of information provided for completing a task and their interdependencies. For example, parts lists can contain information that differs from the corresponding design drawing. In addition, informational complexity results from the dynamics of changes required in the assembly system for the manufacture of new product variants. Therefore, informational complexity is especially triggered by a lack or flood of information [15]. As a result, a poorly prepared information management system can increase informational complexity and lead, for example, to task interruptions, search processes, consultations with designers, or reworking. It must be noted that, from a human-oriented perspective, informational complexity may neither be to low (lack of challenge) nor increased arbitrarily, since the human being requires intermittent periods of redundancy, repetition, and stability. If the complexity in a work system increases, employees perceive this through the sum of cognitive processes of information absorption and processing, selection and decision processes, and resulting experiences of excessive strain and uncertainty [13]. With regard to the evaluation of informational complexity, the extent of uncertainty and unpredictability must be determined in terms of information provision and implementation in particular activities. This objective is also pursued with the questionnaire currently under development.

The design of the questionnaire is based on the theoretical principles of complexity discussed above, individual work analysis processes (questionnaires such as CXI, ISTA, KOMPASS, and TBS-S), expert workshops, and a case study from a cooperating company. As part of the expert workshop, five dimensions were deduced on the basis of the theoretical principles and the case study. To operationalize these dimensions, 25 items could be generated taking the characteristics defined by Klabunde – variety, connectivity, and dynamics – into account and with the help of individual work analysis processes. The items are processed by means of a five-point Likert scale. In the following, the five deduced dimensions (A to E) will be presented with their thematic emphases [13].

- *A. Large number of components and assembly groups:* Major variant numbers and changes lead to performance uncertainty among employees.
- *B. Work execution and conditions:* Work pressure and distraction impair concentration and alertness and hinder decision processes.
- *C. Scope of assembly information:* The number of decisions to be made rises with an increasing scope of information provided.
- *D. Inadequate qualification:* Inadequate qualification, insufficient training time, and a lack of continued education lead to performance uncertainty.
- *E. High quality standards:* The accuracy with which information must be taken in and implemented.

4 Exploratory Factor Analysis

4.1 Principles of Exploratory Factor Analysis

Exploratory factor analysis is used when a newly developed measuring instrument needs to be checked in terms of its underlying theoretical structures. It focuses on evaluating the items generated during the scale construction, which can be grouped into one or more latent factors [16]. This makes it a structure-discovering method, which furthers the formation of hypotheses [16]. The first step of exploratory factor analysis is to analyze the correlation matrix. In it, the interdependencies of the particular indicators become quantifiable by means of correlations [17]. A high correlation of individual variables (items) – interpreted as similarities in content among various items – forms the foundation and condition for discovering and describing a latent factor. The factor loading reflects the relationship between the indicator variables (items) and the latent factor [17]. According to Cohen [18], loadings can be interpreted as follows: 0.10 is equivalent to a weak loading, 0.30 to a medium loading, and 0.50 to a strong loading. In the second step, the number of underlying factors is determined with the help of the eigenvalues (total variance of all variables that are explained by one factor) using the Kaiser-Guttman criterion [19]. For further analysis, only those factors are relevant, whose eigenvalue is significantly higher than one and thus explain more variance than the associated items. With the help of the factor number now determined, the factors can be interpreted in a final step. To do so, a rotated factor loading matrix is examined. In it, each item is assigned to the factor on which it loads the highest. The minimal acceptance for factor loadings is between ± 0.30 and ± 0.40. For a factor to be interpretable, at least four variables must have a loading of ± 0.60 or at least ten variables a loading of ± 0.40 [16]. Exploratory factor analysis thus provides information about whether the theoretical presuppositions about a construct – its informational complexity for instance – can be confirmed by empirically collected data.

4.2 Scope of the Study and Results of the Exploratory Factor Analysis

The following section discusses both the scope of the study and the results of the exploratory factor analysis [13]. The questionnaire conceived is used for the purpose of data generation in two companies. These include a manufacturer of customer-specific truck superstructures and a producer of highly varied series machines. For the survey, different work systems were considered in both companies, in each of which complex products are assembled manually. A total of 61 employees could be surveyed in the area of manual assembly for data collection. At the time of the survey, 85% of the persons surveyed were employees on permanent contracts and 63% worked permanently in the relevant work system. The results of the exploratory factor analysis will now be presented.

The intercorrelation matrix (Table 1) shows that the items correlate predominately at the level of moderate to medium-strong. The items of the indexed dimensions B, C, and E exhibit a closer relationship than the less pronounced relationships of items forming dimensions A and D. The items assigned to A do not only correlate among each other, but also with many other variables with similar strength. On the other hand, items in dimension D exhibit few correlations. In general, it can be inferred from the intercorrelation matrix that a higher level of selectivity between the individual dimensions, and thus between the items as well, must be achieved. In contrast to the preliminary theoretical considerations concerning the characteristics structure, the Kaiser-Guttman criterion used to determine the number of latent factors indicated eight factors, which explain a total of 72% of the total variance.

Table 1. Intercorrelation matrix of the exploratory factor analysis (n = 61) [13].

	A1	A2	A3	A4	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4	C5	C6	D1	D2	D3	D4	E1	E2	E3	E4	E5
A1	–	,08	,31	,30	-,12	-,13	,26	,20	,07	,33	,24	,06	-,05	,02	-,11	,03	,16	,23	,17	,06	,22	,20	,39	,20	,23
A2	,08	–	,29	,59	-,20	-,10	,16	,00	,14	,12	,12	,31	,25	,35	,39	,11	-,25	,06	,13	,34	,09	,42	,20	,05	,02
A3	,31	,29	–	,48	-,07	,03	,00	,18	,18	-,02	,15	,34	,19	,21	,26	-,03	,09	,36	,27	,31	,40	,30	,43	,38	,27
A4	,30	,59	,48	–	-,08	-,10	,22	,21	,22	,18	,30	,37	,29	,37	,31	,32	,24	,29	,38	,15	,19	,36	,30	,19	,28
B1	-,12	-,20	-,07	-,08	–	,22	-,17	,00	,09	,02	,09	,04	-,04	,12	,16	-,04	,19	,29	,03	,09	,19	-,08	,08	-,06	-,15
B2	-,13	-,10	,03	-,10	,22	–	,34	-,04	,27	,03	,13	-,18	-,08	-,15	,33	-,23	-,17	,05	,12	,02	-,16	,19	-,07	,03	-,14
B3	,26	,16	,00	,22	-,17	,34	–	,21	,46	,36	,26	-,10	-,10	-,36	,13	-,01	,07	,12	,22	,12	-,30	,16	-,17	-,11	-,15
B4	,20	,00	,18	,21	,00	-,04	,21	–	,52	,48	,32	,08	,38	,18	,24	,36	,26	,09	,01	,13	,26	,36	,37	,07	,13
B5	,07	,14	,18	,22	,09	,27	,46	,52	–	,43	,44	,16	,20	,06	,39	,12	,10	,17	,07	,18	-,05	,42	,30	-,10	-,13
B6	,33	,12	-,02	,18	,02	,03	,36	,48	,43	–	,33	-,04	,31	,04	,14	,11	,25	-,01	,24	,02	,05	,39	,29	-,05	,13
C1	,24	,12	,15	,30	,09	,13	,26	,32	,44	,33	–	,25	,10	,26	,46	,29	,13	,36	,10	,02	,04	,36	,29	-,13	,07
C2	,06	,31	,34	,37	,04	-,18	-,10	,08	,16	-,04	,25	–	,34	,57	,32	,21	,08	,20	-,08	,23	,22	,13	,15	,15	,17
C3	-,05	,25	,19	,29	-,04	-,08	-,10	,38	,20	,31	,10	,34	–	,49	,29	,33	-,01	,14	-,08	,16	,05	,15	,24	-,09	,14
C4	,02	,35	,21	,37	,12	-,15	-,36	,18	,06	,04	,26	,57	,49	–	,34	,43	,06	,13	-,03	,21	,21	,33	,42	,08	,30
C5	-,11	,39	,26	,31	,16	,33	,13	,24	,39	,14	,46	,32	,29	,34	–	,03	-,04	,18	,00	,19	-,09	,39	,32	,00	,00
C6	,03	,11	-,03	,32	-,04	-,23	-,01	,36	,12	,11	,29	,21	,33	,43	,03	–	,04	,07	-,11	,09	,24	,20	,23	-,01	,31
D1	,16	-,25	,09	,24	,19	-,17	,07	,26	,10	,25	,13	,08	-,01	,06	-,04	,04	–	-,06	,35	-,10	,05	-,06	,00	,06	,07
D2	,23	,06	,36	,29	,29	,05	,12	,09	,17	-,01	,36	,20	,14	,13	,18	,07	-,06	–	,27	,38	,21	,10	,30	-,03	,18
D3	,17	,13	,27	,38	,03	,12	,22	,01	,07	,24	,10	-,08	-,08	-,03	,00	-,11	,35	,27	–	,02	,21	,27	,13	,36	,16
D4	,06	,34	,31	,15	,09	,02	,12	,13	,18	,02	,02	,23	,16	,21	,19	,09	-,10	,38	,02	–	,13	,10	,16	-,11	-,01
E1	,22	,09	,40	,19	,19	-,16	-,30	,26	-,05	,05	,04	,22	,05	,21	-,09	,24	,05	,21	,21	,13	–	,16	,46	,67	,41
E2	,20	,42	,30	,36	-,08	,19	,16	,36	,42	,39	,36	,13	,15	,33	,39	,20	-,06	,10	,27	,10	,16	–	,61	,28	,29
E3	,39	,20	,43	,30	,08	-,07	-,17	,37	,30	,29	,29	,15	,24	,42	,32	,23	,00	,30	,13	,16	,46	,61	–	,35	,47
E4	,20	,05	,38	,19	-,06	,03	-,11	,07	-,10	-,05	-,13	,15	-,09	,08	,00	-,01	,06	-,03	,36	-,11	,67	,28	,35	–	,53
E5	,23	,02	,27	,28	-,15	-,14	-,15	,13	-,13	,13	,07	,17	,14	,30	,00	,31	,07	,18	,16	-,01	,41	,29	,47	,53	–

Legend: ≥ ±0,30 < ±0,40 (light grey) ≥ ±0,40 < ±0,50 (dark grey) ≥ ±0,50 (black) Dimension A to E (dashed box)

Table 2 below shows the resulting factor loadings. The loadings of the items assigned to B, C, and E are predominately satisfactory, i.e. the assumed characteristics of the work performance and conditions, the scope of assembly information, and the high quality standards are confirmed as expected. However, this is not true of the items in dimension A (large number of components and assembly groups) and D (inadequate qualification), which load relatively high on various factors against expectations. The result thus reflects the findings from the correlation analysis.

Table 2. Rotated factor loading matrix (n = 61; main component analysis) [13].

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
A1	,318	,190	,220	,255	-,179	,200	,225	-,518
A2	,031	-,445	-,016	-,248	-,334	,098	,630	,037
A3	,455	-,189	,011	-,130	-,524	,305	,131	-,050
A4	,158	-,486	,102	-,055	-,199	,515	,363	-,270
B1	-,012	-,080	-,019	-,227	-,145	,090	-,803	-,047
B2	-,058	,290	,028	-,768	-,018	,005	-,138	-,002
B3	-,334	,348	,373	-,242	-,151	,294	,359	-,317
B4	,175	-,148	,838	,032	-,070	-,004	-,085	-,043
B5	-,114	-,057	,662	-,424	-,180	,059	-,009	-,170
B6	,041	,029	,770	-,048	,082	,197	,105	-,134
C1	-,039	-,272	,277	-,296	,066	,075	-,073	-,729
C2	,073	-,708	-,071	-,007	-,216	,130	,008	-,085
C3	-,006	-,600	,429	,040	-,160	-,109	,052	,176
C4	,234	-,844	,045	-,045	-,014	-,068	-,048	-,073
C5	-,033	-,458	,154	-,697	-,101	,019	,005	-,112
C6	,131	-,490	,297	,277	,132	-,200	,042	-,312
D1	-,050	-,110	,277	,258	,204	,728	-,337	-,017
D2	,127	-,079	-,066	-,071	-,604	,077	-,248	-,567
D3	,263	,139	,020	-,151	-,079	,748	,064	-,074
D4	-,041	-,157	,114	-,016	-,822	-,090	,024	,022
E1	,777	-,106	,066	,170	-,226	,076	-,208	,050
E2	,438	-,186	,380	-,492	,038	,018	,319	-,177
E3	,683	-,227	,328	-,160	-,143	-,123	-,006	-,256
E4	,831	,057	-,094	-,086	,060	,267	,071	,177
E5	,702	-,184	,008	,146	,129	,024	,121	-,218

□ ≥ ±0,40 < ±0,50 ■ ≥ ±0,50

4.3 Interpretation of Results

The results presented above of the exploratory factor analysis show that the current state of development of the questionnaire does not yet represent a satisfactory result. The manifest variables included (items) do not yet depict the construct of informational complexity with sufficient accuracy, also with respect to the latent structure of the factors. Individual items must be rectified and new items must be integrated into the development of the questionnaire. The theoretically assumed dimensions must be altered or supplemented in this context in order to better operationalize the concept of informational complexity. Regarding this, the results for dimensions A to E are interpreted below with a view to further development:

- *Dimension A:* The results of the factor loading matrix make it clear that the variables load on different latent factors. The theoretical construct of performance uncertainty as a result of an increased number of variants and changes can to that effect not be completely quantified. The low level of selectivity with other latent variables clarify this issue. These problems suggest that the underlying dimension needs to be specified in more detail. The theoretical construct should be reviewed and altered for that purpose. The focus is on the informational effects that arise from an increased number of variants and changes among employees.

- *Dimension B:* On the one hand the results show a close relationship between the items B4, B5, and B6, especially in the case of the factor loading matrix. This partially confirms the assumed model for dimension B. On the other hand, B1 and B3 exhibit no mentionable loading on a factor and thus appear to be irrelevant for the further development of the dimension. The same applies for item B2, which loads on one other factor. Items B1, B2, and B3 are formulated too generally or do not cover the construct with sufficient accuracy. Referring to the thematic emphasis of the dimension, increasingly difficult selection and decision processes must in particular be better depicted by the items.
- *Dimension C:* The construct of assembly information mainly demonstrates weak points in selectivity with other dimensions (see correlation matrix) and thus can be confirmed to the largest extent. Items C1 and C5 deserve special mention in this regard. These have in common that they load on other latent factors (see factor loading matrix). Furthermore, the results reveal high correlations and loadings among items that are concentrated on the procurement of information. This finding should be taken into account in the generation of new items or the alteration of items. The underlying construct of dimension C must also be modified for the procurement of information requirements for this purpose.
- *Dimension D:* The results of the factor loading matrix shows a result similar to that of dimension A. The variables also load on different latent factors. The correlation matrix also demonstrates that the variables correlate significantly neither among one another nor with other variables. The results show that the theoretical construct of inadequate qualification cannot be directly related with informational complexity. In conjunction with the objective of the questionnaire and the low statistical relevance of the latent factor, it suggests that the dimension should no longer be taken into account in the further development of the questionnaire.
- *Dimension E:* The factor loading matrix represents a nearly homogeneous result, which suggests a close relationship between the latent variables and the theoretical construct of the dimension. Nonetheless, external loadings also stand out in the correlation matrix. These must be minimized by adapting individual questions. In this respect, the item formulation must focus more clearly on the accuracy, with which information is taken in and implemented.

5 Outlook

The results of the exploratory factor analysis and their interpretation show that the construct of informational complexity is not yet sufficiently represented with the current questionnaire. In this respect, further studies are necessary so that the true core dimensions of the construct are taken into account in the ascertainment of the informational complexity of an assembly system. Based on this, the questionnaire will be further developed in four steps. The first step will be to re-examine the construct of the questionnaire for the acquisition of all relevant dimensions of informational complexity in the context of an expert workshop. Concretely, individual dimensions must be altered or excluded from further development on the basis of the results and theoretical principles. Building on the revision of the theoretical construct, the second step will be to improve existing items and to integrate further new items into the analysis. The conclusion of the second work step will result in a further developed questionnaire. The third step will comprise the conducting of cognitive interviews with employees from manual assembly as well as their supervisors. Assembly systems should be better depicted this way, and the language of employees better represented in the items of the questionnaire. In a fourth step, a further survey with the improved questionnaire will be undertaken and a further factor analysis carried out. The intermediate steps described above for the further development of the questionnaire form the prerequisite for a subsequent examination of the construct and criterion validity of the questionnaire. Operational users will thus in the near future have the opportunity to (subjectively and objectively) evaluate existing assembly systems on the basis of a valid questionnaire and determine whether potential exists for the use of informational assistance systems.

Acknowledgments. This publication was created as part of the project "Montextas4.0 - Excellence in Assembly" funded by the German Ministry of Education and Research (BMBF) and the European Social Fund (ESF) (Grant no. 02L15A260).

References

1. Bächler, A., Bächler, L., Autenrieth, S., Kurtz, P., Heidenreich, T., Hörz, T., Krüll, G.: Entwicklung von Assistenzsystemen für manuelle Industrieprozesse. In: Rathmayer, S., Pongratz, H. (eds.) Proceedings of DeLFI Workshops 2015, pp. 56–63. CEUR-Workshop, München (2015)
2. Zäh, M.F., Wiesbeck, M., Engstler, F., Friesdorf, F., Schubö, A., Stork, S., Bannat, A., Wallhoff, F.: Kognitive Assistenzsysteme in der manuellen Montage. In: wt Werkstattstechnik online, vol. 97, no. 9, pp. 644–650. Springer-VDI (2007)
3. Bornewasser, M., Bläsing, D., Hinrichsen, S.: Informatorische Assistenzsysteme in der manuellen Montage: Ein nützliches Werkzeug zur Reduktion mentaler Beanspruchung? Zeitschrift für Arbeitswissenschaft **72**(4), 264–275 (2018)
4. Hinrichsen, S., Bendzioch, S.: How digital assistance systems improve work productivity in assembly. In: Nunes, I. (ed.): Advances in Human Factors and Systems Interaction, AHFE 2018. Advances in Intelligent Systems and Computing, vol. 781, pp. 332–342. Springer, Cham (2018)
5. Hold, P., Ranz, F., Sihn, W.: Konzeption eines MTM-basierten Bewertungsmodells für digitalen Assistenzbedarf in der cyberphysischen Montage. In: Schlick, C.M. (ed.) Megatrend Digitalisierung Potenziale der Arbeits- und Betriebsorganisation, pp. 295–322. GITO mbH Verlag, Berlin (2016)
6. Fast-Berglund, A., Fässberg, T., Hellman, F., Davidsson, A., Stahre, J.: Relations between complexity, quality and cognitive automation in mixed-model assembly. J. Manuf. Syst. **32** (3), 449–455 (2013)

7. Bendzioch, S., Hinrichsen, S., Winter, M., Adrian, B., Bornewasser, M.: Wirtschaftliche und humanorientierte Einsatzpotenziale von Montageassistenzsystemen. In: Bericht zum 64. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft, Frankfurt, Germany, 21–23 February 2018. GfA-Press, Dortmund, B.5.2 (2018)
8. Kasselmann, S., Willeke, S.: Technologie Kompendium: Interaktive Assistenzsysteme. http://www.ipri-institute.com/fileadmin/pics/Projekt-Seiten/40ready/Technologie-Kompendium.pdf. Accessed 20 Dec 2018
9. Koch, V., Kuge, S., Geissbauer, R., Schrauf, S.: Industrie 4.0: Chancen und Herausforderungen der vierten industriellen Revolution. In: PwC (2014)
10. Fürntratt, H., Murg, S., Zeiner, H.: DeepLearning basiertes Unterstützungssystem für die Produktion. In: Weyers, B., Dittmar, A. (eds.) Mensch und Computer – Workshopbeiträge, 4–7 September 2016. Veröffentlicht durch die Gesellschaft für Informatik e.V., Aachen (2016)
11. Herder, C., Aurich, J.C.: Bewertungskonzept zur Identifikation von kognitiven Unterstützungstechnologien in der manuellen Montage. In: Weidner, R. (ed.) Zweite Transdisziplinäre Konferenz – Technische Unterstützungssysteme, die die Menschen wirklich wollen, pp. 41–48. Helmut-Schmidt-Universität, Hamburg (2016)
12. Bendzioch, S., Hinrichsen, S., Adrian, B., Bornewasser, M.: Assessing the economic and human-centered potential of assembly assistance systems. In: Villmer, F.J., Padoano, E. (eds.) Production Engineering and Management: Proceedings 8th International Conference, Lemgo, Germany, 04–05 Oktober 2018, pp. 127–136. Publication Series in Direct Digital Manufacturing (2018)
13. Bendzioch, S., Bornewasser, M., Hinrichsen, S., Adrian, B.: Messung der informatorischen Komplexität zur Abschätzung der Einsatzpotenziale von Montageassistenzsystemen. In: Bericht zum 65. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft, Dresden, Germany, 27 February–01 March 2019. GfA-Press, Dortmund, C.9.1 (2019)
14. Klabunde, S.: Wissensmanagement in der integrierten Produkt und Prozessgestaltung – Best-Practice-Modelle zum Management von Meta-Wissen. Deutscher Universitäts-Verlag, Wiesbaden (2003)
15. Mayer, A.: Modularisierung der Logistik: Ein Gestaltungsmodell zum Management von Komplexität in der industriellen Logistik. In: Stube, F., Baumgarten, H., Klinkner, R. (eds.) Schriftenreihe Logistik der Technischen Universität Berlin, Band 1. Universitätsverlag der TU Berlin, Magdeburg (2007)
16. Schwarz, J., Bruderer-Enzler, H., Keller, M., De Simoni, C., Seidmann, S., Westphalen, A.: Methodenberatung. Universität Zürich. https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/interdependenz/reduktion/faktor.html. Accessed 08 Dec 2018
17. Bortz, J., Döring, N.: Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, 5th edn. Springer, Berlin (2016)
18. Cohen, J.: A power primer. Quantitative methods in psychology. Psychol. Bull. **112**, 155–159 (1992)
19. Moosbrugger, H., Kelava, A.: Testtheorie und Fragebogenkonstruktion, 2nd edn. Springer, Berlin (2012)

Automation Configuration Evaluation in Adaptive Assembly Systems Based on Worker Satisfaction and Costs

Peter Burggräf[1,2(✉)], Matthias Dannapfel[1], Tobias Adlon[1], Aaron Riegauf[3], and Jessica Schmied[3]

[1] Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany
{P.Burggraef,M.Dannapfel,T.Adlon}@WZL.RWTH-Aachen.de

[2] Chair of International Production Engineering and Management of University of Siegen, Paul-Bonatz-Straße 9 - 11, 57076 Siegen, Germany
Peter.Burggraef@Uni-Siegen.de

[3] Chair of Production Engineering of E-Mobility Components (PEM) of RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany
{A.Riegauf,J.Schmied}@PEM.RWTH-Aachen.de

Abstract. Nowadays requirements and challenges for production systems have led to an increasing importance of flexible assembly processes. Arising technologies in the context of Industry 4.0 provide previously unknown possibilities for human-oriented automation. To exploit those possibilities effectively, the research project A4BLUE aims to develop and evaluate sustainable and adaptive workplaces that are human-centered and best automated for dynamic environments. As part of the research project, a methodology is developed to determine the optimal automation configuration in relation to context and situation. The present work provides an overview of the methodological approach, implemented as a software prototype within the A4BLUE Adaptive Framework. The methodology takes into account the aspects of worker satisfaction as well as economic factors by adding a systematic, process-related logic to validated approaches. In the present paper, the cost-based module of the methodology is outlined based on process characteristics and manufacturing targets of assembly systems.

Keywords: Assembly · Production planning · Adaptive automation · Cost evaluation · Level of automation · Worker satisfaction

1 Introduction

Manufacturing companies are facing increasing globalization and competition resulting in shorter product lifecycles as well as individualized products [1]. This leads to increasing demands on product quality and functionality as well as higher product variants [2]. These challenges lead to the demand of a high degree of flexibility as well as the reduction of product and production costs [3].

The planning of efficient as well as flexible assembly systems is fundamental, as assembly is a main cost driver in production. Several examples in history have shown

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 12–23, 2020.
https://doi.org/10.1007/978-3-030-20040-4_2

that high automation is not necessarily the solution but may result in poor system performance [4–6].

In order to realize an efficient yet flexible assembly system, the focus needs to lie on the interaction of human and technology instead of purely increasing the automation [7]. Although the factories of the future will be more automated, humans will continue to play an important role. Human-machine interaction provides a way to combine the advantages of both manual and automated work. The efficiency of robots and innovative information technology can support the flexibility of human workers.

Such technical improvements enable new automation configurations involving human workers and therefore require considering two aspects. On the one hand, worker satisfaction needs to be taken into account for sustainable efficiency in human-machine cooperation [8]. On the other hand, costs are a highly influential decision criterion and an essential factor in industrial production planning [9]. Quantifiable measures are needed to achieve an optimum automation configuration. Therefore, a new methodology consisting of different planning modules is presented in this paper.

The methodology is developed as part of the research project A4BLUE (Adaptive Automation in Assembly For BLUE collar workers satisfaction in Evolvable context). A4BLUE aims at the development and evaluation of a new generation of sustainable, adaptive and smart workplaces dealing with evolving requirements of manufacturing processes. For this purpose, automation configurations that are suitable for flexible and efficient task execution in interaction with human workers are introduced. Focus is the optimization of human variability through personalized and context aware assistance capabilities as well as advanced human-machine interfaces.

2 Automation Configuration Evaluation in the A4BLUE Adaptive Framework

The methodology for evaluating optimal automation configurations is implemented in a software application being part of the A4BLUE Adaptive Framework. This modular software framework is designed to provide an open, secure, configurable, scalable and interoperable adaptation management and assistance toolkit. Its components enable the user and the system to adjust certain workplace characteristics according to changing requirements. Within the framework's architecture three high-level functional domains are determined (cf. Fig. 1). The domains define a coarse mapping of system elements to either the Shop Floor, the Enterprise or the Business Domain.

The Shop Floor Domain is the bottom layer of the framework, building the connection to the physical world. It is populated by any kind of device that is connected to the digital world on one side and to the real world on the other side. Consequently, the Shop Floor Domain includes components supporting automated control and adaptation of physical assembly processes, mechanisms and devices.

The middle layer of the framework forms the core part of any A4BLUE-based system. It is represented by the Enterprise Domain. To this end, the domain is populated by several components in charge of managing the logic for adaption management, using an event-driven architecture to provide the assistance services.

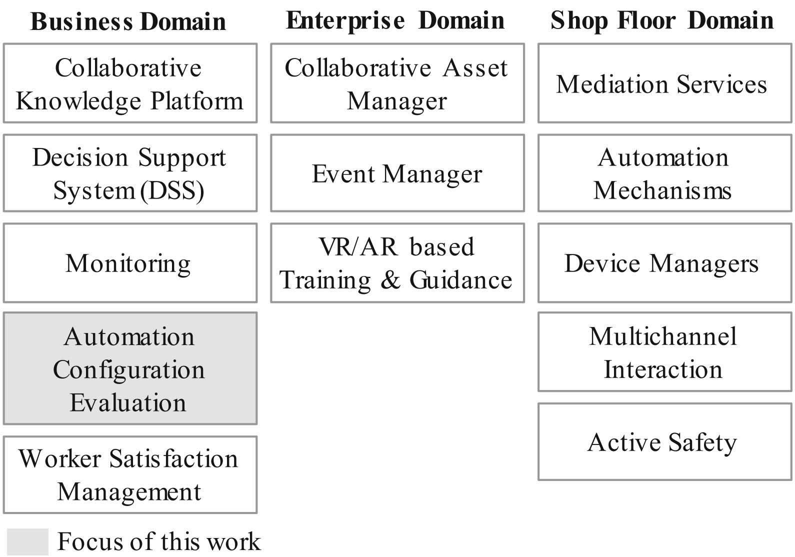


Fig. 1. Domains and components of the A4BLUE adaptive framework

The Business Domain is the upper layer, in charge of supporting the strategic decision-making by connecting white- and blue-collar workers. Components in this domain aim to take benefit from an improved knowledge of the running automation and adaptation processes.

As a part of the A4BLUE Adaptive Framework the optimal automation configuration is determined by means of the component Automation Configuration Evaluation, which is integrated within the Business Domain. All information gathered from shop floor processes shall be used to improve business decisions.

The Automation Configuration Evaluation component can use any data collected and processed in the framework. This provides the opportunity to trigger events for other components in order to adapt the assembly system based on automation decisons.

3 Socio-Economical Evaluation of Optimal Automation Configuration

Different approaches from various research areas with keywords such as cost estimating models, early cost estimation, cost predicting, level of automation decisions, assembly cost, design to cost as well as others have been developed in order to identify the optimal automation configuration. The most relevant methodologies are briefly mentioned in this section since these approaches form the theoretical basis of the presented methodology and will partly be integrated.

The methodologies by Gorlach and Wessel [9] as well as Boothroyd [10] provide basic calculations for the economic efficiency of assembly systems. They are appropriate regarding calculations towards economic optimization of automation configurations. The core idea of these approaches is fundamental for the requirements of an economic automation configuration.

Other methodologies are less focused on costs and automation; instead, they provide an overview of the underlying process tasks in assembly planning through the application context. For this reason, they make an important contribution as an application-oriented, theoretical basis for the creation of the description model. Lotter's [11] methodology however, rather focuses on identifying unfavorable, secondary processes. The approach provides methodological guidance for the planning and improvement of new and already existing assembly systems by an analysis of the determined processes. The aim of Ross' [12] work is to provide a methodology to the assembly planner for determining the economic automation configuration by comparing the currently considered process with already implemented ones based on a threshold value.

The listed methodologies show diverse approaches for optimizing the automation configuration. However, especially in times of rapidly increasing automation, the attention is to be put to the worker, as humans still remain one of the most important resources in assembly [13]. Especially from a socio-economical point of view, worker satisfaction constitutes a crucial factor [8]. Worker satisfaction or job satisfaction is said to be the appraisal of worker's job or job experience and determined by how employees feel about the job [14]. A methodology to optimize the automation configuration has to take into account both economic as well as human factors. Therefore, a modular approach was developed to combine these partly contrasting views.

3.1 Methodological Requirements

As a prerequisite for the targeted and scientific development of the methodology, it is necessary to define the requirements. These are subdivided into formal and conceptual as well as content-related requirements. While the formal and conceptual requirements are based on quality criteria of quantitative research, the specific content-related demands result from the objective of the application.

Three criteria for assessing the quality of research are objectivity, reliability and validity [15, 16]. With regard to this methodology, objectivity requires that the modules have to be designed in such a way that several operators choose the same characteristics when determining an automation configuration. Furthermore, the reliability of an automation configuration as a measuring instrument is to be proven. Thus, a reliable methodology is free of random errors, so that the same measurement result will be reproduced after repeating the application of the methodology under the same conditions [17]. Furthermore, the adequacy and validity of a theoretical approach with reference to the problem needs to be checked with the validity requirement [16]. Therefore, the methodology is valid if the application leads to correct and substantiated results for various manufacturers with different framework conditions.

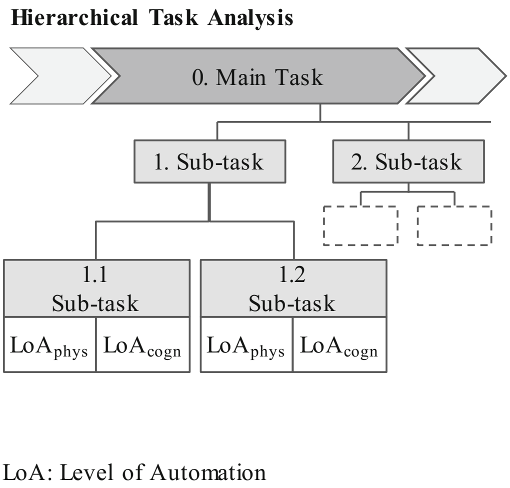
Moreover, the methodology has to be flexible and applicable to various situations, to newly planned as well as existing assembly processes. Decisions made in the context of assembly system planning and automation are occasionally inadequately documented. In these cases, it is not possible to trace or use the experience gained for subsequent planning. However, this experience is an essential basis for successful automation evaluations. [18] The methodology requires good documentation in order to draw on previous experience and especially a possibility to validate a decision made.

To ensure the fulfilment of the listed requirements, the methodology is validated by means of a software prototype before a functional software tool will be implemented within the A4BLUE Adaptive Framework.

3.2 Modular Methodology in Consideration of Worker Satisfaction and Costs

Based on the requirements and the aforementioned existing approaches different modules have been identified. The methodology consists of six modules that are briefly described in this section (cf. Fig. 4).

Within the first module, named Core Module, the optimization scope is defined before the main assembly tasks are analyzed in more detail. Firstly, it is determined which tasks relate to the product structure and design in the sense of the assembly process. This step serves the identification of every single assembly task to reveal potential automation configurations. On the basis of the product components, delivery condition, joining directions and processes, handling properties and quality requirements, the essential assembly tasks are identified. For further detailing and complexity reduction, the main assembly tasks are broken down using a method called Hierarchical Task Analysis (HTA, cf. Fig. 2). The HTA is a flexible methodology [19] that can be used in a variety of use cases aiming to reduce process complexity. This way processes can be decomposed to any level of detail. In the proposed methodology, the tasks are broken down to a level, where only a single resource can be assigned at a time (either a human or an automated piece of equipment).



Levels of Automation

Level	Physical	Cognitive
1	Totally manual	Totally manual
2	Static Hand Tool	Decision giving
3	Flex. Hand Tool	Teaching
4	Autom. Hand Tool	Questioning
5	Static Work Station	Supervising
6	Flex work station	Intervene
7	Totally automatic	Totally automatic

Fig. 2. Process decomposition with HTA and Levels of Automation acc. to DYNAMO [19, 20]

To analyze the resources in use, in a second step, the automation configuration is evaluated for every task on the lowest hierarchy level applying the Dynamo++ methodology [20]. Based on a predefined classification, each task can be rated on a scale from 1–7 regarding its Level of Automation (LoA), beginning with manual processes (1) up to a fully automated task execution (7). Additionally, a distinction between the physical and the cognitive part of a task is made (cf. Fig. 2). Physical automation in the context of this work represents the replacement of physical tasks of a human worker by automation, e.g. the lifting of a work piece by a robot. The substitution of cognitive functions, such as human sensory processes and mental activity by computerization is defined as cognitive automation. A process that is conducted completely manually is classified on the lowest level. When differentiating the cognitive and the physical part of a task, the results can be depicted in a 7 $\times$ 7 matrix [21].

The HTA is performed to evaluate the automation suitability for each task separately. As an output of the core module, the main assembly tasks are divided into various lowest level tasks with the related physical and cognitive LoA for each sub-task.

Within the second module of the methodology, the Constraint Module, the possible Levels of Automation of the assembly tasks are evaluated. To determine which tasks are suitable for automation in general, the theoretical automation expenditure for a single task is evaluated from a technical and economical point of view. Therefore, a catalogue containing a set of criteria for the evaluation of the implementation effort and process characteristics is provided (cf. Fig. 3). As a result, specific sub-tasks might be identified as not automatable with reasonable expenses and thus do not need to be considered in further steps of the methodology. The evaluated automation configuration restrictions lead to a reduced set of tasks and hence form the constraints for determining the optimal LoA.

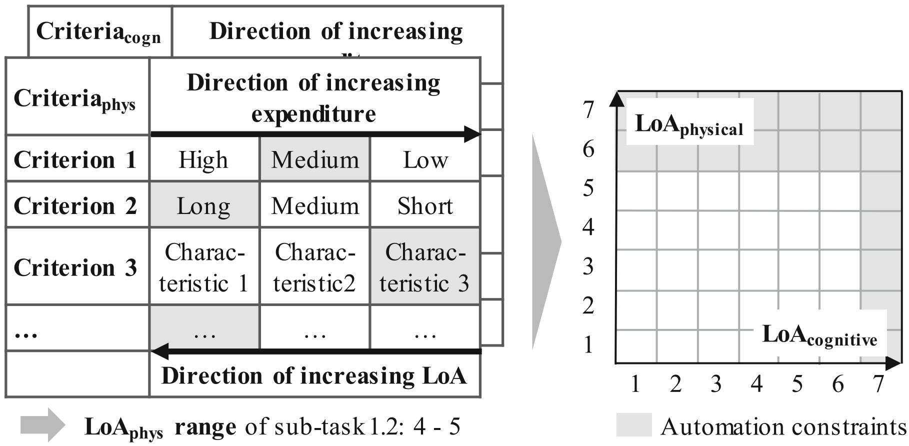


Fig. 3. Exemplary criteria catalogue for determining LoA ranges and constraints

The Cost-based Module constitutes the third module and focuses on the economic aspects of optimal automation configuration evaluations. On this account, automatic and non-automatic resources are identified and assigned to tasks, in order to create different but comparable scenarios for evaluation.

Besides economic factors, social aspects need to be taken into account. Hence, a model for an overall expected measure of satisfaction is applied in the Worker Satisfaction Module. The model has been derived from empirical data of user satisfaction levels measured across key technologies. It is applied for each scenario with regard to the different resources used.

The optimal overall LoA is then determined for each scenario as a combination of the results of the Cost-based and the Worker Satisfaction Module. As an output of the Proposed Automation Module, a new LoA is calculated to quantify the resulting automation solution.

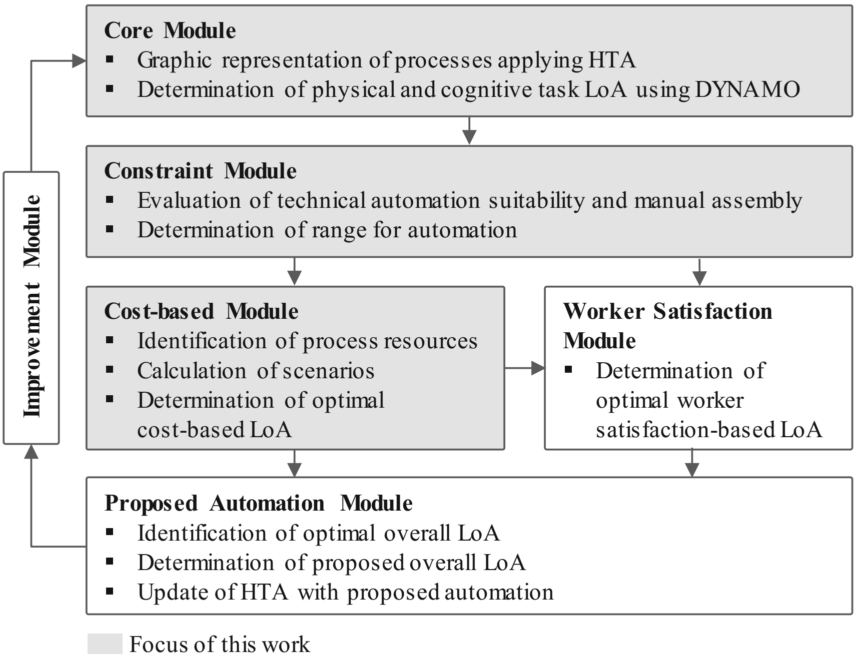


Fig. 4. Modular approach for determining the optimal Level of Automation

The Improvement Module is the sixth module of the methodology and aims to validate and maintain the evaluated optimal automation configuration. After the decision and implementation, the socio-economic improvements of the new configuration shall be re-evaluated on a regular basis to identify further improvement potentials. Additionally, methodologically guided workshops can be used to create new automation possibilities to keep the system updated.

While the Cost-based Module has been elaborated, the Worker Satisfaction Module will be further developed within the scope of the A4BLUE project. Since the cost evaluation is one of the two important aspects for decision making in a socio-economical context, it will be described in more detail in the following section.

3.3 Cost-Based Module

Within the Cost-based Module the impact of different LoAs on the final production costs per unit is assessed. This section introduces the calculation logic of the Cost-based Module under consideration of the required output information from upstream modules. Within a first step of the Cost-based Module, potential resources for the execution of assembly tasks are selected (cf. Fig. 5). These resources can be either automatic or non-automatic. This includes the implementation of new resources as well as resources already being in use at the assembly workplace, hence functioning as a benchmark for modifications.

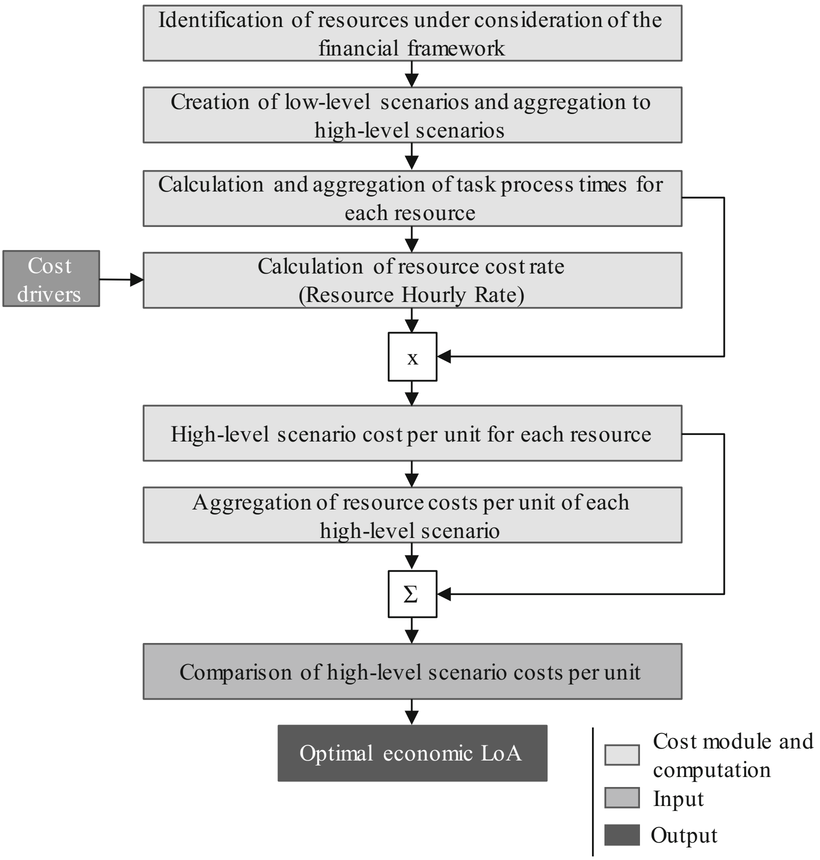


Fig. 5. Approach of the cost-based module

To ensure the overall optimum, the selection of resources must be based on an overall and combined consideration of all processes. Taking into account the resources' available capacity, it is the aim to have a resource executing as many tasks as possible. If each task is assigned to a single resource, this might lead to an excessive increase in investment costs, while individual resources are underutilized. Additionally, the resources have to be selected under consideration of the financial framework, by comparing permitted and resulting investments from a strategic perspective.

In the next step of the Cost-based Module, scenarios are created to achieve the optimal resource utilization. Different resources are assigned to a task, creating one low-level scenario for each combination of resource and task. For an exemplary task that can be executed manually, semi-automatically or fully-automatically, a low-level scenario number of three is obtained. Each sub-task on the lowest level needs to have at least one scenario. Subsequently the low-level scenarios are combined to high-level scenarios for an exhaustive cross task decision.

The determination of the optimal economic automation configuration depends on the resource's performance within the high-level scenarios. Since a resource is assigned to a different set of tasks in the different high-level scenarios, the resource utilization varies in cross scenario comparison. To quantify these differences and to convert them into a monetary measure, the resource cost rate per unit is calculated. The cost rate results from the sum of the cost drivers divided by the usage time per year (Resource Hourly Rate, cf. Fig. 5). The Resource Hourly Rate is then multiplied by the total assembly time of a resource within the high-level scenarios.

Since the defined scenarios generally comprise new automation configurations or new resources, additional assembly times per resource may be required. These can either be measured or estimated from time standards, experience, or methodologies such as Method-Time-Measurement (MTM). The cost drivers required for calculating the machine hourly rate are described in Table 1.

Table 1. Overview and description of cost drivers

Cost drivers	
Personnel costs	Costs for an assembly operator, incurred either for an employee in a completely manual process or proportionately for a machine operator
Energy costs	All types of energy such as electricity, compressed air, water and possibly also gas. Their consumption must be calculated per year. In the case of capital-intensive equipment, this share is usually negligible for an initial rough estimate of the unit assembly costs. Depending on the electricity consumption, the energy costs can also be represented as a variable cost element
Maintenance costs	Maintenance and repair expenses within the useful life in order to maintain the planned functionality of the equipment. For a rough estimation of maintenance costs, a value between 3% to 5% of the replacement value in one-shift operation and a value between 6 to 10% in two-shift operation can often be assumed to be sufficiently accurate [9]
Annual space costs	The space used and the costs per unit of space. The occupied space takes into account the base area of the equipment, the operating space as well as the space required for material supply

(*continued*)

Table 1. (*continued*)

Cost drivers	
Imputed depreciation	The quotient of the replacement value of the assembly system and the useful life. The replacement value consists of the purchase price (this is the investment sum of the assembly system) and the installation and start-up costs. The useful life is usually determined by the production period of the product
Interest costs	The capital invested in capital goods is calculated at 50% of the replacement value, assuming straight-line depreciation. The imputed interest costs are then calculated by multiplying the average fixed capital value with the imputed interest rate

The computation logic and procedures of the Cost-based Module have been implemented in a software prototype. The task execution times and the cost driver values per resource are input parameters of the software application. Based on this, the Resource Hourly Rate and the resources cost rate per unit are calculated automatically for all high-level scenarios. All resource cost rates are then summed up within a high-level scenario resulting in the production costs per unit for this scenario.

Conclusively the costs per unit of all high-level scenarios are compared to identify the most cost-effective automation configuration, the optimal economic LoA. The created scenarios are also used within the Worker Satisfaction Module, to ensure a shared basis for decision-making. After evaluating the worker satisfaction in a quantifiable manner for each scenario, the overall optimal LoA can be determined.

4 Conclusion

The aim of the presented work is to develop a methodological contribution to the determination of the optimal Level of Automation under consideration of worker satisfaction and assembly costs. A systematic, modular-related aspect is added to the approaches pursued in research so far. Therefore, a methodology has been developed which considers different boundary conditions by using a modular approach. The modules have the purpose to analyze the assembly process, to identify restrictions and to support the decision on optimal automation configurations from a cost-based as well as a worker satisfaction perspective.

In this paper the cost-based module is described in detail. This includes the specific cost elements being required to calculate the production costs of each unit as well as a logic for the creation of scenarios, which are the result of the combination of resources and tasks. Each scenario has specific costs, which are the basis for the economic identification of the optimal automation configuration. The evaluation procedure has been implemented as a software prototype to be further validated.

The worker satisfaction module within the overall methodology is still part of further research activities within the A4BLUE project, as well as the determination of the optimal overall decision with regard to the weighting of costs and worker satisfaction. However, the developed methodology is created for general use in use cases that involve humans as well as technology.

Acknowledgements. The research work has been conducted by the Laboratory of Machine Tools and Production Engineering (WZL) and the chair of Production Engineering of E-Mobility Components (PEM) at RWTH Aachen University within the research project A4BLUE, funded by the European Commission within the HORIZON 2020 program; Grant Agreement No. 723 828.

References

1. Kern, W., Rusitschka, F., Kopytynski, W., Keckl, S., Bauernhansl, T.: Alternatives to automobile assembly line production in the automotive industry. In: The 23rd International Conference on Production Research (2015)
2. Wagner, W.: Fabrikplanung für die standortübergreifende Kostensenkung bei marktnaher Produktion. Utz, München (2006)
3. Chryssolouris, G.: Manufacturing Systems: Theory and practice. Springer, New York (2006)
4. Frohm, J., Lindström, V., Winroth, M., Stahre, J.: The industry's view on automation in manufacturing. IFAC Proc. Vol. **39**(4), 453–458 (2006)
5. Endsley, M.: Level of automation: integrating humans and automated systems. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 41, pp. 200–204 (1997)
6. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Trans. Syst. Man Cybern. - Part A: Syst. Hum. **30**(3), 286–297 (2000)
7. Fasth, A.: Quantifying levels of automation: to enable competitive assembly systems. Dissertation, Chalmers University of Technology (2012)
8. Bakotić, D.: Relationship between job satisfaction and organisational performance. Econ. Res.-Ekonomska Istraživanja **29**(1), 118–130 (2015)
9. Gorlach, I., Wessel, O.: Optimal level of automation in the automotive industry. Eng. Lett. **16**, 141–149 (2008)
10. Boothroyd, G., Dewhurst, P., Knight, W.A.: Product Design for Manufacture and Assembly. CRC Press, Hoboken (2011)
11. Lotter, B.: Wirtschaftliche Montage: Handbuch für Elektrogerätebau und Feinwerktechnik. VDI-Verlag, Düsseldorf (1986)
12. Ross, P.: Bestimmung des wirtschaftlichen Automatisierungsgrades von Montageprozessen in der frühen Phase der Montageplanung. Dissertation, Technische Universität München (2002)
13. Co, H.C., Eddy Patuwo, B., Hu, M.Y.: The human factor in advanced manufacturing technology adoption. Int. J. Oper. Prod. Manag. **18**(1), 87–106 (1998)
14. Owusu-Ansah, C.M., Mprah, R.K.: The impact of library automation on the job satisfaction of library staff. Eur. J. Bus. Soc. Sci. **3**(9), 100–113 (2014)
15. Friedrichs, J.: Methoden empirischer Sozialforschung. VS Verlag für Sozialwissenschaften, Opladen (1990)
16. Nachreiner, F.: Grundlagen naturwissenschaftlicher Methodik in der Arbeitswissenschaft. In: Volpert, W., Müller, T. (eds.): Handbuch Arbeitswissenschaft, pp. 82–87. Schäffer-Poeschel, Stuttgart (1997)

17. Diekmann, A.: Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Rowohlt, Reinbek bei Hamburg (1995)
18. Geyer, G.: Entwicklung problemspezifischer Verfahrensketten in der Montage. C. Hanser, München (1991)
19. Annett, J.: Hierarchical Task Analysis (HTA). In: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., Hendick, H. (eds.) Handbook of Human Factors and Ergonomics Methods. CRC Press, Boca Raton (2005). 33-1-33-7
20. Frohm, J., Lindström, V., Winroth, M., Stahre, J.: Levels of automation in manufacturing. Egonomia – Int. J. Ergonomics Hum. Factors **30**(3), 28 (2008)
21. Frohm, J.: Levels of automation in production systems. Chalmers University of Technology, Göteborg (2008)

Digital Assembly Assistance Systems – A Case Study

Alexander Nikolenko(✉), Philip Sehr, Sven Hinrichsen, and Sven Bendzioch

Ostwestfalen-Lippe University of Applied Sciences and Arts, Campusallee 12, 32657 Lemgo, Germany
{Alexander.Nikolenko,Philip.Sehr,Sven.Hinrichsen, Sven.Bendzioch}@th-owl.de

Abstract. As the number of variations increases and batch sizes grow smaller, it can be difficult to fulfill quality and productivity requirements in manual assembly, as employees must record, process and interpret more information, then convert that information into action. Conventional instruction manuals in the form of text, tables or drawings quickly reach their limits. Innovative assistance systems are essential for keeping up with these changes, as they instruct and support employees in line with their specific situations. The goal of this article is to use a case study to illustrate the problems with providing information in manual assembly, and to suggest a potential solution in the form of an informational assistance system. The company considered in the case study stands out for its large production areas where complex, customer-specific truck bodies are mounted manually.

Keywords: Manual assembly · Assembly assistance systems · Industry 4.0 · Information delivery

1 Introduction

Customer-specific product variations with short product life cycles, combined with small batch sizes and short delivery deadlines, are the consequence of increased customer demands and intensifying competition [1]. This development is also associated with a change in requirements for assembly systems. In industrial companies, where a large number of complex products are assembled, the challenge is to ensure that the assembly work is completed in line with requirements. Frequent changes to work orders during the assembly process can lead to an early interruption in the learning curve [2], thereby reducing work productivity and increasing error rates. This situation is why informational assistance systems are so important, to help employees fulfill their duties cooperatively and optimize their activities [3, 4].

The goal of this article is to present a case study for introducing an assistance system in a manual assembly area. The company considered in the case study stands out for its large production areas where complex, customer-specific truck bodies are mounted manually. The case study considers a frame assembly work station.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 24–33, 2020.
https://doi.org/10.1007/978-3-030-20040-4_3

To achieve its goal, the process is divided into five steps. In the first step, a basic theoretical framework is presented (Sect. 2). In the second step, the existing assembly system is analyzed (Sect. 3). The analysis is completed using work and time studies. The paper focuses specifically on information delivery. In the third step, assembly assistance system requirements are identified in order to develop and implement a concept for an assistance system on this basis (Sect. 4). The requirements and the concept are discussed with employees in frame assembly to ensure their acceptance of the solution. Then the evaluation is completed as a fourth step (Sect. 5). Work and time studies are completed once again to review the effectiveness of the assistance system. Finally, the fifth step (Sect. 6) draws conclusions and provides an outlook for the future.

2 Basic Theoretical Framework

Information delivery within a work system is considered dynamic if that information is transmitted to employees on a need- and situation-specific basis. A technical system for dynamically delivering digitally prepared information is called an assistance system. Informational assistance systems record data via sensors and inputs, then process this data to provide employees the right information (“what”) at the right time (“when”) in the desired format (“how”) [5–7]. Information is typically transmitted visually, acoustically or tactilely [8]. The goal of an information assistance system is to avoid uncertainty and mental stress for users, warn them of dangers, or intervene in hazardous situations [8]. Assembly assistance systems support employees in completing their assembly work. Information on the product to be manufactured and the required production steps is provided to employees so they can make the right decisions during assembly [9]. In addition, assistance systems handle communication on operating materials (such as transmitting information on the target torque to the screwdriver) and document process parameters (such as saving the actual torque) [7]. One key goal of the assembly assistance system is to reduce the work required involved in initial training for employees with different professional backgrounds and performance levels. In addition, they are designed to improve employee motivation and minimize error rates [10]. The prerequisite is that the assistance systems are user-centered and context-specific [11].

Assembly assistance systems can be classified by different criteria and summarized in a morphological box [12]. The morphology makes it possible to represent characteristics on a specific assistance system by combining the characteristics of individual features. Changing one or more characteristics generates an analytic solution idea that can be used to further develop the company’s own support system. Table 1 provides an excerpt of key design parameters for assembly assistance systems.

Informational assembly assistance systems can be divided into stationary assistance systems, mobile assistance systems, handheld devices (such as tablet PCs) and wearables. Stationary assistance systems are permanently installed at a work station (such as a mounted projection device). Mobile assistance systems, in contrast, are moved to the assembly object via a mobile solution. Wearables can be classified by the body part on which they are worn (such as “smart glasses,” “smart gloves,” “smart watches”). In

Table 1. Excerpt from a morphology on assembly assistance systems [12].

CHARACTERISTICS	CHARACTERISTIC VALUES			
TYPE OF THE ASSISTANCE SYSTEMS	Stationary (fixed installation)	Mobile (mobile installation)	Hand device	Wearable - Head - Upper body - Arms/Hands - Legs/Feet
HUMAN-MACHINE INTERFACE	Unimodal		Multimodal	
TYPE OF INFORMATION OUTPUT	Visual (optical)	Auditory (acoustic)	Tactile-kinesthetic (tactile)	
TYPE OF THE INFORMATION INPUT/ SYSTEM CONTROL	Manual (via actuators)	Verbal (voice control)	Gesturing (tracking system)	Automatic (sensory)
USER RECOGNITION	None	Registration and uploading of user profiles	Automatic registration and uploading of user profiles	

addition, assembly assistance systems can be differentiated by whether the human-machine interface uses a unimodal or multimodal design. Unimodal means that the human receives information via a single sensory organ (frequently the eye) and that there is only one method (such as a keyboard) to input information into the system. Multimodal interfaces, in contrast, offer different input and output modalities [13]. Information is typically output by assistance systems through primarily visual means (such as via screens, sensor lights for a pick-to-light system, projections on the assembly object). The way in which information is input can be differentiated in the same way. Information is entered manually via control devices (such as keys), verbally via voice input, through gesture recognition, via tracking systems that record human movement [13] or automatically by sensors that monitor the state of the work object or status of the work process. In the past, assembly assistance systems have often been used to create standardized instruction manuals to be used by all employees working on the assembly system – independent of their individual features. In the future, employees will be receiving more and more support from individualized instruction manuals [14]. How much experience the employee has with completing the assembly task is especially important. As the employee gathers more experience, the assistance system must reduce the extent of support it provides so that it is not experienced as disruptive and so that it continues to enjoy a high level of acceptance.

3 Analyzing the Work System

To determine the potential usefulness of an assistance system, work and time studies have been completed in pilot areas of the company. Results of these work system analyses have already been published [15]. Assembly takes place in the company sequentially at several work stations. Typically, one or two employees work on a work station, whereby assembly work for a vehicle at a station can often take more than an hour. The following section will present the results of work studies on the frame assembly work station. Figure 1 shows a mounted frame from a top view. It consists of two frame rails, and typically more than ten cross members. The frame is mounted in a very large number of variations. The truck body is then assembled at the further work stations based on the mounted frame.

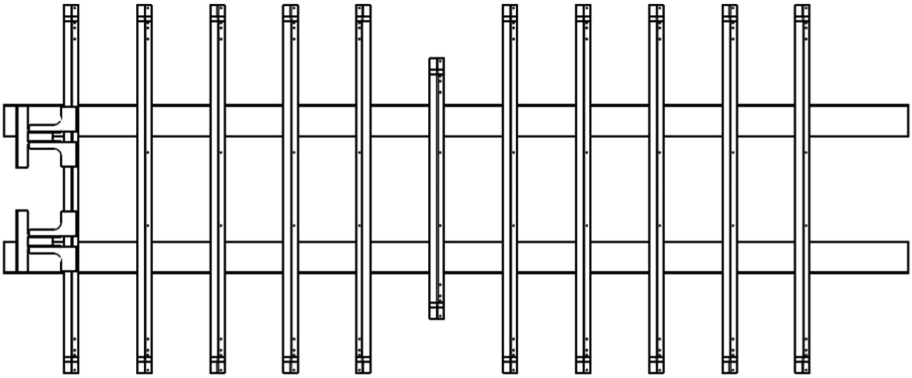


Fig. 1. Example of a frame, top view, consisting of two longitudinal beams and eleven cross members.

The frame assembly work process can be divided roughly into six work steps. The first step is work preparation. First, the employee asks colleagues in the subsequent work station which production order should be processed next. Then he looks for the relevant order documents. The order is registered in the ERP system by scanning the bar code printed on the order documents. The employee then receives an overview of the order by viewing the bill of materials information on the order documents. The employee also looks for the relevant design drawing from the document index on the server. He opens the relevant document with the drawing to obtain distances for the individual cross members and further order information (such as the number and location of attachment brackets, the number and position of drill holes). The employee first notes individual information on a sheet of paper. In the second step, the frame rails are shortened to the specified lengths using a plasma cutter and reworked. In the third step, the employee places the frame rails on the assembly table and attaches them. In the fourth step, a measuring tape and pencil are used to mark the attachment positions for the cross members and the positions of the reinforcing plates and brackets. In the fifth step, the frame is assembled by providing the cross members and other elements. The cross members are aligned using the markings and screwed together with the longitudinal members. In the sixth and last step, the finished frame is moved to the next work station and the order is registered as “complete” in the ERP system.

A time study was completed in order to be able to quantify the potential for implementing an assistance system. The entire work process was divided into process sequences for this purpose. The process sequences were assigned to six process types (see Table 2). An example for a value creating process sequence with the potential for optimization using information technology (process type 3) is determining screw-fitting positions using a tape measure. An example for an avoidable loss of information (process type 6) is displaying and processing information that is not required at the work station. The object of the time study was the assembly of a total of 10 frames. The actual order time for assembling a frame averages 54 min, according to Table 2. The results show that the order time for assembling a frame can be reduced by up to 7 min (13%) by using an informational assistance system (see process types 3 and 6 in Table 2).

Table 2. Percentage of time for assembling frames (n = 10).

No.	Process types	Duration [min]	Percentage [%]
1	Value creation	239	44
2	Value creation with general potential for optimization	60	11
3	Value creation with potential for optimization through information technology	27	5
4	Unavoidable waste	158	29
5	Avoidable non-information-related waste	16	3
6	Avoidable information-related waste	44	8
$\sum$		544	100

There is potential for improvement in providing information primarily in work steps 1 (order preparation), 4 (determining and marking attachment positions) and 6 (reporting the order as complete on the terminal). During step 1, the problem is that it is not immediately apparent which order should be processed next, since the orders are not sorted by priority or due date. Another problem is that design drawings include data for the entire design of the truck body, so that employees have to choose the data relevant for their work station from this complex overall design drawing. In addition, some frame rails must be shortened. The length information is provided on the drawing. The bill of materials to be provided also includes part positions that are mounted at the following work station, so significantly more information is available at this work station than is actually required. Displaying unnecessary information requires the user to search for what they need and results in an increased risk of errors. In addition, employees typically note the information they need on a sheet of paper to avoid having to look at the drawing and bill of materials repeatedly during the assembly process. During step 4, the problem is that information from the model (CAD drawing) must be transferred to the real work object by measuring distances and making pencil markings. Step 6 can be improved, since employees have to log into the terminal again to report the order as complete in the ERP system via a multi-step process.

4 Development of the Assistance System

4.1 Identifying Requirements

With respect to work step 1 (order preparation), the assistance system should automatically sort orders by priority. In addition, the right design drawing must be called up automatically. With regard to the subsequent work steps, only the relevant information is to be displayed to the assembly employee in such a way that he can quickly capture it. In this context, the computer terminal screen should be replaced by a much larger screen so that all information is visible from a greater distance. This allows employees, for instance, to read measurements directly off the screen and transfer them to the frame rails. Overall, the assistance system to be developed should reduce the perceived informational complexity of the assembly task, so that

(a) the resulting cognitive strain on employees is reduced,
(b) initial training times for new employees are shortened,
(c) the number of assembly errors is reduced,
(d) work productivity is increased by reducing the number of decisions and shortening the time spent in searching and orientation.

In addition, a consistent digital process chain should be created where required information is read out automatically from the CAD drawings and displayed to the employee. Furthermore, the needs of the employees should be taken into account so that the user interface for the assistance system software can be individually adjusted.

4.2 Concept

The assistance system is designed in consideration of a user-appropriate development process which focuses on the needs of frame assembly employees. The assistance system concept consists of three levels, which are illustrated in Fig. 2. The level in direct contact with the user is the human machine interface (HMI). All relevant information on the work order is displayed to the employee via the user interface. The user can also access specialized information. The next level is the back end, in this case the terminal PC at the work station. This level involves the actual implementation of system functions. This includes all processes the user does not directly perceive, such as network communication or data processing. The third level of the concept is called the data level. This level includes all the data necessary for the assistance system to function.

The numbered arrows indicate the sequence of the individual work steps. First, the user selects an order that has not yet been processed from the document index (1. order selection). The information for this order is then downloaded from the ERP system. Order information is saved in an XML file in the ERP system, which simplifies algorithmic processing. In the second step, a parser is used to create an informational model from the XML file that contains all important data. A parser is a program that analyzes and processes data, then converts it into a format suitable for further use. The design drawing for an order can be determined from the order information data and can then also be downloaded from the ERP system (3. identify drawing). The design drawing is provided in DXF format.

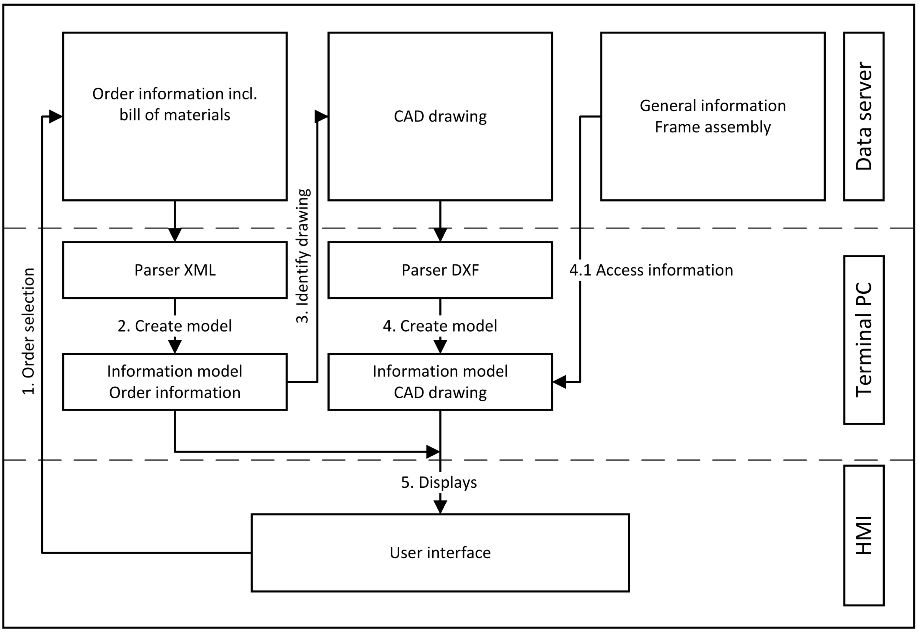


Fig. 2. Assistance system concept

Additional general information on frame assembly is required for the information model of the design drawing. To be able to filter all the relevant parts out of the drawing, the parts relevant to frame assembly must be identified. This additional information is accessed in step 4.1. Then, a parser is used to create another informational model from the additional information and design drawing (4. create model). Once both informational models are available, they are displayed in the user interface.

The user interface is divided into two areas: Order information (A) and design information (B). Both areas are designed for different work steps in the assembly process. Part A makes it possible to view order information and select the next order. In this step, the employee is working directly at the screen. Part B consists of two parts and shows a view of the frame from the side and the top, with relevant information as shown in Fig. 3. A frame is always shown in side and top views in the original design drawing. Employees are familiar with this type of display, so it will be retained.

To mark the measurements on the frame rails, the employee must move away from the screen. In this case, an expanded menu is displayed to the employee where he can individually adjust the degree of detail shown in the design drawing and/or the font size for the dimensions. The menu can be hidden after configuration to enlarge the area where information is displayed. In addition, the old 15″ screen was replaced by a 40″ screen.

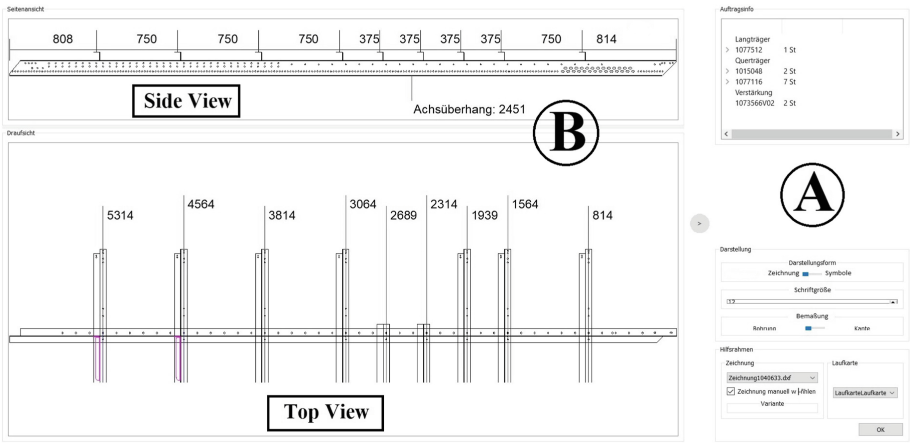


Fig. 3. Informational models consisting of the side view, top view and information on the frame.

5 Evaluation

After introducing the assistance system for testing purposes, another time study was completed. The object of the time study was, once again, assembling ten frames. The results of the time study are shown in Fig. 4 in comparison to the data on the initial situation. Potential areas for using an assistance system after Sect. 4 include steps one, four and six. After introducing the assistance system, most of the time for the first step is eliminated. For instance, it is no longer necessary, to manually search for order data and drawings in the system or note dimensions on a sheet of paper. The assistance system has not yet been able to solve the problems in the sixth step. Therefore, step 6 will not be considered in the following. Through using the assistance system, it has been possible to reduce the average assembly time for the first and fourth step from 7:32 min to 2:26 min. In addition, the scatter of the time values has been reduced (see Fig. 4).

The System Usability Scale (SUS) was used to evaluate the user interface. The questionnaire consists of ten items that are rated using the Likert scale. The scale provides an overview of subject assessments of user-friendliness. The SUS questionnaire was completed by three frame assembly employees who work regularly in the work system. The results of the questionnaire indicated a usability score of 77.5. A reference value for good usability would be 68 or higher. This means the usability of the user interface should be rated as good.

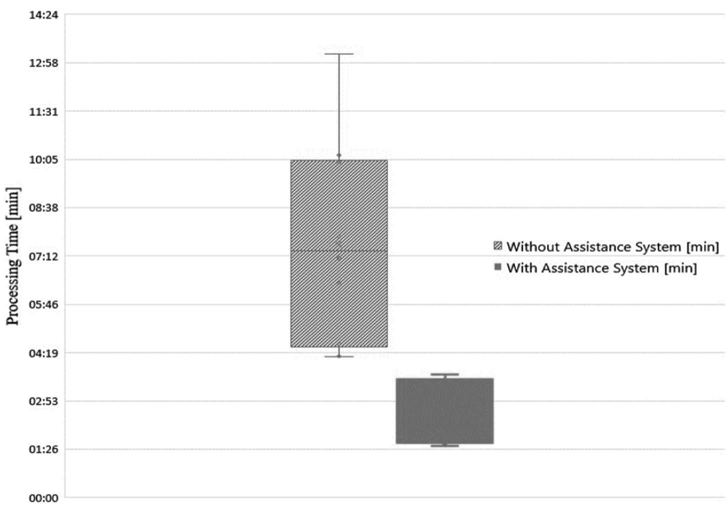


Fig. 4. Comparison of completion times without assistance system (before) and with assistance system (after) for work steps 1 and 4.

6 Conclusion

This article deals with the development and use of an informational assistance system for manual assembly. The system was developed for a company which mounts customer-specific truck bodies. The assistance system filters design data and provides it to the employees on a needs-based basis. The results of work and time studies show that a user-centered assistance system design can help improve both employees' work situation and their productivity in manual assembly. The case study also clearly shows how an assistance system can be used to solve typical problems of information delivery in manual assembly. The assistance system developed will be improved and developed further as time goes on. During the next step, the plan is to display information directly on the work object, for instance using LEDs or projection. This means that it will no longer be necessary to measure and mark. A connection to the ERP system allowing orders to be checked in and out would save additional time. Furthermore, the user interface should be optimized based on another workshop.

Acknowledgments. This publication is part of the "Montextas4.0" project for excellence in assembly funded by the German Ministry of Education and Research (BMBF) and the European Social Fund (ESF) (grant no. 02L15A260).

References

1. Abdi, M.R., Labib, A.W., Delavari Edalat, F., Abdi, A.: Integrated Reconfigurable Manufacturing Systems and Smart Value Chain – Sustainable Infrastructure for the Factory of the Future. Springer, Cham (2018)
2. Hunter, J.E., Schmidt, F.L., Judiesch, M.K.: Individual differences in output variability as a function of job complexity. J. Appl. Psychol. **75**(1), 28–42 (1990)
3. Gerke, W.: Technische Assistenzsysteme – Vom Industrieroboter zum Roboterassistenten. De Gruyter Oldenbourg, Berlin (2014)
4. Unrau, A., Hinrichsen, S., Riediger, D.: Development of projection based assistance system for manual assembly. In: Ergonomics 2016 – Focus on Synergy, 6th International Ergonomics Conference, Zadar, Zagreb, Croatia, pp. 365–370 (2016)
5. Hollnagel, E.: Information and reasoning in intelligent decision support systems. Int. J. Man-Mach. Stud. **27**(5–6), 665–678 (1987)
6. Claeys, A., Hoedt, S., Soete, N., Van Landeghem, H., Cottyn, J.: Framework for evaluating cognitive support in mixed model assembly systems. IFAC-PapersOnLine **48**(3), 924–929 (2015)
7. Hinrichsen, S., Bornewasser, M.: How to design assembly assistance systems. In: Karwowski, W., Ahram, T. (eds.) Intelligent Human Systems Integration 2019, IHSI 2019. Advances in Intelligent Systems and Computing, vol. 903, pp. 286–292. Springer, Cham (2019)
8. Geiser, G.: Informationstechnische Arbeitsgestaltung. In: Luczak, H., Volpert, W. (eds.) Handbuch Arbeitswissenschaft, pp. 589–594. Schäffer-Poeschel, Stuttgart (1997)
9. Bertram, P., Birtel, M., Quint, F., Ruskowski, M.: Informationsmodellierung zur Beschreibung manueller Tätigkeiten an Handarbeitsplätzen. In: Burghardt, M., Wimmer, R., Wolff, C., Womser-Hacker, C. (eds.) Mensch und Computer 2017 – Workshopband, pp. 133–137. BoD – Books on Demand, Norderstedt (2018)
10. Bächler, L., Bächler, A., Kölz, M., Hörz, T., Heidenreich, T.: Über die Entwicklung eines prozedural-interaktiven Assistenzsystems für leistungsgeminderte und -gewandelte Mitarbeiter in der manuellen Montage. Esslingen (2015)
11. Wölfle, M.: Kontextsensitive Arbeitsassistenzsysteme zur Informationsbereitstellung in der Intralogistik. Diss. TUM, München (2014)
12. Hinrichsen, S., Riediger, D., Unrau, A.: Assistance systems in manual assembly. In: Villmer, F.-J., Padoano, E. (eds.) Production Engineering and Management. Proceedings 6th International Conference, Lemgo, Germany, 29–30 September 2016, vol. 01/2016, pp. 3–14. Publication Series in Direct Digital Manufacturing (2016)
13. Schlick, C., Bruder, R., Luczak, H.: Arbeitswissenschaft, 4 Aufl. Springer, Berlin (2018)
14. Kölz, M., Bächler, A., Kurtz, P., Hörz, T.: Entwicklung eines interaktiv, adaptiven Montageassistenzsystems. In: Gesellschaft für Arbeitswissenschaft e.V. (eds.) Verantwortung für die Arbeit der Zukunft. Bericht zum 61. Kongress der Gesellschaft für Arbeitswissenschaft e.V. (GfA), Karlsruhe, 25–27 February 2015. GfA-Press, Dortmund (2015)
15. Hinrichsen, S., Bendzioch, S.: How digital assistance systems improve work productivity in assembly. In: Nunes, I.L. (ed.) Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2018 International Conference on Human Factors and Systems Interaction, 21–25 July 2018, pp. 332–342. Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA. Springer, Cham (2018)

Automatic Particle Classification Through Deep Learning Approaches for Increasing Productivity in the Technical Cleanliness Laboratory

Ronny Zwinkau(✉), Simon Frentrup, Roman Möhle, and Jochen Deuse

Institute of Production Systems, TU Dortmund University, Leonhard-Euler-Straße 5, 44227 Dortmund, Germany
Ronny.Zwinkau@ips.tu-dortmund.de

Abstract. Understanding the properties of particles plays a vital role in assessing the component cleanliness and its origin in the manufacturing process. We propose a classification method using deep convolutional neural networks. Using a dataset of 70,000 annotated images, we achieve a accuracy of 97.7% for a binary classification in metal and non-metal particles comparable to state-of-the-art polarized light microscopy according to VDA 19-1 and ISO 16232. Manual follow-up checks in a cleanliness laboratory are not required due to the robustness of the classification system.

Keywords: Technical cleanliness · Particle classification · Metal/Non-metal classification · Convolutional neural networks

1 Motivation and Focus

The cleanliness of components has become an important quality feature in industry. The industry-specific variety of applications and processes results in different cleanliness requirements relating to the critical contamination type, quantity, size and hardness [1]. Due to their potential to negatively influence technical systems, like blocking of valves or bearings, hard particles are subject to limitations in size and quantity, especially in the automotive and electronics industries. In addition, the number of metallic particles is particularly regulated due to the electrolytic conductivity (risk of short circuits) and hardness. Thus, in industrial production, a reliable distinction between metallic and non-metallic particles during the measurement of the particle load in a cleanliness laboratory is crucial.

2 State of the Art

The cleanliness analysis is generally divided into direct and indirect methods [2]. In the case of three-dimensional geometries of the components, indirect methods are usually used, in which the contamination is extracted from the component by air or a liquid, then filtered and subsequently analysed using various methods [3] (Table 1).

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 34–44, 2020.
https://doi.org/10.1007/978-3-030-20040-4_4

In industrial production, cleanliness analysis with a stereo microscope and automated particle counting and classification into the classes "metallic", "non-metallic" and "fibres" has become the standard analysis [3]. Although a distinct classification of particles in metallic and non-metallic is only possible with advanced methods, the majority of microscopes conducts the classification based on metallic reflections and a manual follow-up inspection.

Most optical light measuring systems available on the market use an incident-particle analysis system with polarization light analysis for classifying particles as metallic or non-metallic. This method irradiates the particle accumulation with linearly polarized light and analyzes the specific reflections. In contrast to non-metals, metals reflect light with the same polarization. The reflections of the metals can be blocked using polarization filters. If there are bright spots on the detected particles and these spots become dark when the polarization filters are used, the corresponding particle is classified as metallically glossy [3, 8].

Furthermore, the metal/non-metal classification depends on technical parameters as well as the expertise and experience of the operating personnel. Parameters that influence the result, respectively the comparability of the results, are listed in VDA 19.1 [3]:

- Type of imaging lens
- Degree of magnification or zoom level
- Type and exact geometry of illumination
- Procedure for characterizing metallic shine and exact parameter settings
- Individual properties of metallic particles and individual particle orientation.

This leads to misclassifications in the application, as shown exemplarily in Table 2.

Due to the risk of misclassifications, VDA 19.1 [3] recommends double-checks and, if necessary, reclassification. This approach requires specially trained skilled staff and thus increases not only the lead-time and utilization of laboratory resources but also the costs incurred.

Rochowicz and Schmauz [9] point out that essential information can be found in the recorded images of the detected particles. The number of such images steadily increases due to the growing relevance of technical cleanliness, the increasing number of laboratories and the falling prices for storage capacity (Fig. 1). Image Data Mining (IDM) approaches are already applied for the search of images, their classification or object recognition in different fields. IDM describes the application of statistical methods to a large collection of images [10, 11]. A suitable database and high-performance hardware are necessary for the training of IDM models, while the application can take place using conventional computers. Thus, there is a great potential to use IDM methods also for the classification of particle images.

Leavers [12] states that IDM techniques produce more consistent, accurate and informative results compared to manual analysis and classification of particle images by experts. For the classification of particles in technical systems caused by wear and tear, features like texture, shape and color are for example used by Myshkin and Grigoriev [13]. They also highlight the potential of neural networks for classification. Albidewi [14] classifies wear particles in engines with single-layer neural networks based on textures. However, the application of so-called "deep learning" methods,

Table 1. Overview of relevant measurement and analysis methods for component cleanliness [3–7]

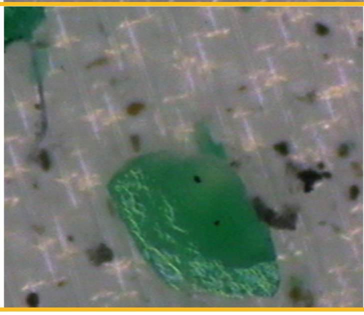
method	**principle**	**measurement result**	**restrictions**	**initial costs (guideline value, without training)**	**running costs (guideline values)**	**duration**
direct methods and quick test (direct)						
direct component inspection	evaluation by eye or microscope	subjective evaluation clean / not clean	simple, accessible geometries, only qualitative measurement result	low	low	short
standard methods according to VDA 19.1 (indirect)						
light-optical microscopy	filter scanning and following evaluation with image processing software, polarization filter for classification into metallic / non-metallic, **imaging**	Number of particles per size class, classification according to "metallic", "non-metallic", "fibre"	No gloss level criterion available, manual verification required	medium	low	short
scanner			Analogue to light-optical microscopy	medium	low	short
gravimetry	Determination of the weight difference of the filter before and after extraction	Mass of contamination	non-imaging	medium	low	short
optical particle counter	number of particles passed through a beam of light	amount of particles	no individual analysis possible	medium	medium	medium

(*continued*)

Table 1. (*continued*)

instrumental special methods (indirect)						
computer tomography	three-dimensional analysis of particles, **imaging**	length, width, thickness and volume	n. s.	high	high	long
(micro-) Raman	comparison of particle to reference spectrum; particle size > 0.5 µm	chemical composition of organic materials	only for organic materials, restriction for black and highly fluorescent materials	high	high	long
infrared spectroscopy (IR)	absorption spectra are used for qualitative and quantitative determination	chemical composition	n. s.	high	medium	medium
Laser Induced Breakdown (Plasma) Spectroscopy	comparison of particle spectrum with reference spectrum	chemical composition	not suitable for organic materials, destructive testing	high	medium	long
scanning electron microscope with X-ray analysis	scanning of the sample surface with an electron beam and evaluation of the generated X-rays, **imaging**	particle size distribution down to the nanometer range, topology and chemical composition	not suitable for organic materials, greyscale images	high	high	long (in combination with LM low – medium)
electron spectroscopy	sample surface is bombarded with ions and the mass of released secondary ions is analyzed, **imaging**	chemical composition and quantity	High initial costs and need for specialists to operate the system	high	high	long

Table 2. Metal/non-metal classification by the VDA 19.1 standard analysis

		classification by standard analysis	
		„metallic“	„non-metallic“
real classification	metallic		
	non-metallic		

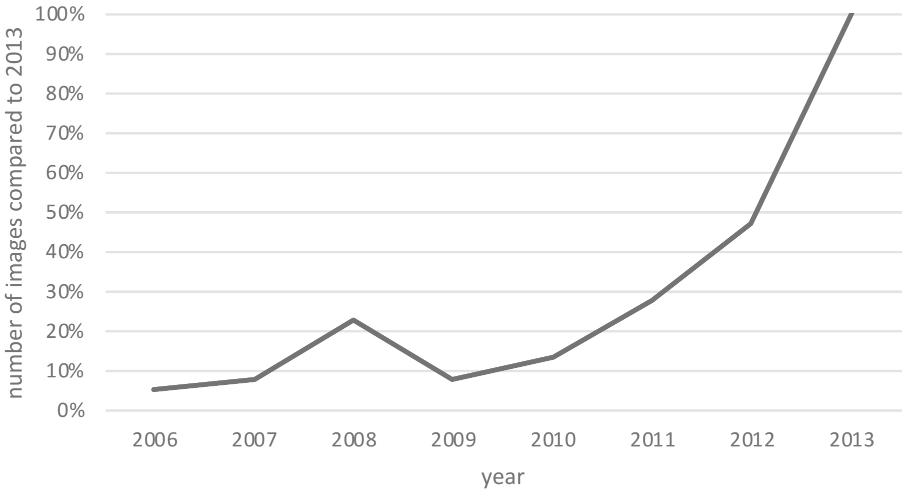


Fig. 1. Percentage increase in particle images from 2006–2013

meaning neural networks with many layers of representation, has not yet been focused on in the context of the particle classification in socio-technical systems.

The main advantage besides the high classification quality of Deep Learning, as a sub-discipline of machine learning, consists in the automated learning of the image characteristics relevant for the classification [15]. In deep learning, model training is characterized by two essential characteristics that significantly increase the performance of the models [15]:

1. Incremental, layer-by-layer approach, in which increasingly complex representations of the image data are generated
2. Each representation layer and its weighting are learned along with the other layers.

For this reason, Deep Learning has experienced a real hype in recent years (Fig. 2) and is used in various applications of daily life, such as voice recognition, translation function or image search.

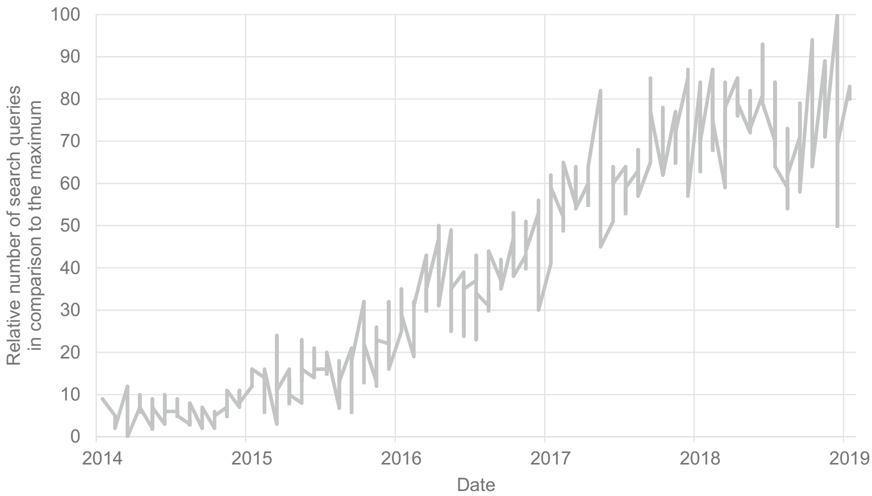


Fig. 2. Worldwide search queries for deep learning from 2013–2018 [16]

The procedure for setting up the deep learning model is based on the phase model of the Cross Industry Standard Process for Data Mining (CRISP-DM). The CRISP-DM includes the phases of business and data understanding, data preparation, modeling, evaluation and deployment [17].

3 Business and Data Understanding

A uniform and error-free database is of great importance for creating a deep learning model. The database on which this model is built is essentially based on two sources:

1. The Institute of Production Systems (IPS) at TU Dortmund University received anonymous particle images from an accredited laboratory containing information on metal/non-metal classification as well as length and width of the particles. The database comprises images of approximately 200,000 metallic and non-metallic particles. The particle images were automatically pre-classified by measuring systems and the results were checked and adjusted by experts. After the data cleansing (particles <200 μm have an unfavourable ratio of particle size on the image and total size of the image due to the standard zoom) approximately 60,000 images of metallic and non-metallic particles have been available for constructing the model.
2. In order to completely exclude misclassifications by the measuring system and the operator, the IPS carried out approximately 100 process investigations. In each case, the images of the 100 largest metallic and non-metallic particles were extracted. Through isolated investigations of separation and forming processes, approximately 10,000 metallic and non-metallic particles with a length >200 μm could be analyzed with a high degree of certainty. The particle suction extraction system CPS^2 of CleanControlling GmbH was used for the extraction. This enabled a targeted extraction of particles directly at the process and thus avoided costly sample preparation. In addition, the risk of invasion of further contaminants not caused by the process itself were minimized. A Jomesa measuring system, which was maintained in May 2018 and was rated satisfactory in the round robin test of the Cleaning Excellence Center [18] with a z score of 1.1, was used to measure the particles.

4 Data Preparation

Images of the file type JPEG have been used for the construction of the model. To increase the industrial applicability of the model, the program accepts any common image format. The choice of the image format or the image compression has no significant influence on the quality of the classification [19].

A data augmentation is performed prior to the training of the model. This step prevents overfitting and enlarges the database. The operations shearing, rotation, zooming, vertical and horizontal mirroring are applied to each image.

5 Modeling

The VDA 19.1 standardizes the evaluation parameters for cleanliness analysis, which is very helpful for creating and applying the model. Important standardized parameters of this analysis are the brightness and the relationship between the zoom of the microscope and the resolution of the used camera. These criteria ensure that clearly visible particles can be used in the learning process and that higher classification accuracies can be achieved. Images that are created using the parameter settings described in VDA 19.1 [3] and saved using the existing image extraction functions thus meet the requirements for image size and brightness.

75% of the available pictures (53,882) were used for the training of the model. The Deep Learning model is based on an InceptionResNetV2 model [20], which has already been pre-trained for the task of image classification over 160 training iterations (so-called epochs) on the ImageNet database. The ImageNet database is a collection of over 1.4 million images divided into 100,000 categories [15]. The model used comprises 54,279,266 trainable parameters and shows a TOP-5 accuracy (correct result is in the top 5 probabilities) of 95.3%. To prevent overfitting of the model, the classification accuracy is tested on a test data set of 5,962 images after each fifty epochs. As shown in Fig. 3, an accuracy of >95% on both the training and the test data occurs after only a few epochs. A small difference between the two graphs indicates a good model without overfitting.

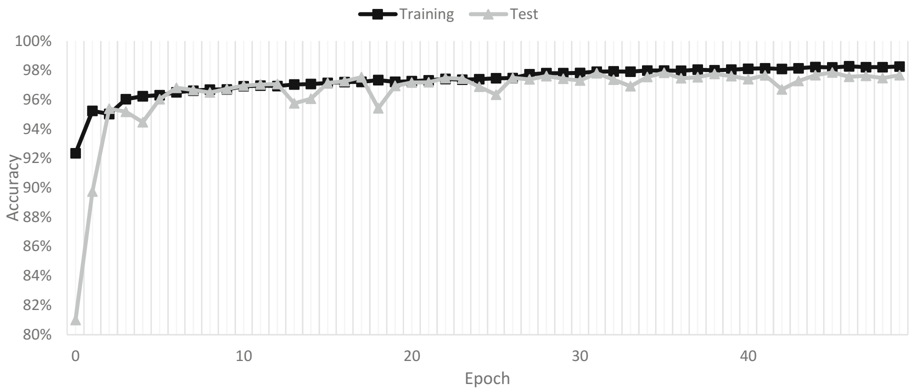


Fig. 3. Training and test accuracy over all epochs

Since the training of the neural network primarily requires matrix multiplications that can be executed very well in parallel, the computing time can be greatly reduced by using suitable hardware architectures. The high-performance computing cluster LiDO3 of the TU Dortmund was used to train the model. This cluster has a total of 8160 CPU cores each with at least 64 GB RAM and 40 Nvidia K40 graphics cards (GPU). With two of these GPUs, a runtime of about three hours could be realized for the training of the model.

6 Evaluation

On the one hand, the practical validation of the developed model was based on the isolated process investigations in order to exclude misclassifications of the Jomesa measuring system. On the other hand, the model was compared with the classification of an accredited laboratory in order to increase the range of particle expressions. The model was validated with 4,147 particle images generated in the isolated process investigations of the IPS and not used in the training of the model (Table 3). This shows an overall high agreement (97.7%) between the classification of the Deep Learning Model and the results of the pre-classification by the Jomesa measurement system used.

Deviations between the model and the measuring system were detected especially for non-metallic particles. In this case, the correct classification rate of the developed model is higher than the rate using the Jomesa measuring system. This can be attributed to the fact that several parameters, such as the color or the surface texture of the particle, are used for classification in the model.

Table 3. Conformity of the metallic/non-metallic classification of the developed model with the results of the isolated process investigations of the IPS (n = 4147 images).

		By model predicted class		
		Metallic	Non-metallic	
Evaluation of the process analysis using the Jomesa analysis system	Metallic	2,000	71	96.6%
	Non-metallic	36	2,040	98.3%
	Concordance	98.2%	96.7%	**97.7%**

Since the particle images used for training the model were classified in advance with established measurement systems, it is also shown that the model also correctly predicts misclassifications of the measurement systems. Dark metallic particles in particular are currently misclassified as non-metallic by both the measurement system and the deep learning model (Fig. 4).

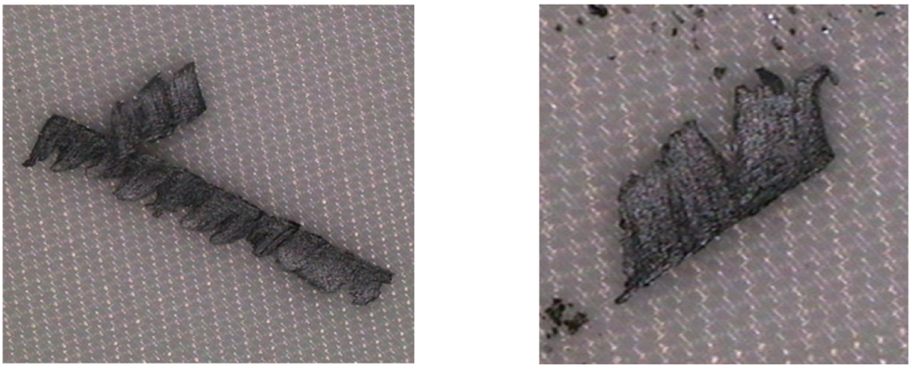


Fig. 4. Metallic particles (milling) misclassified as non-metallic

Further research will identify these misclassifications and adjust the database. This is a decisive advantage compared to the established measurement systems, which do not allow adaptations.

To identify and correct these errors, the developed model for metallic/non-metallic classification is available as a web platform for interested companies. Particle images can be uploaded here with or without the result of a pre-classification and can be classified with the model. All uploads are stored anonymously in a database and the results of the model are compared with the inputs. Thus, weak points of the model can be identified and the model can be trained with new information. This web platform has been made available to an accredited laboratory. The results of the classification of 1,530 particle images (Table 4) show a total agreement between model and laboratory of 96.3%.

Table 4. Validation of the results from the metallic/non-metallic classification model by an accredited cleanliness laboratory (n = 1530 images)

		By model predicted class		
		Metallic	Non-metallic	
Classification by laboratory	Metallic	720	35	95.4%
	Non-metallic	22	753	97.2%
	Concordance	97.0%	95.6%	**96.3%**

7 Deployment

In order to use the achieved results in an industrial context, it is possible to integrate the generated models into the microscope software via the REST interface without additional effort. The web-based demonstrator can be used to reduce rework and interpretation effort in laboratories. The additional information obtained in this way enables a company to directly influence its processes to optimize them. Interested companies can test the web platform with the classification model free of charge. Please contact us to receive your access data. The uploaded particle images are stored anonymously by the IPS and used for the further development of the model. At the same time it serves as an assistance for laboratories and industrial companies to classify particles and to regulate compliance with cleanliness limits.

8 Summary and Outlook

It was shown that the significant increase of available particle images brings potential for the use of image data mining approaches. The achieved classification results of the deep learning model can be evaluated as very good. The high degree of agreement with the process investigations carried out and the classification results of proven experts shows that a promising and cheaper alternative to the established VDA 19.1 methodology has been developed. In contrast to the established measurement systems, the architecture of the model also offers the possibility of continuous adaptation and expansion of the database. Thus, it is possible to react to misclassifications and to eliminate them. Furthermore, the confidence of the classification as well as the simple integration of the model into existing software structures should be emphasized. The classification of an image takes considerably less than a second. Due to the elimination of the polarization filter, a reduction in hardware costs can also be expected.

Further work of the IPS aims at a refinement of the classification with regard to the formation of particles. Against the background of the current discussion on the use of data within the framework of Industry 4.0, this approach creates a bridge from the measurement results to a targeted improvement.

References

1. Kreck, G., Holzapfel, Y.: Reinheitsvalidierung von kontaminationskritischen Produkten, reinraum online (2013). Accessed 31 Jan 2016
2. Verband der Deutschen Ingenieure e. V. 2083: Reinheitstauglichkeit und Oberflächenreinheit, VDI 2083. Beuth Verlag GmbH, Berlin (2006)
3. Verband der Automobilindustrie e. V. 19.1: Prüfung der Technischen Sauberkeit - Partikelverunreinigung funktionsrelevanter Automobilteile, 2nd edn, VDA 19.1 (2015)
4. Zwinkau, R., Krebs, M.: Schlussbericht zu dem IGF-Vorhaben 17495 N/1 Reinheitsgerechte Materialflusssysteme. Technische Universität Dortmund (2014)
5. Kloke, U.: Auslegung von Bauteilreinigungsanlagen mit Hilfe eines Fachinformationssystems. TU Dortmund (2003)
6. König-Birk, J.: Die richtige Analysemethode für filmische und partikuläre Verschmutzungen finden. Fachtagung Technische Sauberkeit, Stuttgart (2014)
7. Valet, O.: Methoden zur Partikelidentifizierung: Analyse der Stärken und Schwächen neuer Technologie. Fachtagung Technische Sauberkeit, Stuttgart (2014)
8. Grüner, S., Lumpe, H., Streibl, D.: Verfahren zur Analyse einer metallische und nichtmetallische Partikel enthaltenden Partikelansammlung und Vorrichtung zur Durchführung des Verfahrens, DE102013106929A1 (2013). https://patents.google.com/patent/DE102013106929A1/de. Accessed 4 Jan 2019
9. Rochowicz, M., Schmauz, G.: Neues Rund um die Technische Sauberkeit: Restschmutz als Qualitätsindikator. J. für Oberflächentechnik (3), 54–56 (2010)
10. Oh, J., Lee, J., Hwang, S.: Video data mining. In: Wang, J. (ed.) Encyclopedia of Data Warehousing and Mining. Idea Group Reference, Hershey (2006)
11. Hsu, W., Lee, M.L., Zhang, J.: Image mining: trends and developments. J. Intell. Inf. Syst. (1), 7–23 (2002). School of Computing. https://eprints.usq.edu.au/5630/. Accessed 30 Aug 2018
12. Leavers, V.F.: Classifying wear debris particles with automatic image analysis. Pract. Oil Anal. **7**(11) (2005)
13. Myshkin, N.K., Grigoriev, A.Y.: Morphology: texture, shape and color of friction surfaces and wear debris in tribodiagnostics problems. J. Frict. Wear **29**, 192–199 (2008)
14. Albidewi, I.A.: Wear particles surface identification using neural network. IJCSNS Int. J. Comput. Sci. Netw. Secur. **8**(1), 262–265 (2008)
15. Chollet, F.: Deep Learning mit Python, 2018 Auflage. MITP, Frechen (2018)
16. Google Trends: Weltweite Suchanfragen für "Deep Learning": 01/2014–01/2019 (2019)
17. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., et al.: CRISP-DM 1.0 - step-by-step data mining guide (2000)
18. Haberzettl, S.: Vergleichsversuche mit partikulären Messsystemen. Fachtagung Technische Sauberkeit, Leipzig (2018)
19. Dodge, S.F., Karam, L.J.: Understanding how image quality affects deep neural networks. CoRR, abs/1604.04004 (2016)
20. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, Inception-ResNet and the impact of residual connections on learning. CoRR, abs/1602.07261 (2016)

How to Combine Lean, Human Factors and Digital Manufacturing – A Teaching Concept

Benjamin Adrian(✉), Sven Hinrichsen, Alexander Nikolenko, and Frederic Meyer

Ostwestfalen-Lippe University of Applied Sciences and Arts, Campusallee 12, 32657 Lemgo, Germany
{Benjamin.Adrian,Sven.Hinrichsen,Alexander.Nikolenko, Frederic.Meyer}@th-owl.de

Abstract. The Toyota Production System became well-known in the 90s and stands for highly efficient processes. The success of the production system stems from its methods and its focus on human factors. For some years, production research has focused on the topic of digital manufacturing. This technology-oriented approach is pursued quite independently of the Toyota Production System. As a result, technical solutions may prove incompatible with the Lean philosophy. Therefore, operational practice must link the Lean philosophy with new technologies in order to make work processes and material flows productive and ergonomic simultaneously. As a part of their education in industrial engineering at the Ostwestfalen-Lippe University of Applied Sciences and Arts, students learn all current Lean methods by means of business games. One of these has been supplemented with information technology components. The objective of this article is to introduce this business game and to explain its didactic concept.

Keywords: Lean · Human factors · Digital manufacturing · Business game · Evaluation · Questionnaire

1 Introduction

For some years, the approach of digital manufacturing has been pursued in production research. This approach is technology-oriented and is pursued quite independently of the Toyota Production System. As a result, there is a risk that individual technical solutions prove incompatible with the Lean philosophy. Therefore, operational practice must link the Lean philosophy with new technologies in order to make work processes and material flows both highly productive and ergonomic at the same time.

As a part of their education in industrial engineering at the Ostwestfalen-Lippe University of Applied Sciences and Arts, students learn all current Lean methods by means of business games. At the same time, demonstrators have been implemented to research various technologies from the context of digitalization in industrial production and to integrate them into education. In the process, special emphasis has been placed on connecting the organizational/personnel approach of the Toyota Production System

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 45–55, 2020.
https://doi.org/10.1007/978-3-030-20040-4_5

with the technology-oriented approach of digital manufacturing. For example, a business game developed several years ago to teach the Kanban method has been supplemented with information technology components. The business game can be used to exemplify how different production philosophies can be combined effectively.

The aim of this article is to introduce this business game and to explain its didactic concept. For this purpose, the article is subdivided into four sections following the introduction. Section 2 presents the theoretical principles of the Lean philosophy and the Toyota Production System, digitalization in industrial production, and human factors in the didactics of industrial engineering. Section 3 introduces the teaching concept developed to impart the Kanban method. The evaluation results of this business game are presented in Sect. 4. Section 5 contains a conclusion and an outlook.

2 Theoretical Principles

2.1 Toyota Production System and Lean Philosophy

Sponsored with 5 million USD, a study was carried out at the Massachusetts Institute of Technology from 1985 to 1991 as part of the "International Motor Vehicle Program" to assess the future of the automobile industry [1]. 54 experts examined production processes in the automobile industry in 15 countries, including all the major automobile companies in the world. The study, led by James P. Womack and Daniel T. Jones, came to the conclusion that the Toyota Production System is far superior to classic mass production. The authors of the study coined the term "Lean Production" in connection with this production system. The production system can be described with the help of five basic principles: value, value stream, flow, pull, and perfection [2]. Another description of the Toyota Production System uses the model of a house, the top of which represents the ultimate aim of high productivity with a maximum of product quality and punctual supply [3]. To reach this goal, the house is supported by two columns built on a common foundation. One column, called Jidoka, represents early error detection and prevention by means of autonomation to meet the highest standards of quality. The other column represents the just-in-time principle, which ensures that only the right parts in the right quantity at the right time are supplied with an orientation towards customer needs. The columns stand on a foundation, which contains, among other things, standardized processes and leveled production. Continuous improvement is pursued by simultaneously concentrating on people and teamwork and by eliminating any form of waste. The success of this production system is based on diverse individual methods such as 5S, Kanban or Andon on the one hand and its focus on human factors on the other. For example, the problem-solving competence of employees and the leadership competence of supervisors are essential components of the production system [4, 5].

2.2 Digitalization of Industrial Production

For some years now, digital manufacturing has been a focus of production research worldwide. Large-scale initiatives have been launched to shape the digital transformation of industrial production. These initiatives originate partly from research-related government programs, and they are partly initiated by large corporations. These initiatives include, for example, “Industrie 4.0” (Germany) [6], “Industrie du Futur” (France) [7], the “Industrial Internet Consortium” (USA) [8], and the “Industrial Value Chain Initiative” (Japan) [9]. The contents of these and other initiatives for the digitalization of industrial production vary widely and are not entirely identical. However, they are all primarily aimed at permeating manufacturing with information technology. An important goal is to create continuous digital process chains in order to bring the complexity of a diverse production – up to a lot size of 1 – under control and to realize high productivity at the same time. To reach this objective, technologies are being developed that will, among other things, contribute to self-configuration, self-diagnosis, and self-optimization of technical systems in addition to their networking, communication, and self-controlling [10]. The result is a smart factory, in which cyber-physical systems autonomously interact with one another in the production process [11].

2.3 Human Factors in the Didactics of Industrial Engineering

There are different methods available in teaching to impart knowledge. These vary by the degree of participation of the student and the extent to which his channels of perception are addressed [12, 13]. Because methods of operational practice play a major role in the subject of industrial engineering (cf. Toyota Production System), business games are especially suitable for conveying teaching content. The teaching and learning method of the business game offers a variety of advantages compared with classic methods, in which communication tends to extend from the instructor to the learner in a one-sided way [14]. Experience-based learning in the context of a business game can link theory and practice, meaning that business games support a holistic understanding of concepts. Business games make it possible for the learner not only to be an observer but to directly recognize the consequences of his own decisions. With the dynamic involvement of participants, business games also provide the opportunity of changing attitudes, reducing resistance, and accepting and internalizing teaching content.

3 Teaching Concept for Imparting the Kanban Method

3.1 Context of the Developed Teaching Concept

The Industrial Engineering Laboratory at the Ostwestfalen-Lippe University of Applied Sciences and Arts contributes to the education of industrial engineers at this university with various teaching modules. One of these is the production systems module, in which the developed teaching concept is applied. Essential teaching contents of the module include the structure, content, and development stages of production systems. In the process, methods of requirement-based design and of maintaining and optimizing production systems are also taught.

On the research side, the laboratory is occupied with the user-centered, customer-oriented, and efficient design of working and production systems. This is done with an emphasis on the development and research of assembly and assembly assistance systems. As part of these activities, different demonstrators have been implemented to research various technologies from the context of digitalization in industrial production and to integrate them into education [15, 16].

In its integration of research into education, the Industrial Engineering Laboratory focuses on linking new digital technologies with the proven methods of the Toyota Production System. For example, a business game developed several years ago to teach the Kanban method, about which an article has already been published [17], has been supplemented with information technology components. The modified business game will be presented below. The extended business game can be used to exemplify how different production philosophies can be combined effectively.

3.2 Basic Structure of the Kanban Business Game and How to Play

As part of the production systems teaching module, students are taught various methods associated with the Toyota Production System. One of these methods is the Kanban method, as it is of major importance for controlling material flow in many operational production systems. The crux of the method is a material supply in manufacturing and assembly that is exclusively oriented towards actual consumption. The developed business game is played in three rounds. On this basis, further expansion levels are designed with other focal points. After every round of the business game and its expansion levels, an evaluation takes place with the help of key figures and/or a discussion of observations, so that the effects of each round or expansion can be seen directly.

The object of the Kanban business game is a diverse vehicle assembly that must be supplied with material from a warehouse. The vehicles are made of LEGO® bricks. There are six different roles in the business game. One role is that of the picker, who sorts the material required at the assembly workplaces in bins in a warehouse. Another player takes the role of the logistics employee, who supplies the assembly workplaces with the filled bins from the warehouse. Assembly consists of a total of three workplaces: a workplace for the assembly of the chassis (assembly group 1), one for the assembly of the vehicle cabin and front bumper (assembly group 2), and one for final assembly. The associated vehicle assembly groups must be manufactured including final vehicle assembly in different versions at all three assembly workplaces with the help of assembly orders and instructions. Each of the three assembly workplaces is taken over by one player. A sixth player assumes the function of disassembly and documentation. Figure 1 shows a graphic representation of the basic structure of the Kanban business game and examples of the LEGO® vehicles to be assembled.

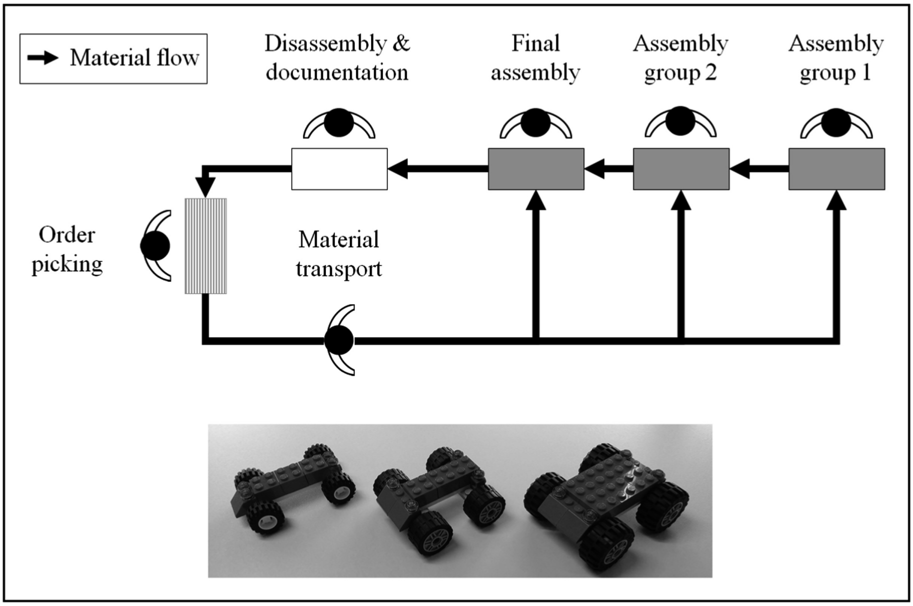


Fig. 1. Basic structure of the Kanban business game and example vehicles

3.3 First Round

The first round simulates a classic material supply scenario. In the process, a conventional production planning and control process is executed along the lines of the "push principle", according to which a central authority specifies the material supply. The flow of information and material proceed in the same direction. The picker takes preplanned and correspondingly sorted material supply orders for the planned production program from an order tray and processes them sequentially. Each material supply order lists the parts required to complete a specific assembly order at one of the three assembly workplaces. These parts are filled in bins in the given amount by the picker. When all the parts have been picked as per the material supply order, the picker places the material supply order on a staging area in the warehouse along with the full bins. The logistics employee then supplies the assembly workplace indicated by the material supply order with material and transports empty bins from the assembly workplace back to the warehouse as required.

The player at the workplace for assembly group 1 takes the top assembly order from the order tray. This order contains information about the vehicle model to be assembled and about the number to be manufactured. Only after the entire lot size specified by the order for the assembly group to be produced at the first workplace has been assembled, the work product is passed onto the following work station along with the order. Then the player at the first assembly workplace starts with the next assembly order. In the meantime, the player at the workplace for assembly group 2 takes the work product from the first workplace, executes the required assembly work as per the order, and, after the entire lot size has been completed, passes it onto the final assembly workstation, where the same principle is applied. Disassembly and documentation are carried out at the final workplace. Completed vehicles are chronologically recorded, disassembled, and the individual components are brought back to the warehouse. An announcement is made when 10 and when 20 completed vehicles are in outgoing goods. Each time, the in-process inventory in the assembly stage is then determined and documented. After 20 vehicles have been completed, the round ends. The change in the in-process inventory is represented graphically, and student observations are discussed as part of the first round. At this phase the students have the essential insight that the in-process inventory continuously increases, which means that individual assembly workplaces are being supplied with an ever-increasing amount of material that is not needed yet. This simultaneously results in downtimes and bottlenecks at individual stations.

3.4 Second Round

In the second round, material is supplied following the Kanban method. Here, the "pull principle" is applied. According to this principle, the material supply is controlled in a decentralized manner according to actual consumption. The flows of information and material proceed in opposite directions. For this purpose, the game structure is altered. A two-bin Kanban system is set up for every part and every assembly group required at the respective assembly workplace. The quantity of one container is determined such that it is sufficient for replenishing the second container with parts. Following the idea of Ohno [18], the bins at the assembly workplaces are sorted into supermarkets, from which the respective assembler can supply himself with the parts he needs at any time. The player roles are the same as in the first run, whereby the logistics employee is designated henceforth as the milk runner.

In the second round the process starts in the final assembly, in which the assembly order tray is now located. The player at this workplace takes the top order from the tray and carries out the final assembly of the product to be produced with the assembly groups and parts provided for him in the supermarket. Every product can be passed on directly to outgoing goods. One parts container is always completely emptied before the second is started. If a container is empty, it is immediately placed on the area for empty bins. Empty assembly group bins are placed directly in the return flow of the supermarket after the assembly group is removed. The empty assembly group container serves as an assembly order for the workplace for assembly group 2. The player at this workplace starts assembling the assembly group as soon as an empty assembly group container of the final assembly workplace is in the return flow of the supermarket. Every finished assembly group is sorted directly into the supermarket of the final assembly workplace. The remaining principles also correspond with those of the final assembly workplace. Empty assembly group bins are placed in the return flow of the supermarket directly after removal of the assembly group and serve as the assembly order for the workplace for assembly group 1.

The milk runner supplies the workplaces with material and carries empty Kanban bins back to the warehouse, where he inserts them into the FIFO track set up there according to the FIFO principle. He transports Kanban containers filled in the warehouse to the corresponding workplaces and sorts them there into the associated supermarket according to the FIFO principle. The picker takes the top empty Kanban container out of the FIFO track. The Kanban is attached to the container. The part indicated on the Kanban is taken out of the parts warehouse in the correct amount and fed into the container. As soon as a Kanban container is full, it is assigned to the associated workplace on a staging area by means of a colored marking.

At the final workplace, disassembly and documentation is carried out in the same way as in the first run: Completed vehicles are chronologically recorded, disassembled, and the individual components are brought back to the warehouse. An announcement is made when 10 and when 20 completed vehicles are in outgoing goods. Then the in-process inventory in the assembly stage is determined and documented.

After 20 vehicles have been completed, the round is ended. The in-process inventory in round 2 is represented graphically after 10 and 20 vehicles are completed and compared with the documentation of the in-process inventory from the first round. Student observations and findings are discussed in the context of the second round as well as in comparison with the first round. One important learning outcome for the students at this stage is that the in-process inventory in the second round levels off at a low level in contrast to the first round. In addition, the students experience through the practical implementation that the problems associated with large inventories no longer arise, and the assembly workplaces are, on the whole, working to equal capacity.

3.5 Third Round

The third round involves the digitization of the flow of information while maintaining the Kanban method. A pick-to-light system is introduced in the warehouse, which leads the picker through the process of filling the bins for the various assembly stations with the help of lamps with sensory properties and with the help of an operating panel. Furthermore, material bins equipped with RFID tags are used. The information stored on these tags can be read by a reading head, which shows the picker the item number, the target quantity, and the recipient of the commission via the operating panel. Requirement messages from individual assembly stations are now conveyed electronically by pressing a button and are visible for the milk runner by means of sensory lamps in the warehouse and at the assembly workplaces themselves.

The new processes in the game are as follows: The picker takes the top empty Kanban container out of the FIFO track and holds it up to the RFID reader. The information is read from the transponder and displayed on the operating panel. A lamp lights up in front of the corresponding parts warehouse. The quantity displayed on the operating panel must be taken out of the storage location displayed by the lamp and filled into the selected container. As soon as one Kanban container is full, the picker confirms this material supply order by touching the sensor lamp at the storage location and places the filled container in a chute with a colored marking for each assembly workplace. As soon as one parts container is empty at the assembly workplaces, it is immediately placed on the area for empty parts bins, and the button located at the workplace is pressed. This sends a requirement message to the milk runner in the warehouse, signaled by the lighting up of a lamp in the warehouse. The milk runner then goes to the workplace where the lamp is illuminated, confirms it at the assembly workplace and takes the empty container with him. After the milk runner has placed the container on the FIFO track in the warehouse, he also confirms the message lamp in the warehouse. If the message lamp goes out, there are no further requirement messages. If it lights up again, there are one or more further requirement messages from the assembly workplaces. In the third round, students learn that, by digitizing the existing business game infrastructure, information can be conveyed more quickly and errors avoided. Figure 2 shows students carrying out the digitized version of the Kanban business game.

Fig. 2. Digitized Kanban business game in action

3.6 Further Expansion Levels

Among other things, assistance systems are introduced in the assembly process in further expansion levels. For example, the students are tasked with carrying out a quality control on different versions of assembled LEGO® vehicles using an image processing system. To do so, they configure the image processing system by means of various software tools in such a way that the assembled vehicles can be identified as accurate or defective with the help of diverse characteristics. The first step for example is to apply a contour tool, a border texture tool, and a color testing tool to check that the tires are assembled correctly. This acquaints them with the strengths and weaknesses of tools and teaches them to identify the tool that is best suited for the application. After further exercise levels, the students identify characteristics, with which assembled vehicles can be distinguished from one another. They then configure the image processing system with the software tools at their disposal so that, among a selection of various vehicles provided by the trainer, only accurately assembled vehicles are identified as correct vehicles by the system. As in the third round, students have the important insight in the expansion level described that digitization in conjunction with the Toyota Production System can contribute to error prevention. Moreover, they test in detail how quality management processes in manual assembly can be supported digitally in a purposeful way.

4 Evaluation of the Teaching Concept and Assessment of Results

The digitized Kanban business game was evaluated in 2018 in two stages. First a test group of 10 students were surveyed for technical experimental purposes after implementing the business game. Then a further survey of a total of 28 students was carried out after playing the business game as a part of regular teaching in the production systems module. The questionnaire was structured in the same way for both surveys. It was subdivided into the components “Organization and structure of the business game”, “Execution of the business game”, “Learning success”, and “Overall assessment of the business game”. The questionnaire contained 27 items in total in all four of the above-mentioned areas. Of these, 20 items were Likert-scaled with five values and four items with three values. Free text answers were queried in three items.

Figure 3 shows an excerpt of the evaluation results of items with five values for use with the business game as part of regular teaching of the production systems module.

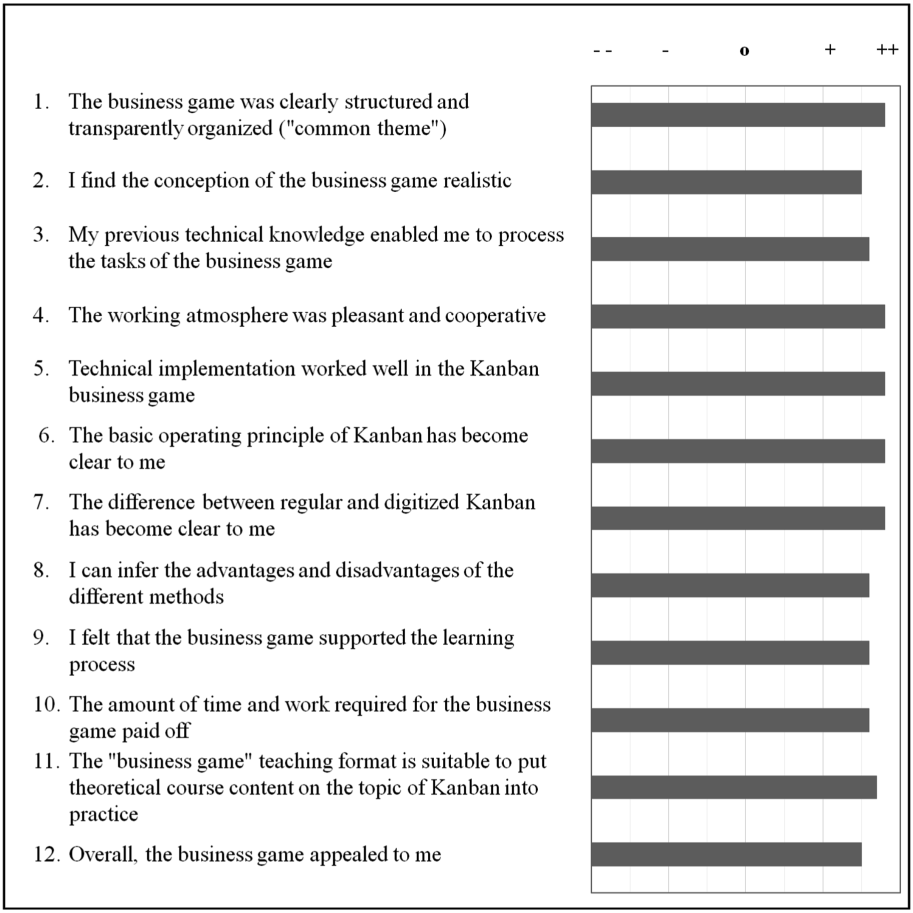


Fig. 3. Excerpt of the evaluation results of the digitized Kanban business game

Items from all four areas of the questionnaire were included. In the items with five values shown in Fig. 3, students found the business game to be clearly structured and realistic. Both the Kanban method itself and the difference between the regular and the digitized Kanban became clear to the students. The business game was perceived as being supportive for the learning process, and the selected teaching format of "business game" was found to be suitable for putting theoretical course content on the topic of Kanban into practice.

In the items with three values, the requirement level (exactly right (100%)) and the group size in the business game (exactly right (85.2%); too small (14.8%)) were, among other things, considered appropriate. In the free text fields, the students positively accentuated, among other things, the business game's well-defined structure, practical relevance and interactivity as well as the link between theory and practice. One negative note was that, in one case, the exercise group was too small, making improvisation necessary. Obligatory student enrollment was suggested. In addition, one free text answer made the critical observation that the role of vehicle disassembly and documentation is not directly integrated into the simulation of material flow and is therefore relatively unattractive.

5 Summary and Outlook

This article has shown, using an example from higher education instruction, how methods from the Toyota Production System can be combined with approaches of digital manufacturing. Students learn by playing, how technological approaches of digital manufacturing can be combined with the established methods of the Toyota Production System in a meaningful way. In the process, reference is made to both columns of the Toyota Production System. On the one hand, the Jidoka principle is taken account of – in error detection by means of image processing, for example. On the other hand, a preliminary stage of the just-in-time principle is realized with the introduction of the Kanban system. By digitizing the flow of information, information can be transmitted more quickly and errors can be prevented.

By enhancing the business game with information technology components, it is shown by example how information technology can support humans in performing activities. The evaluation of the business game that was conducted showed that the developed business game experienced a high level of acceptance and was perceived to be conducive for the learning process. In future, the business game will be supplemented with further topics and expansion levels in order to combine the organizational/personnel aspects of the Toyota Production System with aspects of digitization, thus laying the foundations for both highly productive and humane production.

Acknowledgments. This publication was created in collaboration with the "PRAXIS-OWL" project for practice-oriented and innovative learning in Ostwestfalen-Lippe, funded by the German Ministry of Education and Research (BMBF) (grant no. 01PL17048).

References

1. Womack, J.P., Jones, D.T., Roos, D.: The Machine that Changed the World: Based on the Massachusetts Institute of Technology 5-Million Dollar 5-Year Study on the Future of the Automobile. Rawson Associates, New York (1990)
2. Womack, J.P., Jones, D.T.: Lean Thinking: Banish Waste and Create Wealth in Your Corporation. Free Press, New York (2003)
3. Liker, J.K.: The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer. McGraw-Hill, New York (2004)
4. Atwater, J.B., Pittman, P.: PRODUCTION AND INVENTORY-we want to be Toyota-find out what the Toyota production system can–and cannot–do for your firm. APICS–Perform. Advant. **18**, 32–35 (2008)
5. Kleiner, A.: Leaning toward utopia. Strategy + Bus. **39**, 76–87 (2005)
6. "Industrie 4.0" project of the German Ministry of Education and Research (BMBF). http://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html
7. Alliance Industrie du Futur. http://www.industrie-dufutur.org/
8. Industrial Internet Consortium. http://www.iiconsortium.org/
9. Industrial Value Chain Initiative. http://iv-i.org/wp/en/
10. Oztemel, E., Gursev, S.: Literature review of Industry 4.0 and related technologies. J. Intell. Manuf. (2018). http://doi.org/10.1007/s10845-018-1433-8
11. Kagermann, H., Helbig, J., Hellinger, A., Wahlster, W.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group. Forschungsunion (2013)
12. Schelten, A.: Grundlagen der Arbeitspädagogik. Steiner, Stuttgart (2005)
13. Gasser, J.: Erfolgreich lernen – Praxis-Tipps für sofortigen und nachhaltigen Lernerfolg. Books on Demand, Norderstedt (2014)
14. Deshpande, A., Huang, S.: Simulation games in engineering education: a state-of-the-art review. Comput. Appl. Eng. Educ. **19**, 399–410 (2011)
15. Hinrichsen, S., Riediger, D., Unrau, A.: Development of a projection-based assistance system for maintaining injection molding tools. In: De Meyer, A., Chai, K.H., Jiao, R., Chen, N., Xie, M. (eds.) Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore (2017)
16. Hinrichsen, S., Riediger, D., Unrau, A.: Assistance systems in manual assembly. In: Villmer, F.-J., Padoano, E. (eds.) Production Engineering and Management. Proceedings 6th International Conference, Lemgo, pp. 3–14 (2016)
17. Brown, D., Hinrichsen, S., Paris, M.: Development of a business game for teaching the Kanban method. In: Schlick, C.M., et al. (eds.) Advances in Ergonomic Design of Systems, Products and Processes, pp. 103–114. Springer, Heidelberg (2017)
18. Ohno, T.: Toyota Production System: Beyond Large-Scale Production. Productivity Press, Cambridge (1988)

Towards a Mobile Assistance System to Raise Productivity in Maintenance

Michael Waechter(✉), Thomas Loeffler, and Angelika C. Bullinger

Chair for Ergonomics and Innovation, Chemnitz University of Technology, Erfenschlager Str. 73, 09125 Chemnitz, Germany
michael.waechter@mb.tu-chemnitz.de

Abstract. In the course of digitization, the availability of machinery and equipment data is becoming an increasingly important productivity factor. Mobile devices have the potential to provide maintenance-relevant information at any given time, context-specific and in real time. However, currently available industrial tablets are poorly accepted by maintenance workers. This paper shows that mobile assistance systems created in a user centered design process with maintenance workers included in the design process are significantly better accepted than current commercial alternatives. The presented prototype is a first comprehensive approach to a mobile assistance system that covers the requirements of maintenance workers in a production setting. The paper provides details on how the prototype developed and evaluated in course of the research holds potential to save time and thus increase productivity by obtaining information through the system. Furthermore, it is shown that the prototype provides a basis for useable mobile devices in an industrial context.

Keywords: Human factors · Systems engineering · Tangible HMI · User-centered design · Mobile assistance system · Maintenance

1 Introduction

In digitized maintenance as part of industrial production, status data, relevant information and available resources are exchanged directly between machines and systems. This allows inspections, servicing, and maintenance to be coordinated without human intervention [1]. The associated complexity of machines and production facilities are changing the factors influencing high production productivity. In the future, high productivity will depend less on the operator on the production line and much more on the availability of production equipment. This makes maintenance a key factor for high productivity in production [2]. Currently, relevant information for troubleshooting are often distributed across different databases. Therefore, maintenance workers currently need a lot of time to find causes and possible solutions to resolve unexpected issues [3].

In the wake of Internet of Things (IoT) technologies like mobile devices, new opportunities for increasing productivity in maintenance and challenges for creating human-machine-interfaces arose [4]. With the use of mobile devices on the shop floor, the maintenance engineer of tomorrow receives all relevant information, such as

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 56–64, 2020.
https://doi.org/10.1007/978-3-030-20040-4_6

machine condition, maintenance history and repair instructions, mobile and location-independent [5]. Initially developed for consumer use, the production environment has different requirements for the design of the human-machine interface. Currently, manufacturers of industrial tablets only consider industrial features, such as robust construction, dust and splash water protection [4]. Especially for mobile devices like tablets, tangible user interfaces (TUI), providing the haptic controls for manipulating the graphical user interface (GUI), and in extension tangible human-machine interfaces (tHMI), also including hardware elements for handling and operating the computing device, are to be designed [6].

In this paper, we show the current effort in obtaining relevant information for troubleshooting and show how this process can be made more efficient and thus more productive in the course of digitization. For this purpose, we integrate the requirements of the maintenance workers into the development of a mobile assistance system and show the degree to which currently available industrial tablets meet these requirements. By using a design science based approach, we design and evaluate a mobile assistance system user-centeredly, focusing on the subjective quality measures for the created artefact. Finally, we present the potential of saving time in obtaining information through the use of the mobile assistance system and the resulting increase in productivity.

2 State of the Art

The following section describes the basic tasks of maintenance and, based on this, a study to determine the context of use and the requirements of maintenance workers for a mobile assistance system. Based on these results, existing practical solutions in the form of available industrial tablets are examined and evaluated regarding to the requirements.

2.1 Requirements of Maintenance Workers to tHMI of Mobile Devices

Typical tasks of maintenance include servicing, repair, inspection and the improvement of machinery and equipment [7]. While servicing and inspection are usually predictable, machine and equipment malfunctions are rarely predictable and require a quick response to avoid production downtime. Performed tasks also include planning, execution, functional testing, completion notification and documentation of maintenance activities. This results in a set of requirements from the maintenance workers toward a possible mobile assistance system. In the following, we will focus on the hardware aspect of these requirements, in particular the tangible human-machine interface (tHMI).

To analyze the requirements, a task analysis and context analysis was carried out in companies of the automotive and automotive supply industry. Subsequently, the results were discussed in a focus group with maintenance workers.

The aim of the task analysis is to examine the typical tasks of a maintenance engineer while using a technical system. For this purpose, the workflows and their sequence, interfaces to technical systems and required user information are collected and analysed [8]. The context analysis follows 18 guided interviews [9] with maintenance staff and participating observations in the maintenance process.

Results of task analysis and context analysis can be summarized as follows (for details, cf. [10]): In order to repair a failed machine, maintenance personnel first collects all important information, e.g. previous incidents. Currently, desktop PCs are used to search for relevant information in various databases. In addition, maintenance departments have an offline archive with manuals for the individual machines and systems. As a result of the diverse information channels, maintenance personnel need about 25% of their working time to gather all the information they need to rectify faults [3]. Due to the location-independency of the task, there is currently little possibility to provide necessary information mobile, context-specific and in real time.

Data analyse of the interviews shows the maintenance workers' requirements for an assistance system. A robust construction, i.e. protection against water, oil and drops, intuitive handling and a tablet-like design with a display diagonal of eight to ten inches are the non-functional requirements for a mobile assistance system. In addition, there are the following functional requirements for the tangible HMI: ergonomic and safe handling, operation via touchscreen and physical operating elements, easy transport, the ability to place on flat surfaces and to attach machines and equipment during the maintenance process [11]. As ergonomic and safe handling are listed, too, suitable handles should be added to a mobile assistance system.

Available industrial tablets bring the characteristics of robust construction, oil- and water-resistance to the shop floor, but these do not suffice to reach acceptance by the maintenance engineers. The low acceptance results from the lack of consideration of the user requirements of maintenance workers regarding to functionality and the suitability of the tangible HMI. The following section will detail on this relation.

2.2 Analysis of Relevant Industrial Tablets

Based on the results of the requirements analysis, a systematic search identified existing solutions and experiences from literature and practice. For this purpose, available mobile devices were examined regarding to previously used types of handles and evaluated in the context of a focus group in terms of their advantages and disadvantages in a production use with product designers (N = 5).

The results of the analysis show that the examined industrial tablets fulfil the non-functional requirements of the maintenance workers, but all remaining requirements for the tangible man-machine interface remain widely unfulfilled except for the touchscreen (cf. Table 1). Although there are standards and guidelines in the literature for the design of ergonomic control elements on stationary systems and ergonomic handles of hand tools [12, 13], design guidelines for mobile devices and their specific requirements are still missing [14]. Therefore, a user-centered engineering process was developed and instantiated using the example of the mobile assistance system for maintenance workers.

Table 1. Analysis of available mobile devices for use in maintenance

	Non-functional requirements					Requirements of tangible HMI					
Industrial Tablet	Water protection	Oil protection	Drop protection	Display 8"/10"	Intuitive Handling	Touchscreen	Ergonomic Handles	Physical Controls	Function to transport	Function to place	Function to attach
 Zebra ET50 / ET55	●	●	●	●	◐	●	○	○	○	○	○
 Panasonic FZ-G1	●	●	●	●	◐	●	○	◐	○	○	○
 Xplore Motion F5m	●	●	●	●	◐	●	○	●	◐	○	○
 Kontron KT-RT-I5	●	●	●	●	◐	●	○	◐	○	○	○
 Getac RX10	●	●	●	●	◐	●	○	◐	◐	◐	○

3 Designing a Suitable Mobile Assistance System for Maintenance Workers

This section first introduces the methodological approach - a user-centered process for designing suitable tangible HMI - for developing a mobile assistance system for maintenance workers. Subsequently, the results from the applied design process - the function prototype as a basis for the following evaluation - are described in detail.

3.1 User-Centered Design Process

Following a design scientific approach, the problem of missing practical approaches to user-centered design of suitable tHMI was solved iteratively by the use of existing knowledge with scientific rigor. Derived from the engineers' requirements and existing models for user-centered development in the literature, an engineering process for designing usable tHMI was developed. This consists of the four basic elements of user-centered development analysis, design, prototyping and evaluation. These are iterated in the four phases ideation, conception, concretization and implementation (Fig. 1).

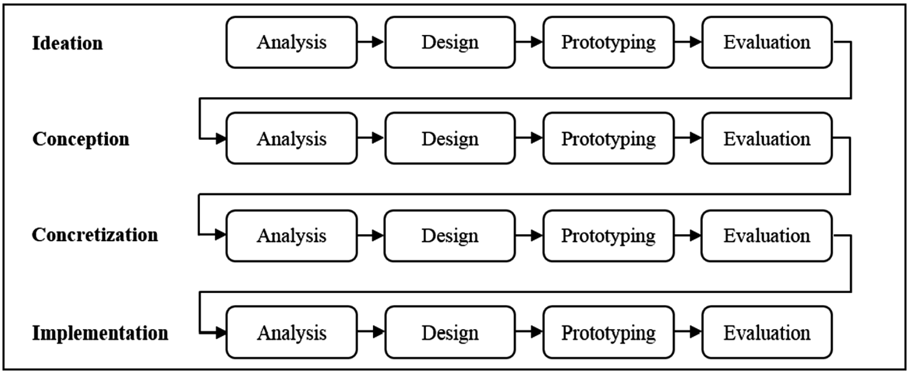


Fig. 1. Phases of the engineering process [11]

In the ideation, all relevant requirements for the tHMI are analysed first and various designs are designed using modelling clay. The concept focuses on designing a handle shape as a basis for safe and usable handling. In the concretization, the final grip form is created and all other requirements, such as adjustment function, are developed. Finally, the implementation develops a working model that prototypically implements all requirements and tests them in the field. Figure 2 shows the resulting prototypes as a result of the instantiated engineering process.

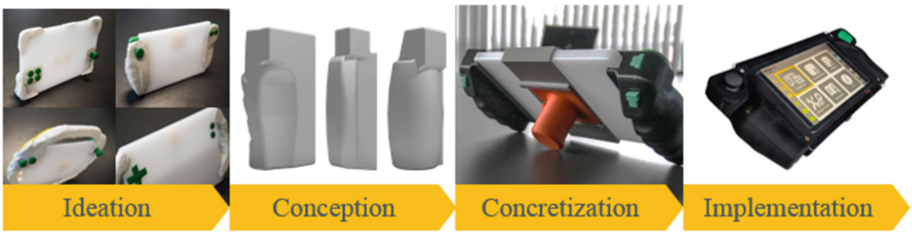


Fig. 2. Prototypes of the instantiated engineering process [11]

3.2 Mobile Assistance System for Maintenance Workers

The resulting functional prototype represents the first comprehensive approach to cover the requirements of the maintenance workers in a production setting as elicited in the requirements analysis. Figure 3 shows the final prototype as result of the applied design process.

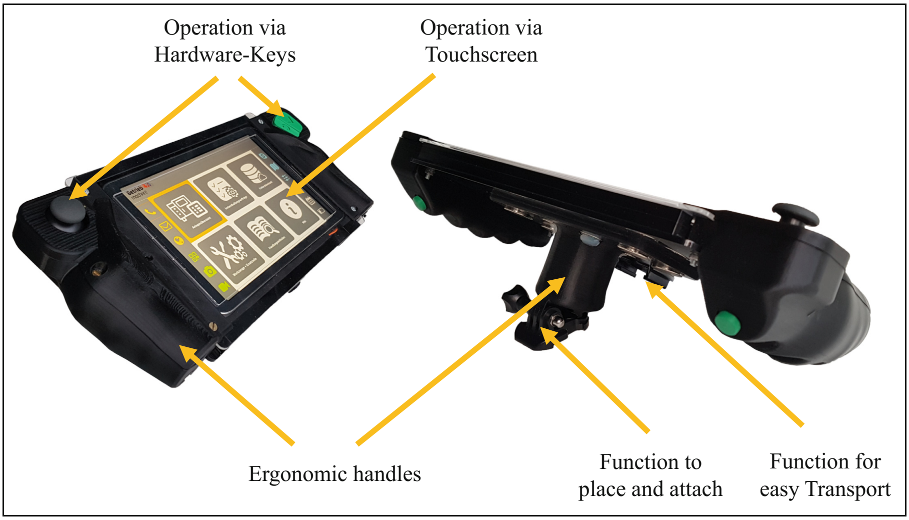


Fig. 3. Final prototype of the design process

In addition to anthropometric variables, the iterative handle design also takes into account comfort elements, such as the ergonomic finger indentation for the fingers, which ensure safe handling. With a removable one-handed handle on the back for touch operation, the mobile assistance system can be parked without tipping over on flat surfaces. A quick-release fastener on the back is used to easily attach the mobile assistance system to a toolbox or cart for transport. For the navigation of the software interface via hardware keys, the maintenance engineers choose a joystick instead of a control pad or thumb touchpad.

4 Evaluation

The evaluation of the developed mobile assistance system is implement in two steps. First, the suitability of the designed tHMI in the maintenance process is investigated. Subsequently, the changes of the maintenance process by using the mobile assistance system and the resulting savings potential are considered.

4.1 Mobile Assistance System

The mobile assistance system was evaluated for usability in a real maintenance setting. To compare the results with conventionally designed tangible HMI, maintenance workers (n = 20) go through the usability test twice. In a Within-Subject design, the subjects start randomized with a standard industrial tablet or the function prototype. After each run, the maintenance worker evaluates the tested mobile assistance system using the questionnaires SUS [15], AttrakDiff [16] and meCUE [17].

The results of the SUS show that maintenance workers perceive the functional model as highly suitable for use and, with M = 88.75 (SD = 7.27), significantly ($p < 0.01$) better than the standard industrial tablet with M = 68.25 (SD = 11.73). The evaluation of the AttrakDiff also results in a better evaluation of the functional prototype compared to the standard industrial tablet for both the hedonic and the pragmatic quality.

Based on the component model of the user experience of [17], the meCUE questionnaire contains four separately validated modules for product awareness, user feelings, consequences of product use and product attractiveness. The individual validation of the various modules makes it possible to use the meCUE flexible by using the modules individually (Fig. 4).

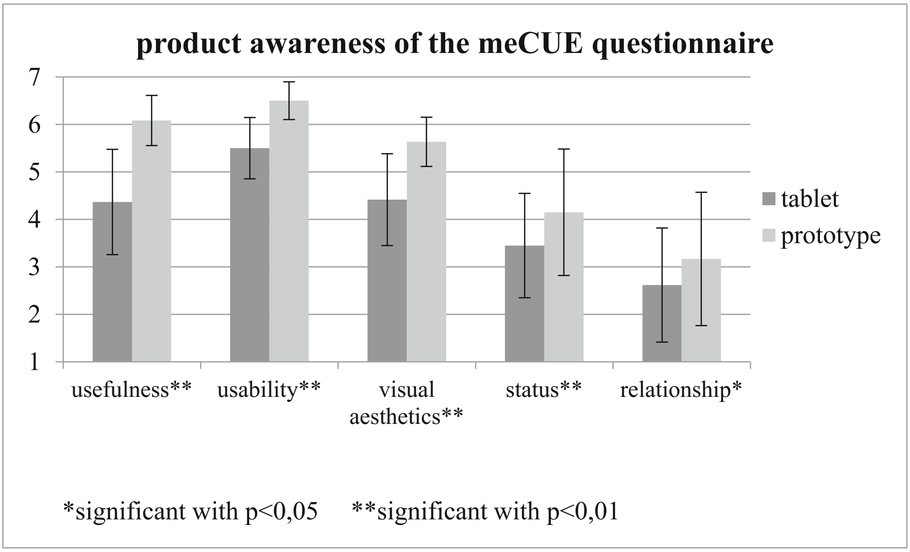


Fig. 4. Final prototype of the design process

The analysis of the meCUE questionnaire also reflect the results of the SUS. In addition to the usefulness (M = 6.08, SD = 0.53) and the usability (M = 6.50, SD = 0.40), the maintenance workers also rate the visual aesthetics (M = 5.63, SD = 0.52) and the status perception (M = 4.15, SD = 1.33) of the function prototype significantly better ($p < 0.01$) compared to the handling of the standard industrial tablet.

The results show a much better assessment of the user-centered mobile assistance system and the practicable applicability of the engineering process, which can also support future developers in the design of ready-to-use tHMI.

4.2 Maintenance Process

The maintenance time consists of the reaction, analysis and repair time. While a manual message is sent to the plant operator in the initial state, the response time can be shortened using the mobile assistance system by sending the automatic message directly to the maintenance engineer. For the analysis of the fault, the baseline scenario considers the search and combination of relevant information in the documentation, which can be depending on the frequency of work with the corresponding system. Using the mobile assistance system, action guidelines can be provided instead, which enables a targeted analysis. The savings potential of the analysis time is estimated to be up to 75% for the test case, but it depends heavily on the starting point of the system documentation and the way in which the information is prepared in the action guide. The recovery time depends on the type of fault and the underlying cause of the fault. If additional coordination with the system supplier is required, these will continue to depend on the availability of the respective contact person. Troubleshooting using the resource cockpit is supported by the guideline and, if necessary, by a wiki, so that in principle no further reconciliation is required. Overall, the reduction in maintenance time and for the scenario under consideration in the demonstrator can be estimated at up to 50% [3].

5 Conclusion and Future Work

With 25% of their working time, this paper addresses the substantial efforts of today's maintenance workers to obtain relevant information. The use of mobile assistance systems can reduce this time, as long as the mobile devices meet the requirements of the users and thereby be accepted. The analysis of available industrial tablets clarifies the gap between the requirements of maintenance workers to the tangible HMI and current designs. The user-centered design and evaluation of the functional prototype of a mobile assistance system shows the influence of the consideration of the users in the design of the tangible HMI on the perceived usability. Finally, the evaluation of the mobile assistance system in the maintenance process shows a savings potential of 75% in the search for relevant information.

Further research is required regarding the design of mobile devices for production in the consideration of age-dependent requirements for tangible MMS. In addition, there are a variety of other mobile technologies, e.g. glasses using Augmented Reality (AR) whose design has potential for discussion in the context of production use. AR glasses are still little explored and by the still high weight and unforeseeable consequences for the eyes, i.e. by accommodation, currently not suitable for use over the entire 8-h shift.

Acknowledgments. This research was partially supported by the German Federal Ministry of Education and Research (project Mittelstand 4.0-Kompetenzzentrum Chemnitz, 01MF16001A). The sponsor had no role in the study design, the collection, analysis and interpretation of data, the writing of the report, or the submission of the paper for publication.

References

1. Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T., Schlund, S.: Produktionsarbeit der Zukunft - Industrie 4.0. [Studie]. Fraunhofer Verlag, Stuttgart (2013)
2. Lee, J., Ardakani, H.D., Yang, S., Bagheri, B.: Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. Proc. CIRP **38**, 3–7 (2015)
3. Hopf, H., Krones, M., Mueller, E.: Modellierung und Optimierung von Instandhaltungsprozessen mit Sozio-Cyber-Physischen Systemen. In: S-CPS: Ressourcen-Cockpit für Sozio-Cyber-Physische Systeme. Abschlussveröffentlichung, pp. 1–22. aw&I - Wissenschaft und Praxis, Chemnitz (2017)
4. Gorecky, D., Schmitt, M., Loskyll, M.: Mensch-Maschine-Interaktion im Industrie 4.0-Zeitalter. In: Handbuch Industrie 4.0 Allgemeine Grundlagen. Allgemeine Grundlagen, pp. 219–236. Springer, Berlin (2017)
5. Scheer, A.-W.: Industrie 4.0. Wie sehen Produktionsprozesse im Jahr 2020 aus? IMC AG (2013)
6. Waechter, M., Bullinger, A.C.: Gestaltung gebrauchstauglicher tangibler MMS für Industrie 4.0 – ein Leitfaden für Planer und Entwickler von mobilen Produktionsassistenzsystemen. Zeitschrift für Arbeitswissenschaft **70**, 82–88 (2016)
7. DIN Standards Committee Services: Fundamentals of Maintenance. Beuth Verlag, Berlin (2012)
8. DIN Standards Committee Services: Resource-Saving Application of Methods and Tools for the Anthropocentric Design of Effective Interactive IT Systems. Beuth Verlag, Berlin (2016)
9. Deutsche Akkreditierungsstelle: Leitfaden Usability. Gestaltungsrahmen für den Usability-Engineering-Prozess (2010)
10. Waechter, M.: Engineering-Methode zur Gestaltung gebrauchstauglicher tangibler Mensch-Maschine-Schnittstellen für Planer und Entwickler von Produtionsassistenzsystemen. Chemnitz (2018)
11. Waechter, M., Hoffmann, H., Bullinger, A.C.: Towards an engineering process to design usable tangible human-machine interfaces. In: Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), vol. 825, pp. 136–147. Springer, Cham (2018)
12. Bullinger, H.-J., Jürgens, H.W., Groner, P., Rohmert, W., Schmidtke, H.: Handbuch der Ergonomie. Mit ergonomischen Konstruktionsrichtlinien und Methoden, Koblenz (2013)
13. DIN Standards Committee Services: Safety of Machinery - Ergonomics Requirements for the Design of Displays and Control Actuators. Beuth Verlag, Berlin (2010)
14. VDI The Association of German Engineers: User-Friendly Design of Useware for Machines - Design of Dialogues for Touchscreens. Beuth Verlag, Berlin (2004)
15. Brooke, J.: SUS-A quick and dirty usability scale. In: Usability evaluation in industry, pp. 4–7. Taylor & Francis, London [u.a.] (1996)
16. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Mensch & Computer 2003, vol. 57, pp. 187–196. Vieweg+Teubner Verlag, Wiesbaden (2003)
17. Minge, M., Thuering, M., Wagner, I., Kuhr, C.V.: The meCUE questionnaire: a modular tool for measuring user experience. In: Advances in Ergonomics Modeling, Usability & Special Populations, vol. 486, pp. 115–128. Springer, Cham (2017)

Smart Devices Evaluation and Dynamic Cognitive Assistance System for Repair Processes in Production

Rainer Mueller[1], Leenhard Hoerauf[1(✉)], Attique Bashir[1], Christoph Speicher[1], Matthias Vette-Steinkamp[1], and Ralf Mueller-Polyzou[2]

[1] Centre for Mechatronics and Automation Technology (ZeMA), Group of Assembly Systems and Automatisation Technology, Gewerbepark Eschberger Weg 46, Geb. 9, 66121 Saarbruecken, Germany
{Rainer.Mueller, Leenhard.Hoerauf, A.Bashir, Christoph.Speicher, Matthias.Vette}@Zema.de

[2] Institute of Product and Process Innovation (PPI), Leuphana University of Lueneburg, Universitaetsallee 1, 21335 Lueneburg, Germany
Ralf.Mueller-Polyzou@stud.leuphana.de

Abstract. Manual assembly needs to be able to react fast and flexible on product modifications at an increasing number of product variants. Cognitive assistance systems and smart devices can increase the efficiency and quality of this product assembly. However, both need to be validated for shop floor application in order to secure the operator's acceptance. Moreover, available assistance systems force operators to a predefined assembly path and do not support dynamic deviations. Hence, the present paper focuses on the evaluation of smart devices for assembly workstations and the development of a cognitive assistance system for dynamic repair processes. In a comparative study, a smartwatch, a wristband with gesture control and a voice control are compared with an IR remote control and evaluated using defined criteria. Furthermore, different smart devices are connected to an assistance system, which reacts flexibly to a specific error and acts situationally to deviations using a precedence graph.

Keywords: Smart devices · Assistance system · Assembly process · Repair process

1 Introduction

In the increasingly globalized business, producing companies can address new markets but also face additional competitive pressure. Both lead to a growing number of product variants, often lower prices, shorter product life cycle times and higher product quality as a distinguishing criteria [1]. In order to manage the high number of product variants and to react quickly and flexibly to product modifications, the assembly is still characterized by a high degree of manual tasks and a confidence in the experience and expertise of the assembly operators. Nevertheless, assembly operators can be assisted

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 65–77, 2020.
https://doi.org/10.1007/978-3-030-20040-4_7

by assistance systems to increase the efficiency and quality of product assembly. Additionally, producing companies expect economic, technical and operator-oriented advantages using intuitive operating concepts in combination with smart devices. Assistance systems are often combined with camera based inspection systems [2–4], so that not only additional information is provided for documentation but also in-line process verification can be performed [5]. These systems ease the work of the assembly operator on the shop floor and are particularly helpful in repair processes where the assembly not only depends on the product variant, but also on the present error [6].

However, the effectiveness, efficiency and user-friendliness of the applied smart devices has to be validated before the assistance systems are put into operation, otherwise assembly operators might not accept them. Moreover, commercially available assistance systems require operators to follow the instructions exactly and do not take any deviations from the specified assembly path into account. These deviations occur especially during repair processes when an assembly operator detects further failures during process performance and needs to deviate from the predetermined repair process.

Hence, the presented paper first focuses on the evaluation of smart devices. In a systematic evaluation study, intuitive operating concepts (a smartwatch, a wristband with gesture control and a voice control) are compared with an infrared (IR) remote control and evaluated in terms of industrial applicability and user-friendliness. The smart devices are used for switching manually trough assembly instructions given by a laser projector at a workstation. In the second part of this paper, the development of a cognitive assistance system for dynamic repair processes is described. This system consists of smart devices and a control unit. The control unit flexibly creates rework plans depending on the structure given by the precedence graph of a product variant and the specific error code. The assistance system is validated with an industrial use case of an automotive gearbox rework process using a laser projector and camera system.

2 State of the Art

The implementation of the evaluation study and the development of the cognitive assistance system is based on existing research work. For this reason, the following subchapters describe the state of the art for these topics. First, general assistance systems for assembly tasks are presented. Afterwards, the basic data used for generating and optimizing work schedules is outlined.

2.1 Human Assistance Systems

A multitude of different human assistance systems and technologies exist that allow adaptation to individual needs on the shop floor in plants. The information delivery and analysis capabilities of these intelligent systems, as well as the automation of simple, repetitive activities, can enable higher production efficiencies [7]. These include reducing errors during process execution, increasing process and product quality, reducing costs, increasing productivity, and shortening the time it takes to train new operators. In some cases, the introduction of technical systems leads to physical relief and even substitution of human workers. Nevertheless, operators will have a central role in future production, where they will be assisted by technical systems in carrying out their tasks [8].

In production, human assistance systems are technical systems that support assembly operators in their actions. By integrating sensors into these systems or networking them with higher-level IT systems, these systems become intelligent. They are able to react to situations independently and can therefore be individually adapted to the operator. In cognitive assistance systems, the perception or decision-making of the operator is supported as the classification in Table 1 shows. On the other hand, physical assistance systems support the execution of operator's work [9]. The focus of this paper is on cognitive assistance systems and especially on decision assistance systems.

Table 1. Classification and characteristics of assistance systems [8]

	Perception assistance system	Decision assistance system	Execution assistance system
Characteristics	Assists five human senses (in the industrial context: visual, auditory, and tactile perception are primarily relevant)	Assists the cognitive skills (supports mainly the development of skills during the process of problem solving and learning)	Assists the manual work of the operator (motion sequences are simplified by machines)
Embodiment	Human assistance systems	Software support of complex decisions, Prioritization of tasks	Exoskeleton, Human-robot collaboration
Classification	Cognitive assistance	Cognitive assistance	Physical assistance

2.2 Precedence Graph and Optimization

As the paper focuses on cognitive assistance in flexible working environments for example the rework area, the information for creating a work plan have to be stored. Part of this information is the product structure, process flow and required assembly resources. Thus, this section shortly introduces relevant theoretical graph and optimization aspects. A graph is a tool to represent relations between units as Fig. 1a shows. According to [10] a Graph G = (V, E) is an ordered pair of sets, with $E \subseteq V^{(2)}$ where V is a set of vertices and E is a set of edges. An Edge is often described by its two connected vertices, denoted as (u, v) [11]. Edges can contain further attributes such as weight (Fig. 1a) or even color [11, 12]. A usual way to depict a graph is using dots for vertices and lines for the edges. Regardless the presentation method, the main concern is to point out the information of which pairs of vertices form an edge and which do not [13]. Matrices or lists represent relations between vertices when it comes to storing data.

A graph with directed edges (Fig. 1b) is called a directed graph or simply a digraph D with D = (V, A) [11]. A is a set of edges also called arcs, whereas an arc **a** is defined as a = (u, v), where **u** is the tail (vertex) and **v** is the head (vertex) [11]. Graphs present

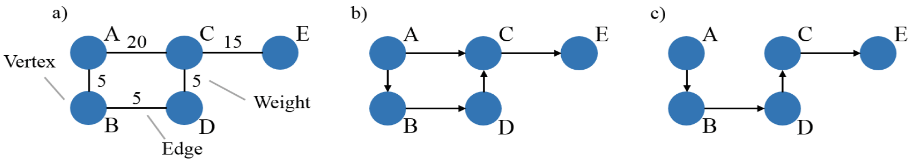


Fig. 1. (a) Directed graph (b) Directed graph (c) Subgraph of left graph

global connections within a limited environment. However, some applications require the consideration of particular parts of the graph. This excerpt of a graph is called a subgraph. A subgraph is a graph that consists of the vertices and edges from the original Graph. In Fig. 1c a subgraph of the graph depicted in Fig. 1b. Nonetheless, it does not need to include all vertices or edges, but a subset such that G'(G) = (V', E') with V' $\subset$ V and E' $\subset$ E [10]. A subgraph in a Graph G in the form S = (V, E) with V = {v_0, v_1,…, v_k} is called a Path in G [11]. A Path does not necessarily have to be a subgraph of G.

To find a path in a graph G, pathfinding algorithms can be used. Pathfinding algorithms use graph structure as well as graph properties to determine the path from a starting point to a finishing point. Common algorithms to determine the shortest path are Dijkstra, A*- as well as Bellman-Ford-Algorithm [14]. Since the Dijkstra algorithm is used in development and evaluation, it is explained below.

Originally, Dijkstra is a greedy single source algorithm and relies on weighted edges. The objective is to find the shortest path from node A to node F as seen in Fig. 2. The algorithm visits all vertices in the one-Step range (C, B) and notes the weights of the edge as cost. In Fig. 2b Path A to B is more cost-effective than A to C. In the next step, the most cost-efficient vertex is selected (A, B) as illustrated in Fig. 2c. Starting from this vertex all vertices in one-step range are visited (D) and again the costs are recorded. The vertices visited in the second cycle are at most two steps away from the starting point. The costs are summated. A check is performed whether the costs are higher than the second best results from the first cycle. If that is not the case, the second best vertex is selected from the first cycle and the algorithm starts anew. In Fig. 2 the path (A, B, D) is more expensive than (A, C). It turns out, that path (A, C, E) is more cost-effective than (A, B, D). The algorithm repeats until the finishing point is reached. The output is the optimal (shortest) path from a defined start to a defined finish. optimal (shortest) path from a defined start to a defined finish.

3 Evaluation of Smart Devices for Assembly Workstations

In the use case described in this paper, laser projection systems are used as an exemplary assistance system providing assembly instructions to the operators at the shop-floor. Laser projectors are mounted at the ceiling of the manufacturing hall or above the assembly workstation. Laser assistance systems consist of one or more laser projectors, a control software installed on a PC, power and data connections. A digital work plan is derived out of CAD data and defined by a sequence of polygons that is projected onto the assembly object [15]. The projection sequence thereby reflects the

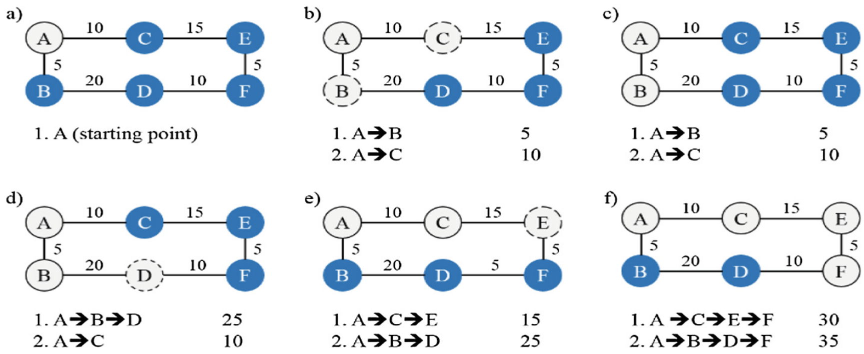


Fig. 2. Dijkstra pathfinding algorithm

assembly process. The laser projection software installed on a computer controls the system. Using the software GUI often requires walking between the product and the computer. The operator loses time and the work process is interrupted. Therefore, IR remote controls became the standard technology for the operation of laser assistance systems. They are cost-effective and robust in industrial use. However, even using the IR remote control still interrupts the work process. In addition, the remote control must be kept close to the operator, especially when working on large objects. The fact that IR remote controls must be oriented towards the laser projectors can also cause non-ergonomic movements. An evaluation study was conducted to evaluate and test appropriate smart devices to replace the IR control concept. The approach of the evaluation study and the study results are reported below.

3.1 Selection of Smart Devices

A market research on smart devices was carried out using previously defined criteria clustered into technical, commercial and user oriented categories:

- Technical criteria: degree of industrial protection, battery life, maximum reach of wireless connection, integration support, availability of API or SDK
- Commercial criteria: price, commercial availability, professional service
- User oriented criteria: user friendliness, supported commands, feedback signals, potential hindering of the work process, hygiene.

An assessment model was developed based on these criteria in order to benchmark the smart devices. Thereby, three smart devices were selected for further analysis:

- A smartwatch for direct and permanent use on the wrist. The smartwatch offers haptic and optical feedback options and is connected via WLAN.
- A bracelet gesture recognition device, which is attached to the forearm of the operator. Electrodes in the bracelet detect the movement of individual muscles and determine them as gestures. The bracelet is connected via Bluetooth. The bracelet gives direct user feedback in the form of vibration patterns.

- Voice control with headphones and microphone. The speech analysis software is installed on the system computer. The headset with microphone is connected via Bluetooth. The system supports language feedback.

3.2 Research Methods Applied

A survey research with a case study was used for the analysis. The case study represents an assembly task using the three smart devices and the existing IR remote control. The case study uses a shop-floor scenario and collects qualitative and quantitative data [16]. Data was collected with one participant expectation questionnaire, one participant questionnaire for every smart device, one final participant questionnaire and one observation questionnaire. The participant expectation questionnaire collected socio-demographic information as well as experience and expectations. The case study questionnaire collected data for the perceived usability, efficiency and effectiveness of the smart devices on an individual level. Additionally, product perception, user emotions and intention of use were analyzed. Specific questions were related to visual appearance, touch and feel, operability and intuitiveness of operability as well as wearing comfort. A combination of the After Scenario Questionnaire (ASQ), the System Usability Questionnaire (SUS) and the meCUE questionnaire with a 5-Likert Scale without N/A option was applied [17–19]. Participants were observed during task completion. The time needed for task completion as well as the number of incorrect inputs, positive and negative comments as well as frustration statements were noted. Potential preferences and improvement suggestions were requested in the final participant questionnaire concluding the case study. The comparative approach allows for the analysis and evaluation of the data including a benchmarking of the smart devices.

3.3 Test Environment and Case Study Execution

The test environment for the case study consisted of an ergonomical assembly workstation (Fig. 3). The three smart devices were integrated into the central projection software of the laser assistance system. The participants were given the task to complete a seven-part square puzzle with the outer dimensions of 50 × 50 cm (19.7 × 19.7 in.). The puzzle with four distinguishable colors is large and straight-edged. The puzzle represents a two-dimensional, reproducible assembly task with a low degree of difficulty in order to focus on the analysis of the system operation. The corresponding digital work plan was stored in the laser projection system.

At the beginning of the case study, the participants were introduced to the smart devices and the laser assistance system. Afterwards, the participants completed the *expectation questionnaire*. The participants then independently tested the control concepts. While performing the task, the participants were observed using the observer questionnaire [23]. The observer also assessed the participant´s appearance during the task in terms of sovereignty and handling. After task completion, the participants filled-in the *operating concept questionnaire*. In conclusion, the participants filled in a *final questionnaire* with their decision regarding one of the operating concepts.

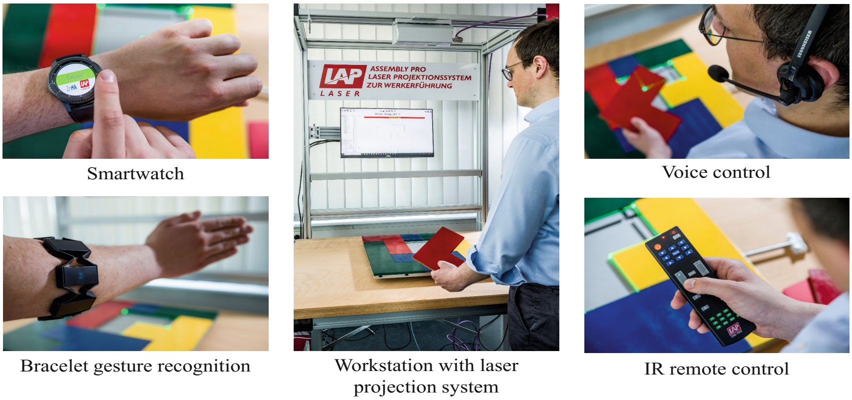


Fig. 3. Test environment for the case study and the used smart devices [20]

3.4 Result of the Case Study

In total 30 employees of four companies participated in the case study. 25 participants completed the questionnaires (sample group n = 25). The average age of the participants was 34 years, among the participants were seven women.

Focusing on the new operating concepts, 80% of the participants were familiar with the use of the smartwatch and therefore it represented the best-known operating concept. 56% of the participants stated that they already had experience with voice control. Before the testing, 45% of the participants opted for voice control, 35% for the smartwatch and 20% for gesture control. Expectations of potential advantages and disadvantages of the operating concepts were stated (Table 2).

Table 2. Advantages and disadvantages of the individual concepts

	Smartwatch	Gesture recognition	Voice control
Advantages	Familiar from using conventional watches	Hands-free operation	Hands-free operation
	No additional device	No additional device	Easy and intuitive operation
Disadvantages	Distracting additional apps	Pressure marks	Wearability
	Battery duration	battery duration	Influenced by ambient noises
	Small display	Misinterpretation	Problems with dialects

The participants needed on average 83 s to complete the task using the gesture control in comparison to smartwatch 61, remote control 55, voice control 60 s. Participants made most input errors using the gesture control (average 4.41). This is also reflected in frustration statements. Most positive statements were about the voice control (average 0.75). The participants worked confidently with the smartwatch and the voice control.

With the exception of the gesture control, the participants indicated for all operating concepts that they could fulfil the task. In terms of visual perception, the participants preferred the smartwatch. The look and feel were rated best for the voice control. With regard to operability, the remote control, smartwatch and voice control hardly differed, whereas the gesture control was rated significantly worse (Fig. 4). In terms of intuitiveness, voice control was preferred by the participants. The smartwatch was rated highest in terms of wearing comfort.

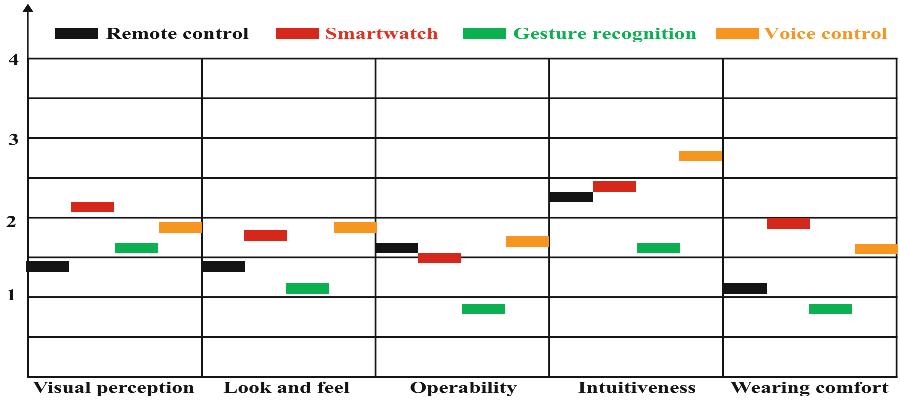


Fig. 4. Result of the evaluation according to the evaluation criteria

Furthermore, the perceived effectiveness (task completion) and efficiency (required time) were analyzed. The effectiveness of the voice control was rated 10% higher in comparison to the IR remote control, and 12% higher in terms of efficiency. The gesture control reached a perceived effectiveness of −47% and an efficiency of −50% compared to the IR remote control.

The total SUS usability score confirms the results. The gesture control reached with 33.3 out of 80 possible points a significantly low value. The three other operating concepts, remote control (51.5), smartwatch (53.5) and voice control (58.4), were perceived good in terms of usability.

To assess the user experience, the participants evaluated the operating concepts in terms of their usability, usefulness and positive and negative emotions. For both operability and utility, the voice control is rated highest with 3.07 and 2.65, respectively, of a maximum of 4 points. Most positive emotions were stated using the smartwatch (1.54 points), most negative emotions by the gesture control (2.08 points).

Overall, the gesture control is on average rated worst with −1.24 points and the voice control is rated best with a mean of 2.52 points. The smartwatch reaches 2.29 and the remote control 1.61 points.

Consistent with the previous results, 44% of the participants opted for voice control in the final questionnaire and 24% opted for the existing IR remote control or the smartwatch. As reasons for their decision for the voice control the participants indicated that the voice control was reliable and allows hand-free work. Overall, voice control is rated highest by the participants, while gesture control is rated worst.

4 Cognitive Assistance System for Dynamic Repair Processes

The use case is settled at the rework area of a transmission manufacturer where faulty products are repaired. Because of the various (error and variant dependent) rework plans, the process is performed at manual workplaces. In contrast to the final assembly, this causes higher process times and lacks process security. To assure that the transmission is in a salable state, repaired transmission are checked again. Due to the time consuming rework, errors need to be detected on the spot rather than in the test bench. To increase productivity and ease the operator's job, cognitive assistance systems are introduced. The following chapters describe the creation of a rework plan based on the overall product graph. The rework plan is a subgraph of the overall graph describing all product states.

4.1 Creation of a Graph for Rework

Since the rework relies on high flexibility and decision-making capability, qualified operators are indispensable for its success. Hence, flexible and cognitive assistance systems are in research focus for this area. The starting point for creating a graph for rework plans is the bill of material (BOM) which considers multiple product levels like assembly groups and parts [21]. In the next step, an assembly order is defined. The resulting graph represents the product structure in the vertices whereas the edges symbolize the process to get into a product state. In rework area, processes are separated in three segments (disassembly, repair and assembly) as shown in Fig. 5a.

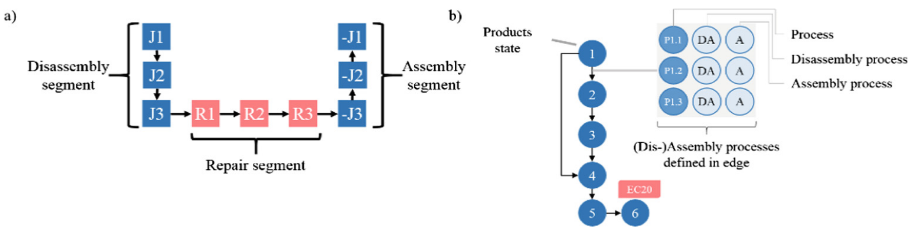


Fig. 5. (a) A standard rework plan consists of three segments (disassembly, repair and assembly), (b) Processes defined in the edges of the product state graph

Before the cause of error is fixed, the product is disassembled to a product state where e.g. the faulty product part is accessible. After executing the repair work, the product is reassembled. The process is fed by process data (Fig. 5b) from the classes:

information, execution, checking and documentation [6]. The data class *information* contains data for the operator, for example a process description or projection data. *Execution* data are machine-readable data which is used to parametrize an intelligent resource with these data to fulfill a (semi)-automated process. The class *checking* contains data to verify a process and product quality. After the verification, the assistance system decides whether a process is finished and starts the next process or the system remains in its current process. The last data class refers to the *documentation* after a successful verification and defines which data needs to be stored. Json-files are used to record this structure, allowing clustering data in objects [22]. In Fig. 5b an error code flag is illustrated (EC20). This is necessary to determine the product state up to which a product is disassembled in order to access the faulty part and perform the repair.

4.2 Derivation of an Initial Rework Plan and Its Dynamical Adaption

In the previous steps, a graph was defined that posed as the basis for a precedence graph (Fig. 5). In most cases, many possibilities exist to reach a specific product status. To select the most efficient path, a shortest path algorithm is applied with weights added to the edges. Paths along assembly groups that are preferred are weighted with lower numbers than the edges describing the handling of single parts within assembly groups. To apply a pathfinding algorithm, a start and finish has to be determined. The rework starts typically with a completed product, which is the initial state. The finish is defined by flags in the graph, hereafter denoted as error codes. By entering the error code the shortest path to the EC corresponding process for disassembly is calculated using the Dijkstra algorithm. The error code is associated with the repair part of the rework plan. Once the disassembly path is known the mirrored processes are selected for the assembly path. Between both paths, the actual repair is inserted.

During a rework plan the experienced operator might detect another failure cause, in that case he can overrule the system and insert a new error code. This changes the way how the rework plan is composed, because a new path has to be calculated. This path might involve going backwards in the product states (assembly) as Fig. 6 shows for an example. Based on a product state reference graph a rework plan for the repair case for error code 20 (EC20) was calculated. The worker decided in process step 5 to cancel the rework plan and insert a new error code (EC30) as finish point. Before applying the shortest pathfinding algorithm according to Dijkstra, the new starting point has to be calculated. A process step is needed which allows to reach process step 10. The algorithm checks whether process step 10 is directly reachable from the current process step (process step 5). If it's not possible the algorithms takes a step back to the previous state unless the process step is reachable. In Fig. 6, process step 10 is not reachable from process step 5, so the system takes three steps back to process step 2. Process step 10 is reachable from process step 2. The new paths therefore consists of the back steps (5-4-3-2), the disassembly (7-8-10), repair (processes associated to R30) and assembly part as well as a path back to the completed product (10-8-7-2-1).

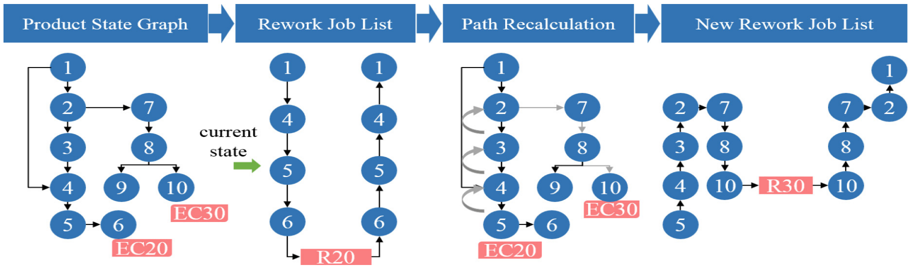


Fig. 6. Recalculation process for new rework plan

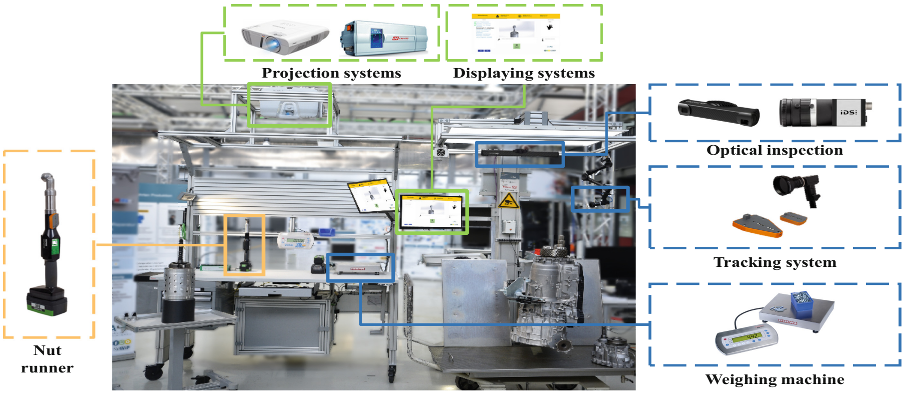


Fig. 7. Assisted workplace with different intelligent resources

4.3 Structure of the Pilot Environment

The pilot environment to verify the approach consists of two workplaces equipped with sensors and (semi-)automated resources (Fig. 7). The desk-less workplaces contains the workpiece carrier with the transmission. At the desk-bound workplace, the handling of assembly groups takes place and disassembled parts can be put down. The resources of the assistance system are classified according to the data classes. It is possible to allocate a resource to multiple data classes. A displaying system is the main information source for the operator. It is realized as worker guidance system: a GUI which lists all processes to carry out, a process animation and further information. It is supplemented with a laser projection system that projects geometries on the product and indicates where parts are to be inserted e.g. bolts. Since the processes are carried out mainly manually there are just a few execution resources. The nut runner, poses as execution resource as well as checking resource. It is parametrized with defined settings to tighten a bolt. It also checks if a tightening process is within the parameters and verifies it. The camera realizes the optical inspection and supports the completeness check. By predefining reference images, the camera captures the live process constantly and checks if a part is inserted or not. The tracking system is used for access control. A scale is used to verify the disposal of a faulty parts, which are not meant to be assembled again.

5 Summary and Outlook

Qualitative and quantitative research results were presented in the area of new operating concepts for manual assembly systems using systematically identified smart devices. A cognitive assistance system was developed and depicted supporting rework processes that are characterized by dynamic deviations.

In a comparative study, a smartwatch, a wristband with gesture control and a voice control were evaluated with an IR remote control. The study indicates that voice control is the preferred smart device for laser projectors in large-scale assembly. Furthermore, different smart devices are integrated into a cognitive assistance system. The system reacts flexibly to a specific error code and acts situationally to deviations using a precedence graph and optimization algorithm to adapt the rework plan.

Summarizing it can be stated that both new control concepts and cognitive assistance systems increase the efficiency and quality of product assembly and repair processes at high product variants and manual assembly. Future research work will focus on validation of the developments in industry and an integration of further functionalities. Furthermore, the results of the study will be integrated in the cognitive assistance systems, for the example the voice control as additional smart device.

References

1. Vogel-Heuser, B., Bauernhansl, T., ten Hompel, M. (eds.): Handbuch Industrie 4.0 Bd. 1. Springer, Berlin (2017)
2. Automations praxis: Handmontage wird digital, Montage-Assistenzsysteme bieten Unterstützung und Qualitätssicherung. https://automationspraxis.industrie.de/industrie-4-0/handmontage-wird-digital/
3. Bosch Rexroth AG: Werkerassistenzsystem ActiveAssist
4. DE software & control GmbH: Intelligent vernetzt – das modulare Assistenzsystem für individuelle Prozesse
5. Taubert, J., Rehe, M., Müller-Polyzou, R.: Application of camera controlled laser projection systems for manual mounting tasks. In: Technische Unterstützungssysteme, Dritte Transdisziplinäre Konferenz, pp. 67–76 (2018)
6. Müller, R., Vette-Steinkamp, M., Hörauf, L., Speicher, C., Bashir, A.: Worker centered cognitive assistance for dynamically created repairing jobs in rework area. Proc. CIRP **72**, 141–146 (2018)
7. Weidner, R., Redlich, T., Wulfsberg, J.P.: Technik, die die Menschen wollen, Unterstützungssysteme für Beruf und Alltag - Definition, Konzept und Einordnung. In: Weidner, R., Redlich, T., Wulfsberg, J.P. (eds.) Technische Unterstützungssysteme, pp. 12–18. Springer, Heidelberg (2015)
8. Bundesministerium für Bildung und Forschung. https://www.wissenschaftsjahr.de/2018/neues-aus-den-arbeitswelten/das-sagt-die-wissenschaft/low-cost-automation-eine-alternative-zur-menschenleeren-fabrik/
9. Reinhart, G. (ed.): Handbuch Industrie 4.0, Geschäftsmodelle, Prozesse, Technik. Hanser, München (2017)
10. Joyner, W.D., Melles, C.G.: Adventures in Graph Theory. Springer, Cham (2017)
11. Rahman, M.S.: Basic Graph Theory. Springer, Cham (2017)
12. Lewis, R.M.R.: A Guide to Graph Colouring. Springer, Cham (2016)

13. Diestel, R.: Graph Theory. Springer, Berlin (2018)
14. Erciyes, K.: Guide to Graph Algorithms. Springer, Cham (2018)
15. Müller-Polyzou, R., Maerterrer, J.: Aus der virtuellen CAD-Welt in die Produktion. Der Konstrukteur **4**, 94–96 (2018)
16. Yin, R.K.: Applications of Case Study Research. SAGE, Los Angeles (2012)
17. Lewis, J.R.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies. SIGCHI Bull. **23**, 78–81 (1990)
18. Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B.: Usability Evaluation in Industry. CRC Press LLC/CRC Press, London/Bristol (2014)
19. Michael, M., Riedel, L.: meCUE - Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In: Boll, S. (ed.) Mensch et Computer 2013 - Tagungsband. 13. fachübergreifende Konferenz für interaktive und kooperative Medien; interaktive Vielfalt, pp. 89–98. Oldenbourg, München (2013)
20. LAP GmbH Laser Applikationen, Borowiakziehe KG, Eric Shambroom Photography
21. Swamidass, P.M. (ed.): Encyclopedia of Production and Manufacturing Management. Springer, New York (2006)
22. Müller, R., Vette-Steinkamp, M., Hörauf, L., Speicher, C., Bashir, A.: Information and data structure to create flexible work plans for worker assistance systems at rework site (2018)
23. Wiesbeck, M.: Struktur zur Repräsentation von Montagesequenzen für die situationsorientierte Werkerführung. Utz, München (2014)

Management of Productivity in Smart Manufacturing/Industry 4.0

Epistemic Debt: A Concept and Measure of Technical Ignorance in Smart Manufacturing

Tudor B. Ionescu, Sebastian Schlund, and Christina Schmidbauer(✉)

Institute of Management Science, Technische Universität Wien,
Theresianumgasse 27, 1040 Vienna, Austria
{Tudor.Ionescu,Sebastian.Schlund,
Christina.Schmidbauer}@tuwien.ac.at

Abstract. This paper introduces the notion of *epistemic debt* as an analytical tool for understanding and managing the effects of technical ignorance in smart manufacturing. Drawing on the concepts of *technical and social debt* from software engineering, the metaphor of epistemic debt refers to the implied long-term costs of rework (e.g., redesign, replacement, reconfiguration or systems and/or organizational structures) caused by a lack of understanding and/or means of knowing the internals of complex software-based manufacturing systems essential to the value chain and core business of an organization. After defining the concept, we identify three of its sources and propose strategies for coping with epistemic debt in manufacturing.

Keywords: Epistemic opacity · Generative entrenchment · CPPS · Secrecy · Collaborative robots · Machine learning

1 Introduction

With the ongoing increase of autonomy through the introduction of cyber-physical production systems (CPPS) and machine learning (ML) in manufacturing, we observe a deepening of the knowledge divide between workers, technical experts, and systems of varying degrees of autonomy induced by the intractability of the operational decision rationales of complex ML models and algorithms. Consequently, the division of tasks between humans and CPPS endowed with artificial intelligence is likely to produce a new kind of ignorance (or non-knowledge) affecting both workers and technical experts, which—we believe—needs to be timely recognized and addressed as a risk for manufacturing companies.

Drawing on the notion of technical debt from software engineering and examples from industry projects and our own research, this paper introduces the notion of *epistemic debt* as an analytical tool and qualitative measure of the degree of technical ignorance entailed by the increased reliance on third-party software in manufacturing. To better defined the concept and phenomenon, we identify and describe three of its sources affecting collaborative human-robot applications: *secrecy, epistemic opacity, and generative entrenchment*. While *epistemic opacity* refers to the intractability of

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 81–93, 2020.
https://doi.org/10.1007/978-3-030-20040-4_8

complex computer models and algorithms in operation, *secrecy* refers to closed source hardware, software, and data, the details of which remain unknown to their users. *Generative entrenchment* refers to the degree to which long-lived software artifacts determine the workings of complex automation systems, whereby the purpose, meaning, and function of such artifacts may be forgotten over time. Each of these categories will be substantiated with examples from the domain of collaborative manufacturing robots. In response to these challenges, we propose some strategies for coping with epistemic opacity in manufacturing companies, which complement existing approaches to managing technical ignorance in organizations.

2 Epistemic Debt: Definition and Sources

The Industry 4.0 vision projects that CPPS will revolutionize the manufacturing domain by enabling a new automation paradigm grounded in the exploitation of the newest software and Internet-enabled smart manufacturing technologies. Within this vision, software plays an essential role—perhaps more than it did before in automation. Software is regarded as an enabler for new sociotechnical production ecosystems, which comprise so-called prosumer actors, both offering and consuming different goods and services that are important for manufacturing value chains. Yet the CPPS paradigm starts with the shopfloor, where two work worlds—that of software and that of industrial engineering—appear to clash. Recent studies have shown that, in practical contexts, the promises of CPPS reach the limits imposed by the physical environment of the factories—the most common sites of manufacturing [1, 2]. Overcoming these physical limitations by technical means is usually deferred to a future in which machine learning technologies will allegedly be able to compensate for that which current CPPS cannot—precision, self-organization, various optimizations, etc.

This apparent clash of engineering cultures provides us with an opportunity to address a series of issues pertaining to what we shall call *epistemic debt* in smart manufacturing. Drawing on the concepts of technical[1] and social[2] debt from software engineering, the metaphor of epistemic debt refers *to the implied long-term costs of rework (e.g., redesign, replacement, reconfiguration or systems and/or organizational structures) caused by a lack of understanding and/or means of knowing the internals of complex software-based manufacturing systems essential to the value chain and core business of an organization*. By analogy with financial debt, technical, social, and epistemic debt generate costs owing to a kind of figurative “interest” which companies need to pay for being indebted—that is for not having eliminated the causes and effects of technical ignorance at the right time. Because of the increased reliance on software in manufacturing, epistemic debt exists in organizations at any given time and cannot

[1] Technical debt (also known as design debt or code debt) is a concept in software development that reflects the implied cost of additional rework caused by choosing an easy solution now instead of using a better approach that would take longer [3].

[2] Social debt is analogous to technical debt in many ways: it represents the state of software development organizations as the result of “accumulated” decisions. In the case of social debt, decisions are about people and their interactions [4].

be eliminated. As a result, smart manufacturing companies become vulnerable to path dependence—for example, on expensive external experts and third party software vendors whenever systems might need to be reconfigured, updated, upgraded or when they fail. Under the current market pressure to further automate manufacturing processes in an increasingly competitive global market, the epistemic debt induced by novel software-intensive systems needs to be managed and mitigated continuously. This calls for a mindset that fosters reflexive design thinking and constructivist learning approaches in order to reduce technical ignorance in the company, as will be discussed in Sect. 4 of this article. Before discussing strategies to cope with the phenomenon, it is important to understand how epistemic debt emerges and which phenomena can be regarded as its sources:

Secrecy—This is perhaps the most common source of epistemic debt. It refers to the inability of the members of an organization to look into the source code and other internal details of the software systems that are commonly used in the company. While this is an accepted business rule protected by intellectual property laws, in the context of manufacturing, software is also used for controlling and actuating robots and other shopfloor technologies. Therefore, not having access to the internals of a software component may prevent its users from leveraging its full potential and preventing eventual undesirable effects, like software errors and defects.

Currently automation software is being at least partly written by the experts of the company that uses the respective software. Examples for this situation are MES systems (e.g., Siemens WinCC OA) and conventional robot programming (e.g., Universal Robots, KUKA, ABB, etc.), which often use proprietary languages. On the one hand, this induces vendor dependencies since companies need to train experts in mastering programming languages used in the context of that particular software alone. On the other hand, these vendors provide guarantees that the APIs (application programmable interfaces) used by company experts are reliable, deterministic, and will not change dramatically from one version to another. These aspects are also ensured by a series of software-related industry standards for the manufacturing domain.

With the current software development and maintenance methods and practices, which invert the control over different products so as to allow vendors to update the software at arbitrary times and for seemingly reasonable yet—for end users—often irrelevant purposes, secrecy becomes a serious matter for the smart manufacturing domain as well. While, automatic software updates are still rare in operational automation software, smartphones and tablet-like devices, which are subjected to automatic updates and data leaks, have become commonplace at least in some branches of the manufacturing industry. This means that, while in general, old technologies are relatively well understood thanks to the experience of experts in manufacturing companies gained over many years of use, newer Internet-enabled software technologies may induce epistemic debt owing to ignorance with respect to what is being processed locally and in the cloud; and how to control data leaks and automatic updates. In other words, epistemic debt emerges seamlessly with the introduction of new software, which is being developed according to the newest trends and practices of the software industry.

In practical terms, owing to the current trends in cloud-based computing towards offering *software as a service*, industrial software is being increasingly operated in the cloud instead of on local devices because the cloud business produces more revenues for software vendors and cloud-based software is easier to maintain. In this context, new paradigms such as *edge* and *fog* computing rely on rather complex, often non-transparent deployment schemes, both on local devices and in the cloud. This may lead to more epistemic debt on the part of customers since end users (i.e., manufacturing companies) have little or no control over the data and algorithms once stored and executed in the cloud. This situation might be one of the reasons for why, in spite of the immense hype created around the Industry 4.0 vision, most manufacturing companies seem rather reluctant to adopt new technologies in spite of the sometimes evident benefits they might provide. Dealing with secrecy as a source of epistemic debt is thus not really a matter of creating awareness about a new phenomenon; for, experienced automation and industrial engineers surely realize the impact of secrecy on the day to day business in an operation factory. Instead, it is a matter of balancing innovation needs with the knowledge in the company and potential risks (such as that of losing control over one's own software infrastructure) in the wake of the introduction of new technologies.

Epistemic Opacity—In the philosophy of simulation, epistemic opacity refers to the inability of a cognitive agent to know all the epistemically relevant elements of a computational process at a given moment because "[t]he computations involved in most simulations are so fast and so complex that no human or group of humans can in practice reproduce or understand the processes" [5]. As Humphreys further remarks [5], the cognitive agent faced with the issue of epistemic opacity may be an individual or a group. Therefore, the concept is also applicable to an entire organization. While the reference to the philosophy of simulation may seem far stretched in the smart manufacturing context, it must be noted that simulation plays an increasing role within the CPPS paradigm and the Industry 4.0 vision. The latter even assigns an important cognitive role to models of real systems, based on which designers can learn more about the system being developed before it is actually physically created and installed in a factory [6]. In some application contexts, such as collaborative robotics, simulation and reality are increasingly being aligned—a practice also known as digital twinning (or, previously, co-simulation). As it is often the case in commercial software development, the first working prototype end up being integrated into customer products. This is to say that, if development starts in simulation, most if not all the components developed in the simulation environment end up being used in operational systems as well. In other cases, it may be the explicit intention of developers that the digital twin of a system share the same components with the operational physical system.

The epistemic debt arising from the duality of simulated vs. operational environments is perhaps best exemplified by path planning algorithms for collaborative robotic arms, such as the Franka Emika Panda or KUKA iiwa. The planning algorithms provided, for example, in the Open Motion Planning Library (OMPL) use different heuristics, which perform many thousands of calculations per second (depending on the capacities of the underlying hardware). These heuristics are neither deterministic nor intuitive for humans because they work with discretized mash or graph like data structures. In this mode of representation, paths represent a series of waypoints, through which the end effector of a robotic arm is supposed to pass. In this context, the same algorithms and code used by various robot simulators (like RViz or Gazebo) is also used to plan real robot movements. As we will discuss in the next section, in operation due to heuristic path planning, collaborative robots often get stuck when carrying out even simple motions, leaving operators with little or no clues as to why this is happening. Because not all robot users are knowledgeable of the internals of motion planning theory and software libraries, they often do not understand why this happens. Motion planning may thus be regarded as an epistemically opaque software function of industrial robots and currently even mature robots warn users of the possibility of robotic arms to choose different paths each time, which shows that robot safety and the epistemic opacity of path planning are intrinsically linked.

Some constructive critics of the newest applications enabled by machine learning (e.g., autonomous driving, predictive maintenance, automated diagnosis of medical conditions, etc.) have pointed out that epistemic opacity is a major obstacle on the path towards certifying such applications in safety-relevant contexts. This arguably also affects the manufacturing domain. In this context, experts identified a new desirable architectural quality of software-intensive systems relying on machine learning and/or complex communication paradigms (e.g., peer-to-peer), called “explainability,” which refers to the epistemic accessibility of the internals of a possibly complex trained machine learning model or another type of (distributed) software component. Explainability is especially relevant for Deep Learning networks comprising millions of interconnected neurons, where the same phenomenon as in the case of complex simulation occurs—a cognitive agent is not able to know all the details of the computational process at any given time during the execution of that process. Moreover, in the case of trained ML models, this problem is amplified by the current lack of formal understanding of how such models make decisions. Explainability, in this context, would require the model to elucidate its own mode of “thinking” in a tractable way, so that the reasoning process can be completely and securely logged in order to facilitate objective post-factum analyses. Ensuring explainable software-based decision processes in all circumstances is a prerequisite for safety certifications.

Epistemic opacity differs fundamentally from secrecy because, in the former case, the source code may be open. This means that, in theory, anyone can look into the code and try to understand what is going on in certain situations. Yet with complex algorithms performing thousands of computations per second, as a phenomenon, epistemic opacity prevents users from knowing all the details of the process as it happens. One way of dealing with this challenge in the motion planning case would be to use precomputed paths plans between predefined locations in space, which can be thoroughly tested before being used in operation. However, this strategy gives rise to another source of epistemic debt—referred to as *generative entrenchment*—also relevant for trained machine learning models used in operational systems.

Generative Entrenchment—The term generative entrenchment has been introduced by Wimsatt [7] in reference to a feature of a structure, a biological system, or culture as one “that has many other things depending on it because it has played a role in generating them” (p. 133). Other scholars (e.g., [8]) have used the notion in relation to software, observing that developmental and evolutionary perspectives on software have revealed processes in software creation and maintenance reminiscent of those widely recognized to contribute to the evolution of biological and cultural organisms, organizations, and structures. Indeed, operational software today encompasses artifacts (code, configuration files, third party components, etc.), the meaning and purpose of which may have already been forgotten by the teams that created them. This may occur in the lack of proper documentation or other means of preserving the understanding or the respective function or component. That is to say that both code and data, which play an important role during the development and operation of a system (i.e., they are not redundant), may—at some point—outlive their creators and slip into oblivion. When that happens, entrenched software components are often regarded by contemporary developers and maintainers of a system as legacy artifacts, which nobody wants to “touch” because of a lack of a deeper understanding of what they do. As we showed elsewhere [9], precomputed parameters in simulation software systems have led to erroneous results produced by a safety-relevant disaster management system in Germany, which remained undetected for years.

In CPPS, software is expected to become much more abundant than in today’s factories. This development coincides with that of what some authors refer to as “Software 2.0” defined, for example, by Meijer as “programming using learned [ML] models” as opposed to “Software 1.0” defined as “programming using handcrafted algorithms” [10]. These concomitant transformations increase the risk of generative entrenchment to contribute to epistemic debt in manufacturing companies looking to keep the pace with Industry 4.0 and other innovation trends. Conversely, generative entrenchment may also contribute to the reasons why some companies are reluctant to introduce new systems in their day-to-day business. Either way, as time goes by and the number of software systems in industrial use increases, epistemic debt due to generative entrenchment is likely to become a more serious issue in future, especially when the first generation of so called self-learning systems will become legacy systems themselves. Who can guarantee that a data-driven system, the workings of which has been determined over generations by the internal features of (deep) machine learning networks or other structures, which may not be transparent even to their creators, who—in term—may not be available to account for anymore? The same question can be posed in relation to simulation software, which is known to be complex and to play an increasing role in “generative design” and digital twins. With the roll-out of more ML models as part of Software 2.0 configurations, generative entrenchment is likely to become an issue equally important as epistemic opacity. That is because data leaves less traces than code since, by definition, data is not memorized by people but by machines.

3 Epistemic Debt in Action: A Hands-On Example

To exemplify the sources of epistemic debt just described, we draw on our experience with an ongoing contract research project for an industrial customer conducted in the Aspern Industry 4.0 pilot factory in Vienna. For this project, a collaborative robot was used to partly automate an assembly process, which is currently being carried out manually by workers. The goal was to assess the dynamic shared tasks approach [11] of task allocation. The manual process consists of several joining and handling tasks. As part of these tasks, four small bolts are screwed on a component to join it with four smaller components. In addition, some handling tasks are required to provide the five components, the four bolts and a screwdriver to enable the joining. A Franka Emika Panda lightweight robot was chosen because it is easy to program, enabling even non-experts to configure and operate it. The robot comes with a task-oriented proprietary programming interface called “Desk” (shown in Fig. 1). In this web-based environment, several basic motion types are provided as applications (“Apps”) that can be combined and executed as a task. Programming the Panda robot is thus similar to operating a smart phone or tablet.

Secrecy—The main advantage of this task-oriented programming system is its ease of use, which comes at the cost of a *closed source* product. It is also part of the business model of the robot’s producer to not provide their software as open source code, as they offer additional and individualized apps on demand. Their standard package, which consists of a standardized robot and its software, only provides the most basic types of movements. The standard apps are described in the manual but only on a very superficial level. Implementation details remain at the producer’s knowledge. This is in fact also to the advantage of users, who are forced to learn by trial and error instead of following the same procedure all the time. Thanks to the high usability of the robot’s multimodal teach-in system, the producer can retain the possibility to upsell additional closed-source apps and hardware components to interested customers, thus creating the premises for vendor lock-in.

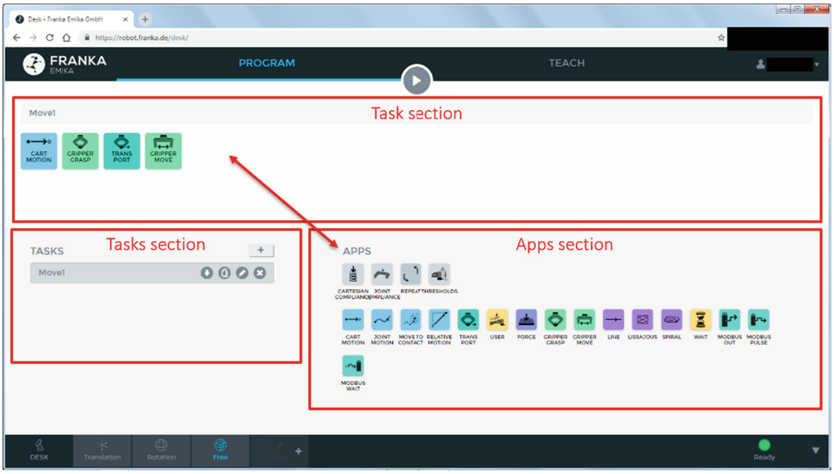


Fig. 1. The “Desk” programming interface of the Franka Emika Panda robot.

The epistemic inaccessibility of the robots soft- and hardware is also a challenge concerning when it comes to changing individual positions of a task or to process the produced data. For example, it is important for every part being manipulated by the robot to have a precisely defined position in order to repeat a task that has been programmed or taught to the robot. Even a small deviation can cause collisions, imprecise end effector positions or a total task failure. To position the parts correctly, some sort of fixture is necessary, e.g., using fitting insets on a shadow board. Problems may occur if some part of the taught-in process changes. In this case, every position has to be taught again, because the programmed process is not transparent to the user, including the taught-in positions and transitions from one motion (i.e., app) to another. The Desk environment of the robot does not allow the extraction of the coordinates of the taught-in points from the programming interface. In this context, more transparency concerning the internal parameters of the robot (notably the end effector positions) would enable a better collaboration between the robot and the human. While the provision of more detailed information regarding the state and position of the robot should not affect usability, users need to have access to these data when problems occur.

This example shows how secrecy induced by closed source apps and purposeful information hiding (i.e., end effector coordinates) may contribute to epistemic debt in a company since users need to pay "interest" in the form of a need to repeat the same teach-in procedure over and over again in order to automate even simple assembly processes. To overcome these impediments, the project team decided to implement a positioning assistance system for the robot, which uses the C++ API of the robot and provides information on a hand-held or wearable device, such as a smartphone or smart watch. The resulting additional development effort may be regarded as a form of interest paid by the team caused by the vendor's strategy to hide the position information and other data in the off-the-shelf Franka Emika Desk environment.

Epistemic Opacity—Another example from the same project underpins the challenges caused by the intractability of the complex path planning models and algorithms used by most robotic arms available on the market, including the Franka Emika Panda. During development, the robot executed a previously taught-in task but stopped immediately for no obvious reason, with the programming interface reporting an arbitrary error and the lights of the robot changing from green to red. In this context, it was not clear why that happened because, previously, the robot was able to reach every position on the taught-in trajectory. With the robot arm being able to move in seven axes, it can reach targets using a variety of possible trajectories for each of its joints. This is an advantage when there is an obstacle in the way. However, in the previous situation, the robot reached the limits of one of its rotation axes. Even though a path works out in the teaching process, it can happen for the path to fail when the robot tries it autonomously. When that happens, the robot arm needs to be placed in a balanced positon or, in the worst case, the path has to be reprogrammed. One workaround to this problem would be to set several waypoints along the path by configuring several Cartesian move apps in sequence for the same movement.

This experience shows that epistemic opacity can cause a less efficient robot programming and operation process. Path planning for collaborative robots and for industrial robots is very different. No ordinary programming knowledge is necessary to plan the path of the collaborative robot Panda. Using any of the provided motion apps (e.g., "Joint Motion", "Cart Motion", "Transport" etc.) individual paths can be planned. The path is defined by at least two points (start and destination). If there are only these two points defined the robot will go the shortest way to reach the destination point. Yet, for users this path is unpredictable since the robot uses heuristic path planning algorithms from the Open Motion Planning Library (OMPL). Finding the optimum number of waypoints is challenges for an operator even if one understands how heuristic path planning works. This is because, when the heuristic executes, the operator (i.e., the cognitive agent) cannot track all the details of the process. The intractability of the algorithm thus forces users into a trial and error mode of teaching the robot, which is seemingly less efficient than one might achieve if path planning would be deterministic (i.e., the exact same path would be computed between two given points).

Heuristic path planning algorithms thus create problems when it comes to certifying human-robot applications in real factories. Yet, a simple solution to this issue is not in sight because the inverse kinematic algorithms used in current motion planning libraries attempt to solve a complex combinatorial optimization problem. Therefore users as cognitive agents cannot know all the details of the path planning process at any given moment of that process. Consequently, they cannot predict if and when the arm might get stuck during operation, especially after several movements have been performed. The interest arising from this problem is paid in form of restricting the use of cobots to very simple tasks and an ineffective cooperation with humans, preventing a more economically-effective use of the robot.

Generative Entrenchment—In the project being discussed, a generative entrenchment scenario may occur due to the limitations of the Franka Emika Desk environment. Here, the different tasks (i.e., assembly sub-processes) are represented as linear chains or workflows of configured apps, the parameters of which—however—are non-transparent to the user. To understand what the workflow does, users need to execute the workflow using the robot, since a simulation environment is not coupled with the Desk environment. In this context, the Desk environment may also induce generative entrenchment because, once implemented, working robot or human-robot processes would be epistemically opaque to robot operators, who would need to re-program a complete workflow in order to change its behavior. Hence, there is a risk that some processes might be left “untouched” for a long time, only to be found unusable at some point. In such a scenario, interest is paid in form of the effort needed to re-implement an entire process due to its lack of transparency. Also, robot operators may be reluctant to change workflows for a long time, although they might not be optimal, in order to avoid “paying” that interest. Hence, some processes may become generatively entrenched, with “new generations” of operators facing the choice of re-implementing them completely or just leaving them untouched.

4 Strategies for Coping with Epistemic Opacity

In spite of the risks they entail, companies need to profit from new technologies, such as Software 2.0 and machine learning. With epistemic debt being unavoidable, individual and organizational users should learn how to minimize interest payments instead of eliminating its sources. At times, it may also be possible to prevent increases in epistemic debt through architectural decisions, such as choosing one system over another as well as isolating epistemically opaque systems so as to avoid ripple effects in case of failure or path dependencies. In this context, this section proposes some strategies for coping with epistemic debt:

Non-Ignorant Design—In the design phase, the functions, interfaces, and design patterns determining the architecture of a system are defined. To reduce system complexity and conjoined risks of overconfidence, the *partitioning* and *modularization* of processes defines potential insecure, critical or unknowable tasks. The *encapsulation* of functionality together with the specification of defined interfaces and monitoring parameters isolates potentially critical functions. As a result, the system contains a smaller number of critical processes, subjected to intensive *monitoring*, *supervision* or (redundant) *safeguarding*. Regarding safeguarding, different approaches are employed to design, develop and implement redundant systems (*dissimilar designs*). As a result, what might be termed *non-ignorant design* enables the monitoring of interfaces between modules of third parties and own modules, respectively between legacy systems and new modules. It reduces complexity and provides a first level protection of modules by using modular monitoring routines. Furthermore, the potentially non-understandable system behaviour is reduced to a small number of (risky) processes. Modularization and structured interconnection allows the use of less complex mental models to understand the subparts and sub-problems of a complex system using a structured learning path.

The exploration of smart manufacturing approaches through pilot applications (pilot use cases or pilot lines) as a first level of implementation represents one way of applying the non-ignorant design strategy in practice. Pilot lines are limited in their functionality, their scope and their impact in terms of potential risks. They do have defined interconnections to their surroundings and allow an observation of a system's performance and impact within a limited scope. A significant number of smart manufacturing projects follow the approach of containing (or isolating) new, potentially risky technologies by evaluating them in a pilot area prior to rolling them out in an entire factory. Like in the example described above, pilot areas may be regarded as constructivist learning environments [12], where instead of conveying information to learners in a positivist manner (i.e., by instructors and learning materials telling learners what to do and how to do it), learners explore solution approaches by themselves or in cooperation with other peers in their attempt to solve the problems inherent to the introduction of new technologies into an operational factory. These environments could benefit from an organization similar to that of public fablabs (fabrication laboratories) or makerspaces, which may require manufacturing companies to designate a permanent piloting area in a factory, where employees can explore and learn about the novel technologies that are about to be rolled out on the shopfloor.

Traceable and Tractable Operations—Focussing on the use of smart manufacturing systems and the underlying risks of secrecy, epistemic opacity and generative entrenchment, strategies of non-ignorant operations—that is, traceable and tractable operations—aim to monitor actions and their impact as well as to adopt different viewpoints on the processes being monitored. The monitoring of events and data storages using log files offers a potential for systematic examination of events and effects and their possible root causes regarding potentially secret system parts (e.g. of third parties). It may be conducted continuously, periodically, or at defined milestones in terms of simulated system behaviour (ex ante) or in terms of a failure analysis or a near accident (ex post). Through different approaches (i.e., prescriptive or descriptive), the use of root-cause-effects-analysis allows to a certain degree the reverse engineering of unknown parameters, thus reducing secrecy and epistemic opacity. It also allows backtracking in case of unexpected system behaviour, localization of critical configurations and their possible impact as well as stepwise expansion of mental models based on single source of documented system behaviour and event analysis. Other non-ignorant operation measures, such as audits, reviews, assessments or so-called 2nd opinion reviews, can also be used mainly ex ante.

Reciprocal Learning—As the expected changes in competency development due to smart manufacturing are widely discussed, there is a need to transfer novel training concepts into industrial use. Besides fostering of self-confidence to retain and improve learning curves for blue-collar and white-collar employees, the following situations represent challenges to learning in smart factories:

- *Larger scope*—Due to higher automation and increasingly autonomous technical systems, the average staffing per machine decreases. Hence, the number of processes to be mastered by the remaining employees is increased.
- *Fewer learning opportunities*—Due to the fact, that machines take over routine tasks and the resulting focus of humans is put on non-routine tasks, less learning opportunities with respect to routine processes exist for human operators [13].
- *Uncertain role of human work in hybrid (human-machine) settings*—Due to collaborative tasks with machines and algorithms, additional requirements in terms of learning emerge (Fig. 2).

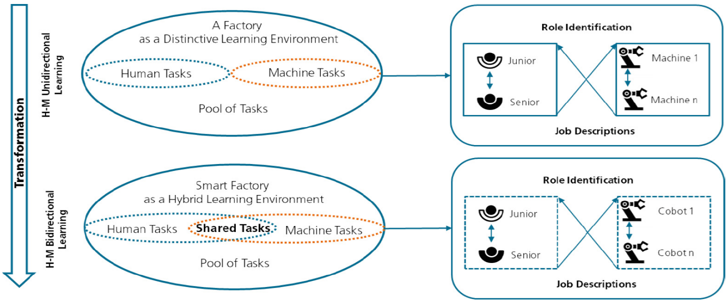


Fig. 2. The concept of reciprocal learning [14].

The *reciprocal learning approach* [14] might become necessary in hybrid (human-machine) settings in a smart factory. This approach uses human experience and tacit knowledge to train machine data sets (machines learn from humans) and on the other hand, employs data-based learning that is guided by smart algorithms (humans learn from machines). As depicted in the following figure, the latter divides the entire pool of tasks into 3 clusters [14]: (i) tasks assigned to the human workforce, (ii) tasks assigned to (intelligent) machines, and (iii) shared tasks assigned to both human workforce and intelligent machines such as collaborative robots or assistance systems. Reciprocal learning between humans and machines working on a set of shared tasks extends traditional (positivist) training approaches (usually based on predefined training routines and learning paths) and enables human-machine-conversations. It fosters awareness of secret or hidden features, while enabling confident actions in unexpected, challenging, or inexplicable situations; and problem solving in unstructured, unknown, or uncertain situations.

5 Conclusion

This paper argued that epistemic opacity and technical ignorance are emerging issues in the context of manufacturing, which need to be addressed as part of an effort to realize the 'smart factory' vision. Whilst not claiming to be exhaustive, the proposed analytical framework be used by researchers and practitioners from the smart manufacturing domain to manage and mitigate the sources of epistemic debt in their own organizations. In this sense, we proposed strategies to prevent overconfidence in software-intensive, Internet-enabled shopfloor technologies. In out future work, we aim to further elaborate the proposed constructivist learning approaches to the end of enabling researchers and end-users to cope with epistemic debt in practical smart factory settings. In particular, our goal is to explore the potential and limits of the proposed analytical framework in the Aspern Pilot Factory Industry 4.0 in Vienna together with leading manufacturing companies from central Europe.

Acknowledgements. This research was supported by the by the Austrian Research Promotion Agency through the HCCPPAS grant and the "Trust in Robots" Doctoral College of the TU Wien.

References

1. Ionescu, T.B.: When software development meets the shopfloor: the case of industrial fablabs. In: Proceedings of the 40th International Conference on Software Engineering: Companion Proceedings (ICSE 2019). IEEE (2019)
2. Ionescu, T.B.: Developing software for the shop-floor on the shop-floor. In: Proceeding of IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE). IEEE (2019)
3. Kruchten, P., Nord, R.L., Ozkaya, I.: Technical debt: from metaphor to theory and practice. IEEE Softw. **29**(6), 18–21 (2012)
4. Tamburri, D.A., Kruchten, P., Lago, P., Van Vliet, H.: Social debt in software engineering: insights from industry. J. Internet Serv. Appl. **6**(1), 10 (2015)
5. Humphreys, P.: The philosophical novelty of computer simulation methods. Synthese **169** (3), 615–626 (2009)
6. Ionescu, T.B., Merz, M.: Cyber-physische Produktion: Modelle und Inszenierung der Smart Factory. Arbeits- und Industriesoziologische Studien **11**(2), 247–261 (2018)
7. Wimsatt, W.: Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality. Harvard University Press, Cambridge (2007)
8. Kelty, C., Erickson, S.: The durability of software. In: Kaldrack, I., Leeker, M. (eds.) There is no Software, There are Only Services, pp. 39–56. Meson Press, Lüneburg (2015)
9. Ionescu, T.B.: Simulation, epistemic opacity, and 'Envirotechnical Ignorance' in Nuclear Crisis. Minds Mach. **29**, 1–26 (2018)
10. Meijer, E.: Behind every great deep learning framework is an even greater programming languages concept. In: Proceedings of the 26th ACM European Software Engineering Conference and Symposium on the Foundations of Software Engineering, p. 1. ACM (2018)
11. Ansari, F., Erol, S., Sihn, W.: Rethinking human-machine learning in industry 4.0: how does the paradigm shift treat the role of human learning? In: 8th Conference on Learning Factories 2018 - Advanced Engineering Education & Training for Manufacturing Innovation. Procedia Manufacturing, vol. 23, 117–122 (2018)

12. Wilson, B.G.: Constructivist learning environments: case studies in instructional design. Educational Technology (1996)
13. Baxter, G., Rooksby, J., Wang, Y., Khajeh-Hosseini, A.: The ironies of automation … still going strong at 30?. In: Turner, P., Turner, S. (eds.) Proceedings of the 30th European Conference on Cognitive Ergonomics, New York, USA, pp. 65–71 (2012)
14. Ansari, F., Hold, P., Mayrhofer, W., Schlund, S., Sihn, W.: AUTODIDACT: introducing the concept of mutual learning into a smart factory industry 4.0. In: Proceedings of 15th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2018), Budapest, Hungary (2018)

Industry 4.0 Technology and Employees Behavior Interaction in Serbian Industrial Companies

Vesna K. Spasojevic Brkic(✉), Zorica A. Veljkovic, and Andrija Petrovic

Faculty of Mechanical Engineering, University of Belgrade, Belgrade, Serbia
{vspasojevic,zveljkovic,aapetrovic}@mas.bg.ac.rs

Abstract. On the threshold of the 4th industrial revolution this survey is aimed to survey technology and employees behavior interaction, due to the fact that previous research has shown that this topic is not surveyed enough. For instance, recent advances in ICT have allowed manufacturing enterprises to use remotely controlled decentralized manufacturing processes and have led to development of different types of interfaces in the framework of industry 4.0. This paper examines data collected in 119 industrial companies. Modern technology latent variable is described by technological dimensions such as level of automation, level of IT application, and technological level against competitors, while employees behavior concept is described by employees' pro-activity/reactivity, interpersonal relations, collectivism/individualism and decentralization. Statistical examination included descriptive statistics, correlation analysis, factor and reliability analysis. Results show that variables are appropriately described with significant correlation between examined constructs. It could be concluded that upgrading of technological levels forces employment of proactive people with soft culture in industrial companies.

Keywords: IT · Automation · Technological level · Employee behavior

1 Introduction

Ever since the beginning of industrialization, technological leaps have led to paradigm shifts so today we are in the 4th industrial revolution period. It means that on the basis of an advanced digitalization within factories, the combination of internet technologies and other future-oriented technologies in the field of "smart" machines and products makes a new fundamental paradigm shift in industrial production [1].

Accordingly, in order to remain competitive in a turbulent and very competitive environment, contemporary industrial companies are forced to search for solutions in terms of new business models and strategies, management principles, organizational models, processes and technological capabilities in order to satisfy their customers. In order to achieve those employees behavior is crucial factor. Soft dimensions which are basis to implementation of advanced digitalization and ubiquitous manufacturing until today are rarely surveyed, together with relation to culture and similar constructs, both input and output [2–5].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 94–103, 2020.
https://doi.org/10.1007/978-3-030-20040-4_9

This paper examines data collected in more than 100 Serbian industrial companies to check in which manner the upgrading of technological levels is connected to characteristics of employees' behavior.

1.1 Previous Research

The upcoming and ongoing industrial revolution 4.0 is triggered by the internet, which allows communication between humans as well as machines in Cyber-Physical-Systems throughout large networks, the Internet of things (IoT), and the Internet of services [6, 7]. Also, traditional single site manufacturing companies now are converted to decentralized multi-site manufacturing networks [8]. The growing demand of customized products in combination with decreasing product lifecycles forces further transformation towards new organization structures, which are capable to cope with increased complexity [6].

As the most flexible entity in cyber-physical production systems, employees are faced with a large variety of very demanding, different jobs that are completely different that are previously seen [7, 9]. Therefore, new additional qualification strategies for the current workforce are required, while employees need to become enabled to take on more strategic, coordinating and creative activities [10]. Barnum [11] points out that today the focus of contemporary IT technologies must be on the user, not on the product, that means that user must be in the center rather than technology – it forces workforce also to change. The fact is that in a contemporary manufacturing system, effective human communication is vital, not only for its operation, but also for its design and any further developments and changes [12].

The remote control of industrial systems today raises a growing interest and enables an easier and cheaper way for management of complex, dispersed or dangerous systems [13], while the success of those systems depends on the group characteristics and dynamics, the social and organizational skills and the positive and negative effects of technology on the group's tasks and processes characteristics [14]. Sometimes, the technology shift causes operators to focus only on their screens, reducing interaction between crewmembers and in those cases communication, vigilance and mutual awareness become extremely important [15].

New context and new skills further influence different interface design in aim to make better connection between technology and people. Today the interconnection between machine health analytics through a machine–cyber interface at the cyber level is possible, similarly to social networks, which means that when the cyber level infrastructure is in place, machines can register into the network and exchange information through cyber interfaces [16]. Also, few our previous research, such as Spasojevic Brkic et al. [2, 8] and [17], have described and analyzed different aspects of distributed remote user controlled manufacturing system that included "client" user interface for the distributed manufacturing system as collaborative environment, shown in Fig. 1.

Human – machine communication channel, representing the basic architectural pattern for remote control of complex distributed manufacturing systems, has been in the focus of those surveys. The user interface for the remote controlling functionality consisted of control panel for remote machine controls, communications controls, panel to see absolute and relative positions of each axis as the feedback information from the machine movements, and video frame to get live video feeds. Our previous research results in the field of testing group and individual working options in distributed remote user controlled manufacturing system show that group work consumes less time [2]. Even interface design issues point out the importance of technology and employees behavior harmonization.

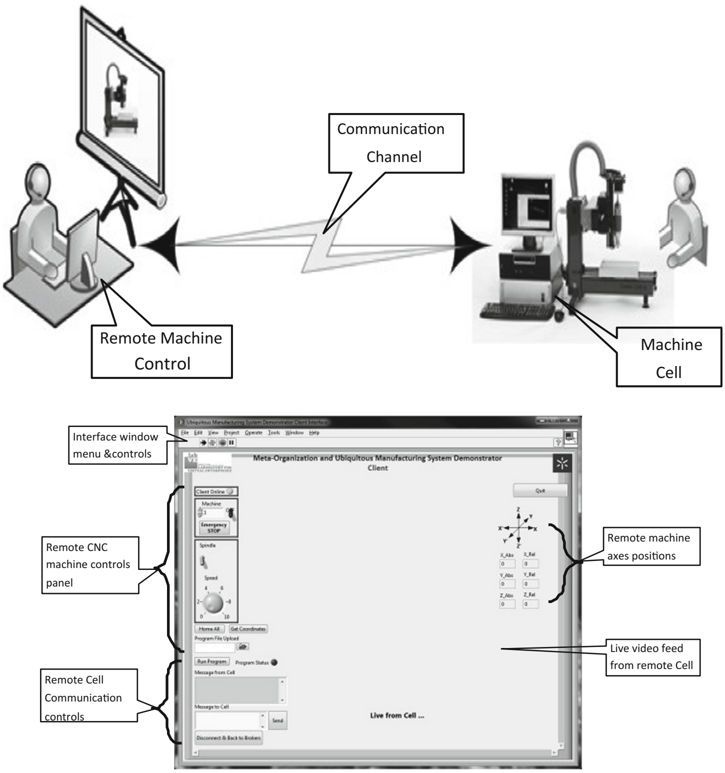


Fig. 1. An example of interface applied in I4.0 framework [4]

But, until now soft dimensions which are critical to implementation of ubiquitous manufacturing are very rarely surveyed [3]. Dubey et al. [3] discuss enablers of ubiquitous manufacturing and propose further research to examine relations between technologies and soft, human dimensions.

This paper is aimed to give those answers through problem introduction, identification of variables defining technology and employees' behavior described using previous research, followed by methodology and results on examined data which yields to adequate conclusions.

2 Technology and Employees Behavior Variables

Khandwalla [18] long time ago has noticed that technology is defined by level of automation, level of IT application, production program, diversity and product/service complexity. Decades later Escanciano et al. [19] point out that it is useful to take into account the technological level against competitors within a branch of industry, and thus separate capital-dependent from nondependent branches. Accordingly, in today environment actual technological dimensions that should describe technology as construct have to be defined and further analyzed (namely level of automation, level of IT application, and technological level against competitors).

Fok et al. [20] put employees' behavior parameters into the following groups:

1. proactive/reactive culture (orientation to quality, innovation, pro-active thinking),
2. soft/hard culture (describes employees' interpersonal relations: good, informal, fellowship atmosphere),
3. collectivism/individualism (collective spirit and good cooperation) and
4. power decentralization/centralization (power participation and decentralization).

Kuei and Madu [21] define the following dimensions:

1. The degree of autonomy of individuals;
2. The degree of correlation between earnings and quality of work;
3. The degree of risk taking in the performance of work;
4. The level of friendly team spirit;
5. The sense of belonging to the enterprise;
6. Employee participation rate and
7. Collective spirit of employees.

Combining dimensions of employees' behavior from previous research, the following dimensions appear and are will be analyzed in this paper:

1. Employees pro-activity and reactivity respectively;
2. Employees interpersonal relations;
3. Employees collectivism versus individualism and
4. Power participation and decentralization versus performing tasks by orders and centralized power.

3 Methodology and Results

To examine interaction between technology and employees behavior, variables data were collected, using a questionnaire formed for this purpose. Respondents were required to mark the value of technological and employee behavior dimensions rating at a Likert scale from 1 to 5. The survey population consists of 969 ISO 9001 certified industrial companies. Based on the experience of other researchers such as Casadesus et al. [22] and Simon et al. [23], in order to represent population in good manner, 250 industrial companies were chosen for the survey, including 40 large, 80 medium-sized and 130 small companies. Distribution of the survey was conducted via e-mail.

Responses were received from 119 industrial companies certified according to ISO 9001 management standard, which comprise more than 10% of the Serbian industrial company population with that characteristic. There was a 47.6% response rate. Average number of employees in companies in the sample was 370.91 with standard deviation 267.98. Answers have come from persons employed in Serbian industrial companies with an average of 16.98 years work experience. The 41% of employees that answered were quality department directors, 32% were directors of other technical departments, 22% employees work in quality departments or quality representatives and 5% of them come from other managerial positions.

Descriptive statistics on examined variables is shown in Table 1, while Table 2 shows partial correlations between examined variables, with significance level at $p \leq 0.05$.

Table 1. Descriptive statistics on examined variables

	Mean	Min	Max	*SD*
LA	2.97568	1.000000	5.00000	1.29096
IS	3.50455	1.000000	5.00000	1.14238
TL	3.18182	1.000000	5.00000	1.09362
PR	3.50000	1.000000	5.00000	1.15536
IR	2.40909	1.000000	5.00000	1.28002
CI	2.46364	1.000000	5.00000	1.21668
DE	3.46847	1.000000	5.00000	1.13474

Table 2. Partial correlations between examined variables (significant at $p \leq 0.05$ level)

	PR	*IR*	*CI*	*DE*
LA	0.11	0.5	0.5	0.44
IS	0.06	0.5	0.2	0.29
TL	0.24	0.5	0.4	0.45

Examined variables represent two groups of latent variables – technology and employees' behavior constructs. Hagenaars and McCutcheon [24] describe latent variables/constructs (opposed to observable variables) as variables that are not directly observed but are rather inferred (through a mathematical model) from other variables that are observed (directly measured). Factor and reliability analysis are appropriate methods for further examination of latent variables.

Factor analysis is used for a variety of applications as the default method of factors extraction and very efficient data reduction method [25]. The aim of factor analysis is to reveal any latent variables that cause the manifest variables to co-vary [25, 26]. Accordingly, factor analysis could be seen as the fundamental tool in multivariate statistic to summarize several (continuous) measurements through a small number of (continuous) latent traits very good method of checking dimensionality [27].

When variables developed from summated scales are used as predictor components in objective models, reliability of the scale becomes important. Since summated scales are an assembly of interrelated items designed to measure underlying constructs, it is necessary to check Cronbach's alpha coefficient [28, 29]. Cronbach's alpha is a measure of internal consistency or a measure of the scale reliability and it is calculated as a function of the number of test items and the average inter-correlation among the items [28, 29].

Table 3. Analysis of reliability for variable technology

Techology variable	Mean value after deletion	Variance after deletion	Standard deviation after deletion	α after deletion	$\overline{x}$ =9.97 *SD*=3.33 *Cronbach* α=0.634 *Stand.* α=0.640
LA	9.498991	5.924154	2.433959	0.426100	*Cronbach* α=0.634 *Stand.* α=0.640
IS	8.946697	7.959112	2.821190	0.618706	
TL	9.281560	6.953778	2.637002	0.477513	

Table 4. Exploratory factor analysis by the principal components method for technology variable

Variable technology	Factor 1	"Eigen" value 1.914273
LA	0.815296	
IS	0.563908	
TL	0.796535	
Explained variance	1.914273	
Percent in total	0.545073	

Communalities
Rotation: without rotation

	From factor 1	R^2
LA	0.664708	0.351132
IS	0.317993	0.115347
TL	0.634468	0.333989

The data in Table 3 show that there is adequate reliability of the scale. Factor analysis data (Table 4) show adequate values.

Table 5. Analysis of reliability for variable behavior of employees

Employee behavior variable	Mean value after deletion	Variance after deletion	Standard deviation after deletion	α after deletion	$\bar{x}$ =11.82 *SD*=2.81 *Cronbach* α=0.366 *Stand.* α=0.353
PR	8.327272	5.670165	2.381211	0.098882	After *DE* delectations *Cronbach* α=0.567 *Stand.* α=0.549
IR	9.418181	4.370578	2.090593	0.079110	
CI	9.363636	4.304132	2.074640	0.031970	
DE	*8.372727*	*7.001983*	*2.646126*	*0.332536*	

Table 6. Exploratory factor analysis by the principal components method for employees behavior variable

Employees variable	Factor 1	"Eigen" value
PR	0.739121	1.778144
IR	0.539898	
CI	0.613083	
Explained variance	1.778144	
Percent in total	0.592715	

Communalities
Rotation: without rotation

	From Factor 1	R^2
PR	0.781948	0.596770
IR	0.583408	0.297173
CI	0.617588	0.351490

Data in Table 5 show that the reliability of the scale can be significantly increased with 0.366 by ejection of the power decentralization dimension. In Table 6, it is noticeable that the remaining factors have an adequate loading value, Eigenvalues and communalities.

Since technology and employees variables are subjected as constructs, further correlation analysis could describe their interaction. There exists significant correlation that is described by correlation coefficient that has value 0.402, with p level equal 0.003.

4 Conclusion

This paper examines data collected in Serbian industrial companies to check in which manner the upgrading of technological levels in industry 4.0 context is connected to characteristics of employees' behavior. It is based on data collected in 119 Serbian industrial companies. Modern technology latent variable is described by technological dimensions such as level of automation, level of IT application, and technological level against competitors, while employees behavior concept is described by employees' proclivity/reactivity, interpersonal relations, collectivism/individualism and power participation and decentralization. Statistical examination included descriptive statistics, correlation analysis, factor and reliability analysis and results show that latent variables are appropriately described and that there is significant correlation between technology and employees behavior constructs.

Since Serbia is not industrially advanced country, in other industrially more advanced countries even stronger correlation could be expected.

It could be concluded that upgrading of technological levels forces employment of proactive people with soft culture (good, informal, community atmosphere and collective spirit) in industrial companies and the proposal for further research is to check results obtained herein in other industrial contexts and countries in the future.

Table 7. Nomenclature

Abv.	Meaning
ICT	Internet computer technology
IT	Informational technologies
IoT	Internet of Things
LA	Level of automation
IS	Information System
TL	Technological Level
PR	Proclivity
IR	Interpersonal relations
CI	Collectivity/individuality
DE	Decentralization
SD	Standard deviation
$\bar{x}$	Sample mean
p	Level of significance
R^2	Coefficient of determination
α	Cronbach alpha

Acknowledgements. This work is supported by grant TR35017, Serbian Ministry of Education, Science and Technological Development (MESTD).

References

1. Lasi, H., Fettke, P., Kemper, H.G., Feld, T., Hoffmann, M.: Industry 4.0. Bus. Inf. Syst. Eng. **6**(4), 239–242 (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. Dubey, R., Gunasekran, A., Chakrabarty, A.: Ubiquitous manufacturing: overview, framework and further research. Int. J. Comput. Integ. Manuf. **30**, 1–14 (2015)
4. Spasojevic-Brkic, V.K., Putnik, G.: User evaluation of interfaces for remote control manufacturing systems. Serb. J. Manag. **8**(2), 201–212 (2013)
5. 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)
6. Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M.: How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective. Int. J. Mech. Ind. Sci. Eng. **8**(1), 37–44 (2014)
7. Gorecky, D., Schmitt, M., Loskyll, M., Zühlke, D.: Human-machine-interaction in the industry 4.0 era. In: 2014 12th IEEE International Conference on Industrial Informatics (INDIN), pp. 289–294 (2014)
8. Brkic, V.S., Putnik, G., Veljkovic, Z.A., Shah, V., Essdai, A.: Group characteristics and task accuracy in distributed remote user controlled manufacturing as collaborative environment. In: International Conference on Applied Human Factors and Ergonomics, pp. 32–42. Springer, Cham (2017)
9. Roblek, V., Meško, M., Krapež, A.: A complex view of industry 4.0. Sage Open **6**(2), 2158244016653987 (2016)
10. Hecklau, F., Galeitzke, M., Flachs, S., Kohl, H.: Holistic approach for human resource management in Industry 4.0. Procedia CIRP **54**, 1–6 (2016)
11. Barnum, C.M.: Usability Testing Essentials: Ready, Set… Test!. Elsevier, Pennsylvania (2010)
12. Brkić-Spasojević, V., Putnik, G., Shah, V., Castro, H., Veljković, Z.: Human-computer interactions and user interfaces for remote control of manufacturing systems. FME Trans. **41** (3), 250–255 (2013)
13. Szafnicki, K., Noterman, D., Guillemot, M., Louail, G.: Remote control of manufacturing systems-a collaborative course between engineering schools. IFAC Proc. Volumes **33**(31), 79–84 (2000)
14. Antunes, P., Herskovic, V., Ochoa, S.F., Pino, J.A.: Structuring dimensions for collaborative systems evaluation. ACM Comput. Surv. (CSUR) **44**(2), 8 (2012)
15. Savchenko, K., Medema, H., Boring, R., Ulrich, T.: Comparison of mutual awareness in analog vs. digital control rooms. In: International Conference on Applied Human Factors and Ergonomics, pp. 192–199. Springer, Cham (2017)
16. Lee, J., Bagheri, B., Kao, H.A.: A cyber-physical systems architecture for industry 4.0 - based manufacturing systems. Manuf. Lett. **3**, 18–23 (2015)
17. Brkić, V.K.S., Putnik, G.D., Veljkovic, Z.A., Shah, V.: Interface for distributed remote user controlled manufacturing: manufacturing and education sectors led view. In: Handbook of Research on Human-Computer Interfaces, Developments, and Applications, pp. 363–391. IGI Global (2016)
18. Khandwalla, P.N.: Properties of competing organizations. Handb. Organ. Des. **1**, 409–432 (1981)

19. Escanciano, C., Fernández, E., Vázquez, C.: Linking the firm's technological status and ISO 9000 certification: results of an empirical research. Technovation **22**(8), 509–515 (2002)
20. Fok, L.Y., Fok, W.M., Hartman, S.J.: Exploring the relationship between total quality management and information systems development. Inf. Manag. **38**(6), 355–371 (2001)
21. Kuei, C.H., Madu, C.N., Lin, C., Lu, M.H.: An empirical investigation of the association between quality management practices and organizational climate. Int. J. Qual. Sci. **2**(2), 121–137 (1997)
22. Casadesus, M., Gimenez, G.: The benefits of the implementation of the ISO 9000 standard: empirical research in 288 Spanish companies. TQM Mag. **12**(6), 432–441 (2000)
23. Simon, A., Karapetrovic, S., Casadesús, M.: Difficulties and benefits of integrated management systems. Ind. Manag. Data Syst. **112**(5), 828–846 (2012)
24. Hagenaars, J.A., McCutcheon, A.L.: Applied Latent Class Analysis. Cambridge University Press, Cambridge (2002)
25. Costello, A.B., Osborne, J.W.: Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract. Assess. Res. Eval. **10**(7), 1–9 (2005)
26. Mulaik, S.A.: Foundations of Factor Analysis. Chapman and Hall/CRC, USA (2009)
27. Cai, L.: Latent variable modeling. Shanghai Arch. Psychiatry **24**(2), 118–120 (2012)
28. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika **16**(3), 297–334 (1951)
29. Santos, J.R.A.: Cronbach's alpha: a tool for assessing the reliability of scales. J. Extension **37**(2), 1–5 (1999)

Holistic Productivity Management Using Digitalization

Tim Jeske[1(✉)], Marc-André Weber[2], Frank Lennings[1], and Sascha Stowasser[1]

[1] ifaa – Institute of Applied Industrial Engineering and Ergonomics, Uerdinger Straße 56, 40474 Düsseldorf, Germany
{t.jeske, f.lennings, s.stowasser}@ifaa-mail.de
[2] Department of Management Studies, Institute for Supply Chain and Operations Management, Kiel University of Applied Sciences, Sokratesplatz 2, 24149 Kiel, Germany
marc-andre.weber@fh-kiel.de

Abstract. Holistic approaches for designing and developing enterprises and production systems are well-proven for years. They are still crucial when digitalization is introduced. Digitalization brings new opportunities regarding the collection, transfer, processing, providing and usage of data. An integrated application of these new opportunities for handling data can support holistic approaches and leads to what is defined as Industry 4.0, Industrial Internet of Things or Smart/Advanced Manufacturing. This contribution emphasizes on describing three basic elements of holistic approaches for managing productivity in production companies: Deriving goals for all management levels, designing work for reaching defined goals and continuous improvement for considering further potentials. The opportunities of digitalization for supporting each of these elements are elaborated in detail. Finally, a holistic implementation of digitalization by the help of a stepwise approach is described which can be supported by a detailed checklist.

Keywords: Holistic management · Productivity · Digitalization · Industry 4.0 · Industrial Internet of Things · Smart manufacturing

1 Opportunities of Digitalization for Holistic Approaches

Holistic approaches are characterized by considering more than several details of a system: They are based on considering ideally all parts of a system with all dependencies and interdependencies between them. This leads to an integrated view which makes holistic approaches successful. Consequently, the more information about a system, its parts, their characteristics and state are available, the more precisely any derivable conclusion or decision will be. This leads to a trade-off between the effort for collecting and updating more information about a system and the benefit generated regarding a more precise conclusion or decision base. This trade-off has to be solved according to the requirements of each application/use case.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 104–115, 2020.
https://doi.org/10.1007/978-3-030-20040-4_10

Since the character of digitalization is to make information easier accessible and to enable the handling of a huge amount of information within little time and with high precision, it becomes clear that digitalization can help gaining more and more precise information about whole systems. This decreases the effort for getting the same amount of information about a system while getting conclusions of the same precision. Conversely, the same effort for getting information about a system leads to more and more precise information and as a result to more precise conclusions. Consequently, there are large opportunities to be expected from introducing digitalization into a system. However, it has to be kept in mind that the complexity of today's systems may not be fully covered even by highly developed digitalization technologies. But still it seems worthwhile to deal with the chances given in regard of the opportunities aforementioned.

2 Productivity and Its Management

Productivity describes the ratio between the output and the input of a business process [1]. It is applicable to any kind of business process: direct as well as indirect processes and for all processes along the value chain. Independently from the kind of process, its influencing factors can be clustered into three groups: factors influencing the output, factors influencing the input and factors influencing both [2]. For production companies these three factor groups are illustrated in Fig. 1 including examples. Thus, productivity depends on many parameters which can be influenced specifically. For increasing productivity, parameters influencing the output have to be adjusted to increase the output, parameters influencing the input have to be adjusted to decrease the input and those which are capable for influencing both have to be adjusted depending on the individual circumstances of each company and each application/use case. Due to this, there are many opportunities for choosing parameters for influencing productivity. In addition to that, each parameter generally can be adjusted in many ways and by different activities.

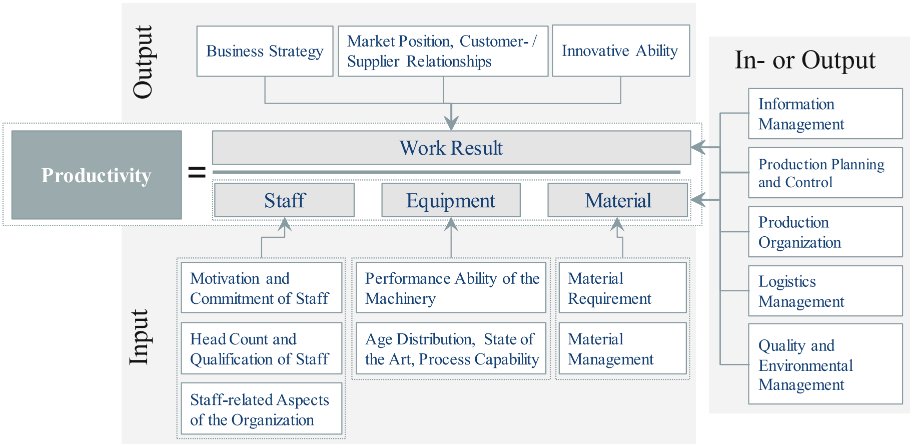


Fig. 1. Examples of influencing factors on productivity [2].

All activities aiming on adjusting parameters for influencing productivity – including systematical planning, steering, realizing and controlling – are subject of productivity management [3]. Productivity management is basically designed as a control loop (see Fig. 2). It starts at the systematical *productivity planning,* which is subdivided into strategical and operative planning. Strategical productivity planning focuses on decisions affecting a whole company and determines long-term productivity goals. Operative productivity planning concretizes the determined goals of the strategical planning by deriving target values for specific parameters on a short- and mid-term basis, which will be used for *productivity steering* activities as the second step of the productivity management loop. Afterwards, activities are executed for *realizing productivity measures*, respectively for reaching the target values within the affected departments of a company. Finally, *productivity controlling* compares targeted and reached values of the relevant parameters for handing over the result to the productivity planning and thus closing the control loop. This enables the productivity planning to recognize potentials for further improving measures, to implement corrective measures and to adjust goals.

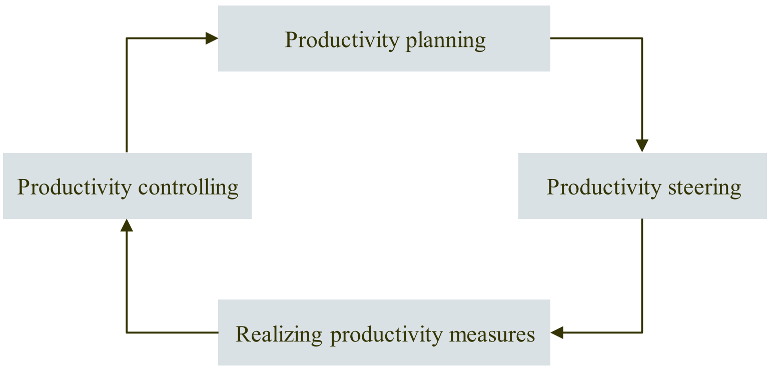


Fig. 2. Control loop of productivity management [3].

3 Holistic Productivity Management

3.1 Defining Goals

Since productivity management is designed as a control loop (see Fig. 2), it is essential to define goals to which the management activities aim on. As described before, these goals have to be concretized by target values for parameters (see Fig. 1) influencing productivity.

Defining goals which meet the claim to fit into holistic approaches requires to start on a normative management level. This is where companies define major decisions like their vision, purpose and mission [4]. Based on this, strategic goals have to be derived for the strategic management level. They are long-ranging and focus on the long-term development of a company. Thus, strategic goals are the basis for strategic productivity planning and allow deriving goals for the operative management level. These goals are concretized target values for selected productivity-influencing factors and focus on the short- and medium-term development of a company. Thus, on the operative management level the operative productivity planning can be executed to prepare steering and realizing productivity goals. In this way, goals for the operative productivity management are derived and concretized step by step down from the normative to the operative management level (see Fig. 3).

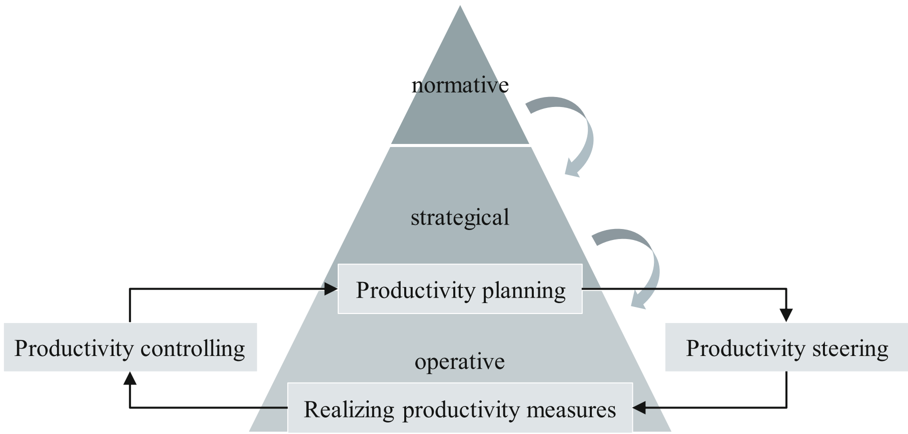


Fig. 3. Deriving goals from the normative to the operative management level and placement of the productivity management loop within the management levels.

On the operative level it has to be taken into account if the influencing of a parameter aims on changing the process' input or the process' output (see Fig. 1). Additionally, it has to be considered if this shall happen by a quantitative or a qualitative improvement of productivity. A quantitative improvement is given if the amount of input factors decreases or the amount of output produced increases. A qualitative improvement is given if the amounts are unchanged, but the output achieves a higher level of quality or the input needs a lower level of quality [5]. This is important since the selection of adequate measures for improving productivity mostly depends on quantitative as well as qualitative influences. Due to this, goals for increasing productivity can be differentiated generally according to four types:

- Qualitative Output (up/increase)
- Quantitative Output (up/increase)
- Qualitative Input (down/decrease)
- Quantitative Input (down/decrease)

3.2 Designing Work

Reaching defined goals and the derived target values for certain parameters requires adequate designing of all parts of a company's value creation process. Nevertheless, this contribution focuses on manually executed parts within the value creation process: human work. Due to this, the automation of processes (e.g. for increasing productivity or compensating a lack of workforce) is not considered in this paper.

Human work is usually designed by industrial engineers and ergonomics experts. To do so, basically three strongly interdependent design aspects are considered: technical aspects, organizational aspects and human aspects. A common approach for designing work and workplaces – derived from safety engineering – is to start adapting technology to the needs of human workers. Afterwards, organizational aspects are adapted in the same way to the human worker's needs. Finally, human workers are equipped and taught how to execute their work at the new or new designed workplace.

The way human workers perform their work tasks strongly impacts their productivity. In this context, the chosen work method usually impacts productivity stronger than the level of performance it is carried out with [6]. Thus, the most effective and efficient method has to be elaborated and defined by means of a standard. These standardized work processes ensure a high productivity and can work as a guideline for employees – especially for those, who are new to the described work task.

This procedure can be applied to any type of work. It is applicable for energetic work, which requires human workers to apply their muscles and physical strength, as well as for informatoric work, which requires human workers to perceive information, mentally process them and link them to conclusions (which are new information). Since any kind of work is composed of the afore described types of work [7] (see Fig. 4), also any kind of work can be designed within technical, organizational and human aspects.

Type of work	**Energetic work**				**Informatoric work**
Kind of work	mechanical	motor	reactive	combinative	creative
What requires the completion of the task to do?	Use force	Execute movements	Reacting and acting	Combining information	Creating information
	"Mechanical work" according to physics	Accurate movement with low force	Perceive information and react to it	Link information to memory content	Link information to "new" information
Which organs or functions are stressed?	Muscles, tendons, skeleton, breathing	Sensory organs, muscles, tendons, circulatory system	Sensory organs, reaction and memory ability, muscles	Thinking and memory ability, muscles	Thinking, memory and concluding ability
Example	carrying	assembly	car driving	engineering	inventing

Fig. 4. Kinds of work and their composition out of energetic and informatoric types of work [7].

3.3 Continuous Improvement

When goals are defined and work is designed (or redesigned) for reaching those goals, there is usually still potential leftover. This is due to several reasons. The main reasons are: (1) The method firstly elaborated and standardized might be suboptimal. A better method is often found based on the experience gained during the application of the standard, respectively the standardized work process. (2) The circumstances of the work system might have changed. This can affect all parts of a work system from input specifications over work equipment to environmental effects and required output. (3) New opportunities might have been developed which naturally could not have been considered while the process was standardized for the first time. This can affect materials, equipment, technologies etc. Digitalization is one of these developments and has increasingly created new opportunities in the last few years.

This potential should be considered by the help of continuous improvement processes. They aim on the same goals derived from the normative level and support approaching them stepwise. This requires (1) a broad knowledge of the normative goals within a company and its departments, (2) an implementation of continuous improvement processes in all these departments and (3) the continuous adaptation of already defined standards. Furthermore, continuous improvement should be anchored within the company's corporate culture.

4 Opportunities of Digitalization for Holistic Productivity Management

4.1 Digital Supported Data Handling

Digitalization affects fundamentally the way data and information are handled. It requires all kind of information that has to be available either to be digitized or to be *collected* already digitally. This is the first of five steps for data handling and allows the non-physical data to be *transfered* via digital networks and thus to be available very quickly – in principle – everywhere within a company or around the world (so-called real-time data). This high availability of data opens up huge opportunities for *processing* them in order to support all kind of activities and decisions. Additionally, information can be *provided* via different kinds of displays according to current needs – for example, in assembly areas according to the current progress within the work process. Finally, data can be *used* e.g. for implementing safe human-robot-collaboration. These five steps (collection, transfer, processing, providing and usage of data) are illustrated and complemented by examples in Fig. 5.

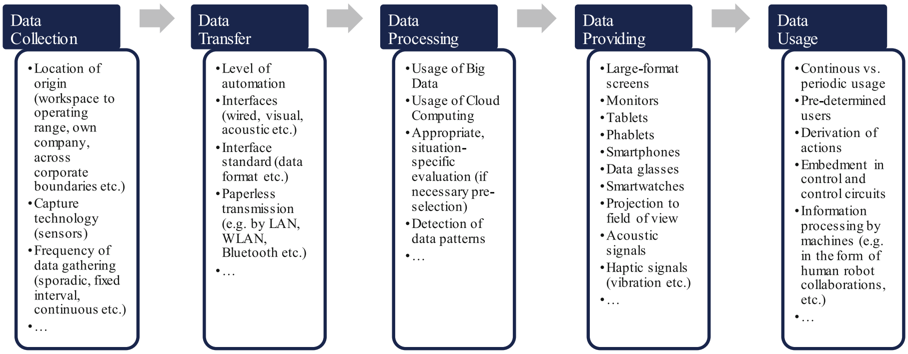


Fig. 5. Steps of digitalized data handling with examples for each [7].

Since digitalization automates large parts of the handling of data and thus accelerates the involved processes, it is important to ensure the effectiveness of each process that shall be supported digitally. This requires processes which are designed according to the principles of lean management, e.g. the 5S methodology, the elimination of waste (muda) and the elimination of information barriers. It is strongly recommended to gain any potential for improvement within a process prior to using digitalization for its further improvement.

4.2 Data Supported Defining of Goals

Digitalization and the consequently improved availability of data can support the definition of goals in two major ways: When goals are derived from the normative to the operative management level and when already existing goals need to be adjusted.

The derivation of goals from higher levels is basically supported by more and more precise data about the actual state of the company (e.g. highly aggregated real-time key figures) and about development tendencies as well as expected changes in its environment (e.g. market demands, availability of workforce etc.).

Regarding already existing goals which are operationalized by target values digitalization supports the adjustment within the control loop on the operative management level. Therefore, productivity controlling (see Fig. 2) takes advantage of the improved availability und quality of data for determining the currently reached state of the parameters for which target values were defined. This leads to a more precise detection of the difference between reached and targeted values. Additionally, the improved data availability can support the investigation of reasons why target values could not be reached (regardless of whether they were over- or under-fulfilled). Both information – the current value as well as the reason for it – support the productivity planning step within the control loop (see Fig. 2).

4.3 Data Supported Designing of Work

Digitalization and the consequently improved availability of data can support the design of work by enabling more individualization – regarding the variety of products, processes and human workers.

The variety of products is supported by an improved management of product specifications and its presentation/illustration according to each work task. Especially work instructions and bills of materials which are always up to date, easily available and presented according to the current work progress facilitate human work and thus increase productivity.

The variety of processes is supported by an enhanced production planning which analyzes different possible production sequences for each customer order and selects the most suitable sequence for each order according to current needs and capacities. This can support even work load at different work places (when work is executed at all workplaces) as well as uneven workload at different workplaces (when work is executed only at some workplaces – maybe due to workplace-specific shift planning or due to a lack of workforce).

The variety of human workers is supported by more individualized design of workplaces (e.g. height of a worktable or individual preferences in workplace illumination) or work instructions (e.g. adapted to individual experiences). Additionally, job rotation can get more ergonomic by being set up based on ergonomics assessment methods and individual characteristics of each worker [8].

4.4 Data Supported Continuous Improvement

Digitalization and the consequently improved availability of data can support continuous improvements of a process. Thus, more data and key figures can be collected and tracked to get evidence for improvement measures. Additionally, these data can get analyzed by the help of correlations and other statistical methods and approaches that can extract "hidden" information out of large amounts of data (e.g. big data analysis). This helps to find the right starting points for improvement measures and allows quick feedback on conducted improvements to adjust these measures if necessary.

Therefore, digitalization supports continuous improvement measures by (1) deepening knowledge and experiences about work processes, (2) adapting these experiences quickly when circumstances of a work system change and (3) targeted implementing of new opportunities into an existing system.

5 Holistic Implementation of Digitalization

Process models for the implementation of digitalization can be found variously in the literature [e.g. 9–12] and focus on different approaches. Since companies show an individual degree of digitalization, which results from the already existing use of different information technologies, it is suggested to evaluate the current infrastructure of the company in scope before conducting further activities. The implementation-approach presented here explicitly takes the perspective of productivity into consideration, integrating the elements of holistic productivity management presented before.

As discussed earlier, processes based on set standards are a prerequisite for a functioning lean management and successful digitalization. In this context, they undergo a steady improvement. Based on continuous improvement and its anchoring in the mindset of the staff, a holistic process orientation can be achieved by using digitalization. The achievable additional added value from the application of digitalization, which goes beyond traditional lean management, seems to be of central importance from the point of view of business process management. However, one should keep in mind that many companies still have potential to increase productivity through “traditional” methods of lean management that should be fully employed before the use of digitalization.

For a structured implementation of digitalization in a company, a procedural model is presented in Fig. 6. It is designed in several stages and depicts the entire process from determination of suitable company-specific approaches of digitalization to a holistic implementation of these approaches.

The starting point (1a) is the idea that the use of digitalization affects both business models and business processes. Thus, as a higher-level measure, companies first have to review their existing business models and the associated corporate strategy and, if necessary, adjust them before adjustments can be made in subordinate processes. It is therefore necessary to examine whether products and operational processes can or should be optimized or redesigned by means of digitalization. Approaches to do so need to be defined in the next step (1b).

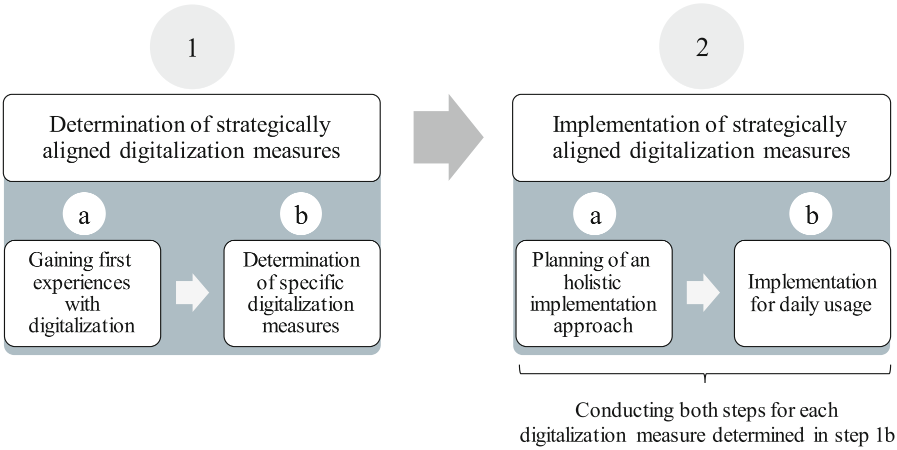


Fig. 6. Holistic procedural model for implementing digitalization

The latter can be supported by using a framework introduced by Weber et al. [5] which has been filled with many examples [13–15]. This determination of specific digitalization measures should represent a selection of the most concrete technologies that can be introduced in practice and used in operational processes. In this regard it is important that potential digitalization measures promise advantages for business processes. Especially in this step, the mentioned framework can be used to find suitable technologies. The selection of technologies and their integration into a company should ideally be integrated into a holistic concept, e.g. the use of digitalization must support strategic goals of the company in scope. Thus, it has to be considered both for business models and for the processes necessary for their implementation.

Afterwards, the selected strategical options for the use of digitalization, in particular for processes, have to be successively broken down to the operational level and concretized in terms of a structured implementation (2a). This needs to be done for each digitalization technology selected in (1b). The procedure in the model is based on the assumption that a holistic approach to digitalization must apply both to individual workplaces and to an entire production system and thus goes beyond the mere consideration of technical aspects. This implies that digitalization is changing work content, processes and environments, and that, in addition to technology, business, human factors and ergonomics implications must be considered.

Due to these interdisciplinary requirements, questions about the implementation of digitalization need to be discussed and answered (2b), ideally between members of a cross-departmental team. It is recommended that a group of representatives of different divisions is assigned to take responsibility for the success of implementing a specific part of digitalization in the context regarded. It may also be advisable to use external specialists if know-how of a company should be supplemented by non-company competences and experiences.

For supporting all these steps, a detailed checklist has been developed [16, 17]. Steps 1a and 1b are covered by orientation questions regarding circumstances and tendencies of the company's customer markets as well as the company's current process organization and state of digitalization. Steps 2a and 2b are covered by questions regarding eleven fields of action. These are: Work design, work organization, health and safety at work, qualification and qualification measures, operation and working hours, wage, privacy, data security, employee participation, external support as well as profitability and success.

6 Summary

Even though holistic approaches are well-proven for years, they still can take advantage of digitalization. This is due to the fact that digitalization supports the main character of holistic approaches: considering ideally all parts of a system and their state.

This applies also for the digitalization's support for holistic productivity management. The main elements of holistic productivity management are supported by enhanced and facilitated handling of data – from collection and transfer to processing, providing and usage: (1) deriving goals for all management levels – regarding the derivation of goals from the normative to the operative management level as well as the adjustment of existing goals on each level, (2) designing work for reaching defined goals by supporting the handling of variety in products, processes and human workers, and (3) continuous improvement for considering further potentials by deepening knowledge, quickly adapting knowledge to changes and targeted implementation of new opportunities.

This underlines the importance of a successful implementation of digitalization for companies and their further development. For supporting this, a holistic implementation approach for digitalization is presented. It consists of four steps from gaining first experiences with digitalization to implementation of digitalization measures. All steps are supported by a detailed checklist which covers three fields of orientation as well as eleven fields of action.

Acknowledgements. The presented results were elaborated in the research project “TransWork – Transformation of Labor by Digitalization”, which is funded by the German Federal Ministry of Education and Research (BMBF) within the “Innovations for Tomorrow’s Production, Services, and Work” Program (funding number 02L15A164) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

References

1. Wöhe, G.: Einführung in die Allgemeine Betriebswirtschaftslehre, 21st edn. Vahlen, München (2002)
2. Nebl, T.: Produktivitätsmanagement: theoretische Grundlagen, methodische Instrumentarien, Analyseergebnisse und Praxiserfahrungen zur Produktivitätssteigerung in produzierenden Unternehmen. Carl Hanser, München (2002)
3. Dorner, M.: Das Produktivitätsmanagement des Industrial Engineering unter besonderer Berücksichtigung der Arbeitsproduktivität und der indirekten Bereiche, Dissertation an der Fakultät für Maschinenbau, Karlsruher Institut für Technologie (KIT), Pforzheim (2014)
4. Bergmann, R., Bungert, M.: Strategische Unternehmensführung. Springer, Berlin, Heidelberg (2012)
5. Weber, M.A., Jeske, T., Lennings, F., Stowasser, S.: Framework for the systematical design of productivity strategies. In: Trzcielinski, S. (ed.) Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future, pp. 141–152. Springer, Berlin (2017)
6. Bokranz, R., Landau, K.: Produktivitätsmanagement von Arbeitssystemen. Schäffer-Poeschel, Stuttgart (2006)
7. Schlick, C., Bruder, R., Luczak, H.: Arbeitswissenschaft. Springer, Heidelberg (2010)
8. Jeske, T., Brandl, C., Meyer, F., Schlick, C.M.: Personaleinsatzplanung unter Berücksichtigung von Personenmerkmalen. In: Gesellschaft für Arbeitswissenschaft e.V. (ed.) Gestaltung der Arbeitswelt der Zukunft – 60. Kongress der Gesellschaft für Arbeitswissenschaft. GfA-Press, Dortmund, pp. 327–329 (2014)
9. Appelfeller, W., Feldmann, C.: Die digitale Transformation des Unternehmens. Systematischer Leitfaden mit zehn Elementen zur Strukturierung und Reifegradmessung. Springer Gabler, Berlin (2018)
10. Fleischmann, A., Oppl, S., Schmidt, W., Stary, C.: Ganzheitliche Digitalisierung von Prozessen. Perspektivenwechsel, Design Thinking, Wertegeleitete Interaktion. Springer Vieweg, Wiesbaden (2018)
11. Schallmo, D.R.A.: Jetzt digital transformieren. So gelingt die erfolgreiche Digitale Transformation Ihres Geschäftsmodells. Springer Gabler, Wiesbaden (2016)

12. Weinreich, U.: Lean Digitization. Digitale Transformation durch agiles Management. Springer Gabler, Berlin, Heidelberg (2016)
13. Weber, M.A., Jeske, T., Lennings, F.: Ansätze zur Gestaltung von Produktivitätsstrategien in vernetzten Arbeitssystemen. In: Gesellschaft für Arbeitswissenschaft e.V. (ed.) Soziotechnische Gestaltung des digitalen Wandels – kreativ, innovativ, sinnhaft – 63. Kongress der Gesellschaft für Arbeitswissenschaft. GfA-Press, Dortmund (2017)
14. Jeske, T., Weber, M.A., Klues, J., Lennings, F.: Strukturierung und Analyse von Praxisbeispielen zur Nutzung der Digitalisierung für das Produktivitätsmanagement. Z. Arb. Wiss. 72, pp. 190–199 (2018)
15. Lennings, F., Jeske, T., Meyer, K.: Konzept zur Auswahl und ganzheitlichen Planung von Digitalisierungsmaßnahmen. In: Gesellschaft für Arbeitswissenschaft e.V. (Hrsg) Arbeit interdisziplinär analysieren - bewerten - gestalten – 65. Kongress der Gesellschaft für Arbeitswissenschaft. GfA-Press, Dortmund (2019)
16. Weber, M.A., Terstegen, S., Lennings, F.: ifaa-Checkliste: "Digitalisierung & Industrie 4.0 in der Praxis". Betriebspraxis & Arbeitsforschung (232), p. 57 (2018)
17. Weber, M.A., Terstegen, S., Lennings, F., Institut für angewandte Arbeitswissenschaft (ed.): Checkliste Digitalisierung & Industrie 4.0 in der Praxis. Geschäftsstrategie und Prozesse ganzheitlich gestalten. ifaa, Düsseldorf (2017). www.arbeitswissenschaft.net/checkliste-digitalisierung

Enabling Smart Workplaces by Implementing an Adaptive Software Framework

Peter Burggräf[1,2(✉)], Achim Kampker[3], Kai Kreisköther[3], Tobias Adlon[1], Aaron Riegauf[3], Benjamin Dorn[3], Patricia Casla[4], Angelo Marguglio[5], and Giuseppe Donvito[6]

[1] Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany
{P.Burggraef,T.Adlon}@wzl.rwth-aachen.de
[2] Chair of International Production Engineering and Management of University of Siegen, Paul-Bonatz-Straße 9 - 11, 57076 Siegen, Germany
Peter.Burggraef@Uni-Siegen.de
[3] Chair of Production Engineering of E-Mobility Components (PEM) of RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany
{A.Kampker,K.Kreiskoether,A.Riegauf, B.Dorn}@pem.rwth-aachen.de
[4] Fundación Tekniker, Calle Iñaki Goenaga 5, 20600 Eibar, Gipuzkoa, Spain
Patricia.Casla@tekniker.es
[5] Engineering Ingegneria Informatica S.p.A, Piazzale dell'Agricoltura 24, 00144 Rome, Italy
Angelo.Marguglio@eng.it
[6] ILLOGIC S.r.l., Corso Stati Uniti 57, 10144 Turin, Italy
Donvito@illogic.xyz

Abstract. Today's requirements and challenges on production systems have led to an increasing importance of flexible assembly processes. At the same time, arising technologies in the context of Industry 4.0 offer the opportunity to tap previously unknown potentials. These provide both opportunities and risks for the most flexible and important resource in the production: the workers. Within the EU research project A4BLUE, Smart Workplaces are developed in order to raise productivity and quality, while sustaining and enhancing worker satisfaction. Enabled through both a software framework for adaptation management and contextual worker assistance as well as hardware solutions (e.g. automation mechanisms such as smart tools or co-operative robots) for physical support, the workplaces adapt to the workers' specific characteristics considering both process and environmental variability. The presented research work provides an overview about the Frameworks' software components and the application to a use case at the RWTH Aachen University.

Keywords: Smart Manufacturing · Smart Workplaces · Smart assembly · Worker assistance · Adaptive Automation · Software reference architecture

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 116–127, 2020.
https://doi.org/10.1007/978-3-030-20040-4_11

1 Introduction

In course of advancing and gradually establishing Industry 4.0 technologies, nearly every industrial sector is affected or being transformed somehow. Those technologies such as Virtual Reality (VR), Augmented Reality (AR) or Collaborative Robots, can be both enablers and catalysts for the transformation towards Smart Manufacturing. Smart Manufacturing can provide a company's crucial competitive edge by shortening times for production, implementation and information processing, thus allowing faster reactions to continuously changing markets [1].

Nevertheless, the conversion to Smart Manufacturing leads to a certain amount of investment costs. In order to optimize efficiency in the manufacturing ecosystem, it is crucial that all resources are used to their full potential. This is particularly significant on the shop floor level, where humans directly interact with automated systems. Consequently, Smart Workplaces – the shop floor level subdivision of Smart Manufacturing – need to ensure the accurate application of the individual strengths of both humans and automation mechanisms.

Collaboration of humans and robots for example allows the usage of the precision, the ability of repeatability as well as the physical strength of robots combined with the humans' flexibility and capability to react to unpredictable changes. Therefore, Smart Workplaces do not primarily aim either to replace the shop floor workers (blue collar workers) with automated systems or to monitor and control the workers. Instead, the purpose of Smart Workplaces is to enable a more efficient use of the workers' strengths by assisting them with the targeted use of hardware support and real-time data processing. This will enhance the process of physical task execution with an increasing part of coordinating tasks and distributed decision making [2].

To enable Smart Workplaces, real time collection, analysis and distribution of digital information between and within the manufacturing ecosystem are required [1]. With a digital image of the production processes, adaptation measures can be executed automatically, information can be displayed to the workers and the current process efficiency can be evaluated. To manage the large amount of data generated on the shop floor, the use of key performance indicators (KPIs) could be an appropriate approach [3]. KPIs, such as process time or rework rate, are used by automated systems and can be either directly displayed to the worker or indirectly as recommendations or notifications after further analysis.

In any case, it has to be ensured that these KPIs are available in an appropriate level of detail depending on the respective addressee. This means, the workers need to be provided with the required information at the right time. Furthermore, the visualization has to be configured individually with regard to the workers' preferences and experience [4]. By doing so, experienced workers are provided with different types and information details than less experienced ones.

In a Smart Workplace, the workers are given the opportunity to check work performance by receiving individual feedback on tasks and potential errors. Having the feeling of being an irreplaceable resource in the manufacturing ecosystem and provided with the possibility of a high self-reliance degree, the workers are motivated to achieve a continuous improvement of themselves as well as of the processes [5]. It is assumed, that having committed shop floor workers, improvement in business KPIs, such as productivity or quality, will take place.

Physical world and virtual world need to be connected to enable a Smart Workplace. Therefore, a digital infrastructure for the processing and provision of data is needed, which is aligned to worker support. For those reasons, dedicated information and communication technology tools and methodologies have been developed within the EU research project A4BLUE (Adaptive Automation in Assembly For BLUE collar workers satisfaction in Evolvable context), realizing the so called A4BLUE Adaptive Framework.

2 Software Framework for Adaptive Assembly

The A4BLUE Adaptive Framework provides an open, secure, configurable, scalable and interoperable adaptation management and assistance toolkit. This allows to adjust workplace characteristics according to context information, such as dynamic production processes, environment changes and varying workers' characteristics. The Framework's concept includes elements of three existing Industry 4.0 reference architectures: Reference Architectural Model Industrie 4.0 (RAMI 4.0) [6], Industrial Internet Reference Architecture (IIRA) [7] and FIWARE Smart Industry Reference Architecture [8].

The main principles of the Framework's design are integration, virtualization, adaptation management, context aware worker assistance and monitoring. Three high-level functional domains (Shop Floor, Enterprise and Business) group those principles into a logical three-fold breakdown structure containing different components as outlined in Fig. 1.

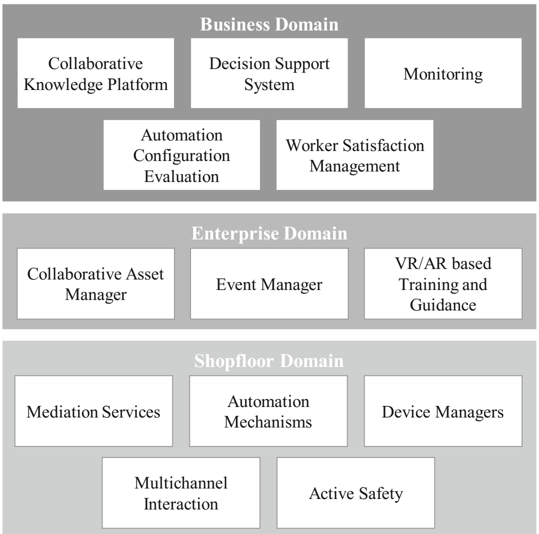


Fig. 1. Domains and components of the A4BLUE Adaptive Framework

The domains define a coarse mapping of system elements either to the factory – Shop Floor Domain - or to the broader field of corporate IT - Enterprise or Business Domain. This structure empowers the direct connection of cyber-physical systems with humans in the different areas of the production ecosystem.

The Shop Floor Domain is the bottom layer of the Framework and populated by any kind of device that is connected to the physical world on one side and to the virtual world on the other. Therefore, this domain includes functionalities supporting automated control and adaptation of physical production processes as well as the main human-machine interactions. The domain consists of the software components shortly described in Table 1.

Table 1. Components of the Shop Floor Domain

Shop Floor Domain	
Mediation Services	The Mediation Services enable the integration of existing enterprise level legacy systems (e.g. Manufacturing Execution Systems) into the Framework
Automation Mechanisms	The Automation Mechanisms component includes physical automation technology such Collaborative Robots or Smart Tools
Device Managers	The Device Managers support a plug and produce approach by enabling discovery and integration of external automation mechanisms using existing and common standards (i.e. OPC UA)
Multichannel Interaction	The Multichannel Interaction component enables the workers to interact with the automation mechanisms in an intuitive way through verbal (e.g. voice commands or natural speaking) and non-verbal (e.g. gestures) channels, as well as to receive appropriate feedback through visual (e.g. graphical user interfaces, lights) or auditory (e.g. voice messages, sounds) channels
Active Safety	The Active Safety component adapts the behavior of automation mechanisms considering the workers activity

The Enterprise Domain is the middle layer and represents the core part of the A4BLUE Adaptive Framework. To this end, this domain consists of several components in charge of managing the logic for adaptation management and the provision of assistance services using an Event Driven Architecture. The domain consists of the software components briefly described in Table 2.

Table 2. Components of the Enterprise Domain

Enterprise Domain	
Collaborative Asset Manager	The Collaborative Asset Manager provides virtualization capabilities by managing the semantic virtual asset model as a knowledge base of the ecosystem. It includes reasoning capabilities and contains static representations of both tangible assets (physical assets) and intangible assets (non-physical assets) as well as dynamic context representations
Event Manager	The Event Manager continuously captures and analyzes relevant events from the shop floor as well as the legacy system. It reacts to an event, by triggering the required adaptation or assistance actions or by enabling the information flow among the different components of the Framework
VR/AR Based Training and Guidance	The Virtual and Augmented Reality component provides off- and on- the job training and guidance for the operator. Furthermore, VR/AR interaction devices allow capturing events and data, which are used for real-time context aware reasoning in collaboration with other components of the Framework

The A4BLUE Business Domain is the upper layer, in charge of supporting the strategic decision-making process by targeting both, blue and white-collar workers. All components in this domain aim to extract and take benefits from the information basis of the production's running automation and adaptation processes. The domain consists of the software components shortly described in Table 3.

Table 3. Components of the Business Domain

Business Domain	
Collaborative Knowledge Platform	The Collaborative Knowledge Platform transfers knowledge, especially implicit and informal knowledge, from skilled to less experienced workers as well as to various departments of an organization
Decision Support System	The Decision Support System supports operators on relevant decisions (e.g. for assembly, maintenance or inspection operations) by providing insights on processed data
Monitoring	The Monitoring component enables the collection and calculation of KPIs for the assessment of production processes and performances
Automation Configuration Evaluation	The Automation Configuration Evaluation component allows the assessment of different levels of automation from a socio-technical as well as from an economical perspective to identify the most efficient Level of Automation
Worker Satisfaction Measurement	The Worker Satisfaction Measurement component assesses the workers satisfaction quantitatively in relation to human-automation systems and work environment characteristics

As outlined in the component's descriptions, every component is designed to serve a specific purpose, which differs from each other significantly. Some components enable the connection of the A4BLUE Adaptive Framework to the existing software and hardware infrastructure as well as the seamless connection of new equipment. Others are designed to manage processes and derive knowledge from information exchanged within the Framework. Finally, the actual analysis of data and information is performed to support several decision making processes.

Therefore, some of the components are operating in the background most of the time, invisible for the end user, or are seldom used for general configuration activities such as the integration of new automation mechanisms. Others have a perpetual interface to the end user for the continuous processing of input and output data in order to enable the direct worker support at the workplace.

3 Adaptive Framework Application to an Assembly Use Case

In order to test and evaluate the hardware and the software developments, two industrial as well as two laboratory use cases were set up within the A4BLUE project. The use cases' purpose is to validate the technical feasibility and functionality as well as to obtain objective feedback on the socio economic impact of the A4BLUE solutions.

One of these use cases is set in the Ramp-Up Factory of RWTH Aachen University, which has been developed in cooperation with shop floor experts of the factory. Within this use case the assembly processes of a rear light and a brake module of an electric vehicle are performed in a representative working environment. Both processes are conducted by a shop floor worker, while organizational, quality and steering operations are undertaken by a production supervisor. For the use case implementation a set of components of the A4BLUE Adaptive Framework has been selected with regard to process and production side framework conditions. Within this chapter only an extract of the overall use case is described. It provides the current status of the concept, while the technical implementation is still in progress. The entire implementation includes additional components and hardware solutions such as the Automation Configuration Evaluation or an autonomous tool trolley.

3.1 Main Components of the Adaptive Framework Applied to the Use Case

To empower the Smart Workplace within the use case, data needs to be processed according to the three steps of the data value chain: data collection, data interpretation and visualization [9].

Subsequently only those components, that interact directly or indirectly in regular operation with the shop floor worker as well as with the production supervisor are described in detail. Nevertheless, further components of the Framework such as the Event Manager or the Collaborative Asset Manager are mandatory to assure the workplaces adaptability capabilities. Within the use case three components are mainly necessary for the interaction among workplace and workers. These components are VR/AR based Training and Guidance, Monitoring and Decision Support System (DSS).

The first step (data collection) is realized with the VR/AR based Training and Guidance component. To connect this component to the physical world, the worker wears Augmented Reality glasses (Microsoft® HoloLens™). Via the device, it is possible to collect data and information from the assembly process by triggering certain actions that are proceeded as events. A back-end server application acts as a bridge between the device and the A4BLUE Adaptive Framework, where the Event Manager and the Collaborative Asset Manager provide the capabilities to exchange and process data. With this link, data (e.g. 3D-models, assembly instruction or statistics) are made available dynamically and contextually to the worker. Additionally, actions triggered by the worker (e.g. gestures or voice commands) or by external impulses (e.g. markers or a task completion event from a given work order) are captured and sent back to the Framework for further elaboration. The front-end modules are outlined in Fig. 2.

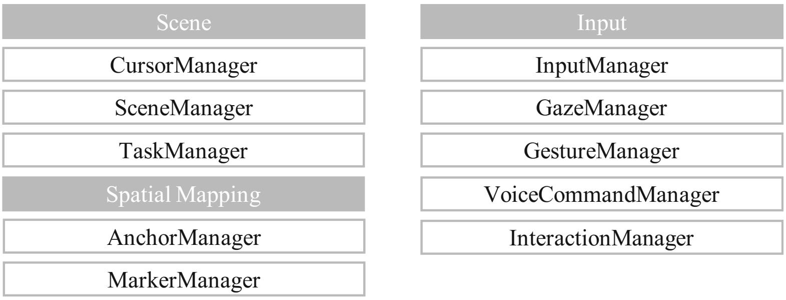


Fig. 2. Categories and modules of the front-end application

The modules are classified according to the categories Scene, Input and Spatial Mapping. Modules within Scene enable the static and dynamic visualizations of all necessary objects to create the virtual elements. The category Input contains modules for the capture, the interpretation and the management of actions via different channels. The modules of Spatial Mapping ensure correct position and size of virtual objects as well as the processing of external impulses with markers. Table 4 provides an overview of the modules.

The Monitoring and the DSS component are implemented within the second step of the data value chain (data interpretation), on basis of the Open Source Knowage Suite[1]. The purpose of the Monitoring component is to make processes and information from the production ecosystem understandable and manageable. Within this component, KPIs are calculated by using data from the production processes managed within the Framework in order create a concise representation of certain parts of the ecosystem. For example, if the worker classifies a part as scrap, further information about the quality issue can be captured with the device. This way the bare identification is enriched with additional contextual information to foster the understanding of decision makers to be alerted.

[1] https://www.knowage-suite.com/.

Table 4. Modules of the front-end application

Scene	
CursorManager	Handles cursor type and position in a scene. The cursor can be visualized in different forms for different interaction states
SceneManager	Manages consistency check, loading and unloading of all 3D objects in a scene
TaskManager	Evaluates user operations against a given work order (or task list)
Input	
InputManager	Collects all input actions (gaze, gesture, voice commands) and raises events/callbacks
GazeManager	Raycasts the 3D scene at the position the user is looking at and triggers events when capturing interactive objects
GestureManager	Adapter that interprets and translates a subset of gestures as input events
VoiceCommand Manager	Adapter for interpretation and translation of a subset of voice commands in input events
Interaction Manager	Guides the camera by using head tracking, manages the animation timing of interactive objects and translates input actions in possible events and commands
Spatial Mapping	
AnchorManager	Retrieves, clears and stores anchors (the location of physical objects in the real world) locally on the device, linking and updating an associated real transformation matrix on the scene graph
MarkerManager	Recognizes an AR marker (e.g. QR-code) in the scene, triggers an event when the marker is on line of sight and identifies if the marker is a new one or already registered

KPIs calculated in the Monitoring module are used for further analytic steps by being stored in the A4BLUE Adaptive Framework and processed by the DSS component, which is shown in Fig. 3. Input for the DSS are both historical production data and events having their root cause in the use cases assembly ecosystem. Data and information are displayed to the production supervisor through a web-based graphical user interface allowing the user to access relevant information for the support of decision making (DSS Dashboard).

In the background, the Event Manager processes events related to specific interventions, such as collaboration, maintenance or inspection intervention requests. These results are adapted by the DSS Event Adapter to translate them into notifications and to identify the correct receiver as well as the appropriate notification channel. The DSS Notification Management implements the business logic to notify the target user via an interaction device by selecting the most appropriate notification channel (e.g. graphical user interface, push notification, email).

Visualization, the third step of the considered data value chain, therefore is partly covered by the DSS components' Dashboard as described above. Furthermore, the AR device is used for the visualization as well. This involves the component VR/AR based

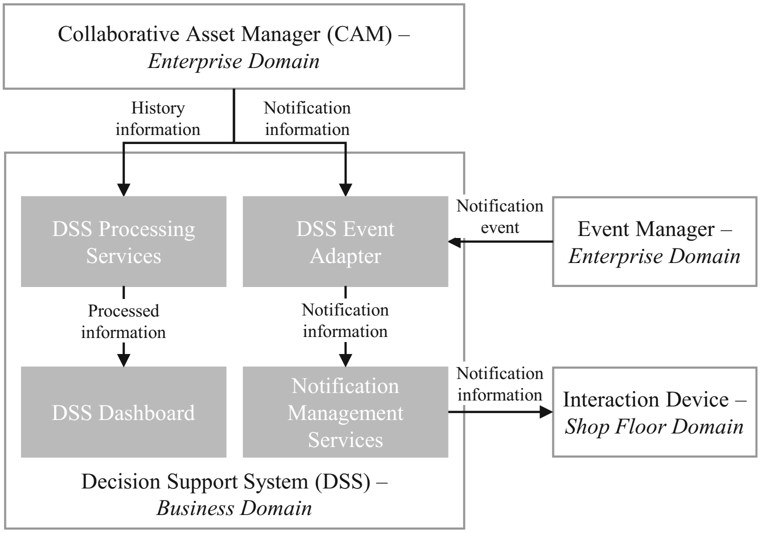


Fig. 3. Decision Support System (DSS) high-level decomposition

Training and Guidance again, which enables the display of notifications directly to the worker on the shop floor. In addition to event driven notifications the component is used to display assembly instructions with a varying level of detail depending on workers' characteristics and experiences, both stored and managed in the A4BLUE Adaptive Framework.

3.2 Exemplary User Story

Subsequently, a representative user story is described in order to outline how the information flow between workplace, worker and production supervisor is supposed to be managed for both assembly processes. The user story is visualized in Fig. 4 and shows how the worker is supported in the case of rework. This user story represents a small extract of the use cases overall capabilities under consideration of the A4BLUE solutions.

Firstly, the use case is configured by connecting the Automation Mechanisms to be used via the Devices Managers. Additionally, a complete digital image of the assembly ecosystem is generated by the production supervisor using the Collaborative Asset Manager. This image includes a semantic representation of all the tangible (e.g. tools, equipment, personnel, etc.) and intangible assets (e.g. processes, knowledge, etc.) involved in the use case to provide it throughout the Framework. In case of changes or modifications of the use case, the representation can be adapted via the Collaborative Asset Manager.

In regular shop floor operation, the worker initiates the assembly process manually, by looking at a marker, making a voice command or a gesture being captured by the AR device. This way the data and information generated in the assembly procedure is assigned to the specific process and can be stored as well as retrieved from the

Framework for further processing. The process initiation is executed manually, since the use case is designed without the involvement of a legacy system, such as a production planning and control system. If a legacy system is used, the process initiation can be executed by the system automatically. To close an operation within the assembly process the worker again uses a marker, makes a voice command or a gesture. The assembly process itself, which includes an optical and functional inspection of the assembly object, is closed automatically by the system after the last operation. After closure, a query is displayed to the worker whether rework on the product is required. The query is answered by the worker throughout a voice command or a gesture. If rework is required, the worker is requested to give further information about this specific quality issue via predefined comments. Within the Monitoring component, the rework rate, the percentage of defective products in the total number of

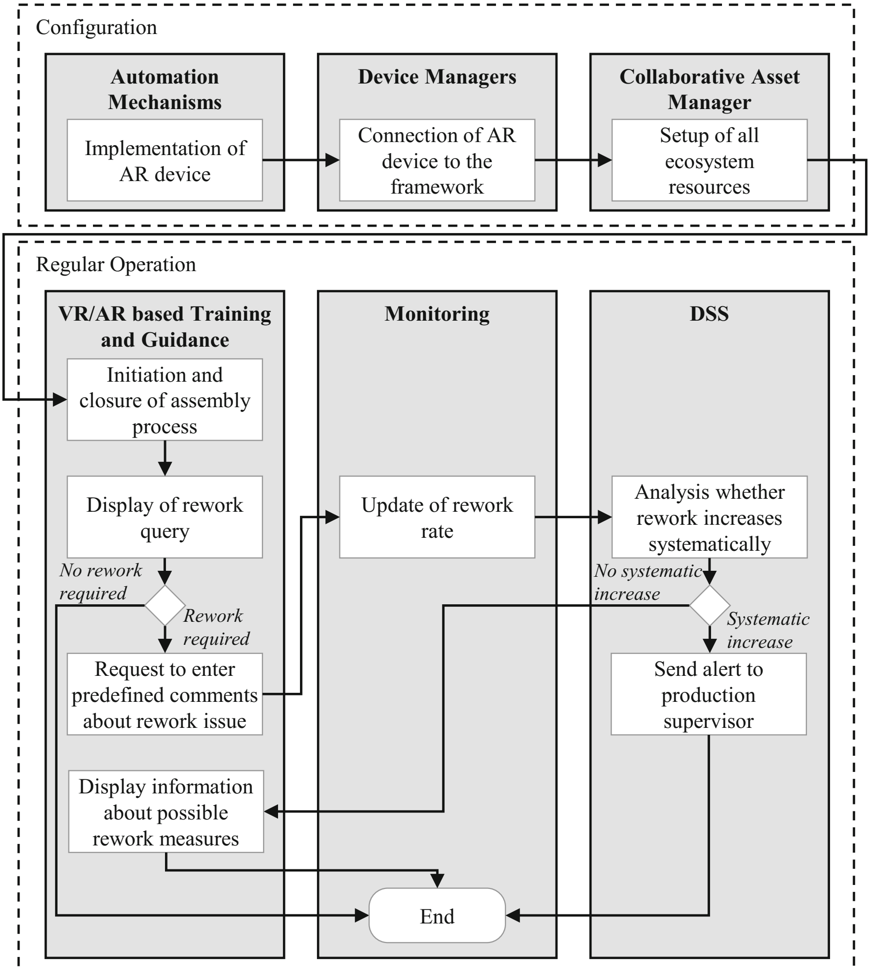


Fig. 4. Exemplary information process flow of an assembly rework process (simplified diagram)

products produced, is recalculated. Subsequently, the DSS analyses whether there has been a systematic increase of rework over the last processes.

If a systematic increase is identified, an alert notification (email or push notification) is displayed on the production supervisors smartphone to inform him to inspect the issue at the assembly object in person. Beforehand, the supervisor can retrieve the information about the issue entered by the worker that is provided by the DSS. If no systematic increase in rework was concluded, possible solutions for rework measures are provided to the worker. Afterwards the initiation of the next assembly process can be executed.

4 Conclusion

In the presented research work, concept and implementation of a digital infrastructure, the A4BLUE Adaptive Framework is presented. This Framework is the basis for the realization of Smart Workplaces as part of Smart Manufacturing. By connecting physical and virtual world, it enables the adaption of workplaces to the workers' specific profile. Additionally this empowers process as well as environmental variability. The goal is to combine the strengths of humans and automation mechanisms to achieve an increase in efficiency and to sustain worker satisfaction.

The A4BLUE Adaptive Framework contains a set of software components that are structured in three domains: Shop Floor, Enterprise and Business Domain. Each component is designed for a different purpose. Some components manage processes and databases, derive knowledge, trigger actions or analyze data, altogether realizing a Smart Workplace.

The Framework is implemented, amongst others, in an assembly use case at the Ramp-Up Factory at RWTH Aachen University. The use case provides an exemplary Smart Workplace in a real production environment using for example Augmented Reality, monitoring via key performance indicators and a system for decision support. Furthermore, hardware solutions and automation mechanisms can be connected to the Framework, while a complete digital image of the assembly ecosystem is managed and provided within the Framework.

As next steps the technical implementation will be finalized. Following experiments will be conducted to assess economic efficiency as well as worker satisfaction of the A4BLUE Adaptive Framework. To create a quantitative performance reference, the experiments in the first instance are conducted to a conventional workplace. Afterwards they are repeated with the implemented Framework. The results are expected to proof the potential of Smart Workplaces. Therefore, key performance indicators are used, being significant for a company's short- and long-term success, with a focus on economic efficiency as well as worker satisfaction.

Acknowledgements. The research work has been conducted by the Laboratory of Machine Tools and Production Engineering (WZL) and the chair of Production Engineering of E-Mobility Components (PEM) at RWTH Aachen University within the research project A4BLUE, funded by the European Commission within the HORIZON 2020 program; Grant Agreement No. 723 828.

References

1. Lu, Y., Morris, K.C.: Current standards landscape for smart manufacturing systems. National Institute of Standards and Technology (2016)
2. Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T., Schlund, S.: Produktionsarbeit der Zukunft. Industrie 4.0. Fraunhofer IAO, Fraunhofer Verlag (2013)
3. Lado, M.: Building a Showcase Culture: Powerful and Practical Keys for Manufacturing. Global Manufacturing Services (2018)
4. Rasmussen, J.: The role of hierarchical knowledge representation in decisionmaking and system management. IEEE Trans. Syst. Man Cybern. **SMC-15**(2), 234–243 (1985)
5. Agustiady, T.: Communication for Continuous Improvement Projects. CRC Press, Boca Raton (2014)
6. Zezulka, F., Marcon, P., Vesely, I., Sajdl, O.: Industry 4.0 – an introduction in the phenomenon. IFAC-PapersOnLine **49**(25), 8–12 (2016)
7. Lin, S.-W., Crawford, M., Mellor, S.: The Industrial Internet of Things - Volume G1: Reference Architecture. Industrial Internet Consortium. https://www.iiconsortium.org/IIC_PUB_G1_V1.80_2017-01-31.pdf
8. FIWARE: Smart Industry. Data is driving Industrie 4.0. https://www.fiware.org/community/smart-industry/
9. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)

Analysing Body Motions Using Motion Capture Data

Martin Benter(✉) and Peter Kuhlang

MTM-Institut, Deutsche MTM-Vereinigung e. V. (DMTMV), Eichenallee 11, 15738 Zeuthen, Germany
{Martin.Benter,Peter.Kuhlang}@dmtm.com

Abstract. Analysing manual work is an important task in industrial companies with high labour costs and labour intensive work processes. Industrial Engineer s can use the results of these analyses to identify potentials and improve productivity to maintain and improve competitiveness. To obtain goal-oriented results the processes need to be analysed in detail. One method that yields detailed information is the MTM-1-method. However, it requires a lot of effort and special knowledge. This presents a big hurdle, especially for companies with small production quantity. One option to reduce the effort and the required knowledge is using motion capture technology. This technology is capable to record human motions and to translate them into data that can be processed digitally. Known representatives are 3D cameras and motion capture suits. They track positions and postures of the human body, thus allowing conclusions about human movements. This paper presents an approach to detect body motions in accordance to MTM-1 using motion capture data from 3D cameras. The detected motions are then analysed with respect to productivity to show improvement potential.

Keywords: Motion capture · Labour productivity · Body motions

1 Introduction

Production Costs are one of the key aspects for manufacturing companies when facing national or international competition. One major portion of these costs are labour costs, especially if the production requires a high amount of manual processes and the wages are high. The most important key indicator to control these costs is labour productivity. Analysing labour productivity allows the identification of improvement potentials, thus helpings in choosing appropriate measures to increase productivity.

In general, productivity is the ratio between output and input of a system. The output in manufacturing areas usually consists of produced goods while the input consists of necessary resources to create the output. The main resources in most companies are human and mechanical work. Labour productivity is therefore the relation between produced goods and staff working time [1, 2]. As the output of a manufacturing company is usually determined by the market, the working time is

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 128–140, 2020.
https://doi.org/10.1007/978-3-030-20040-4_12

analysed in detail to control and improve labour productivity. One established method to analyse working time in detail is the MTM-1-method [3, 4]. It yields reliable times for analysed processes and creates a foundation for a target-oriented improvement process.

2 Assessment of Manual Working Time

2.1 Recording Working Time Using the MTM-1 Method

The MTM-1 method divides work processes into basic motions. It thus describes work in a very detailed way. Afterwards, influencing factors are determined for every motion. With that information, the MTM-1-time can be derived [3, 4].

MTM-1 Motions. There are 25 different basic motions. They can be divided into *Basic Motions*, *Visual Functions*, and *Body Motions* [4]. Figure 1 shows an overview of these motions.

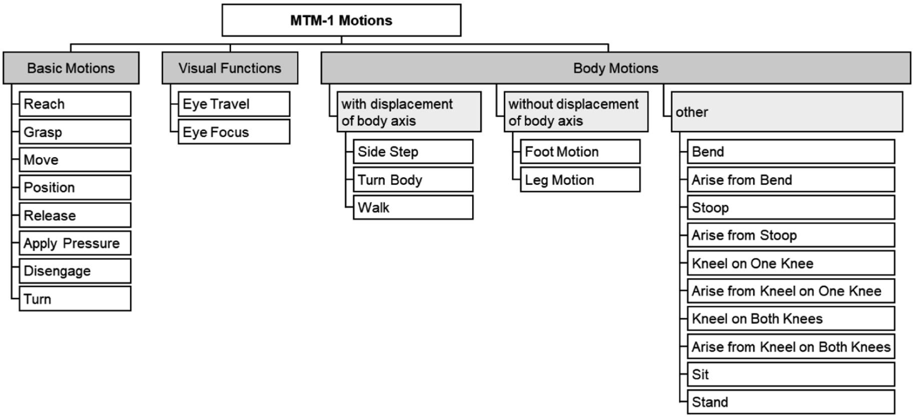


Fig. 1. Overview of the MTM-1 motions

In large-scale production, Body Motions usually do not play a crucial role, especially for assembly lines. In the case of single and small series production as well as logistics processes however, Body Motions like movements between workstations become more relevant.

Influencing Factors. MTM-1 defines specific influencing factors for each Motion type, which the user must select to determine times. Below are the factors of the MTM-1 Motion 'Walk' [3, 4].

Number of Paces describes the distance travelled during the Motion and has a major impact on the required time. In general, Motion Length plays an important role for many Motion types within the MTM-1 method.

The second influencing factor is *Type of Walk*. 'Walk' has two Types: With, and without obstruction. An example for an obstruction is uneven or slippery ground.

MTM-1 Coding. The influencing factors can be coded to ensure a short and clear documentation. Figure 2 shows an example for the movement 'Walk'. In this case, 'W' stands for the Motion type 'Walk', followed by the Number of Paces and the Motion Case. Here, 'PO' stands for pace obstructed.

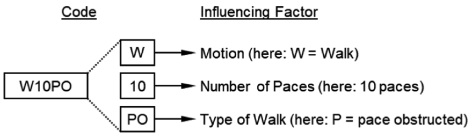


Fig. 2. Coding example for the MTM-1 motion walk

Deriving MTM-1 Times. One can now use the determined influencing factors to derive the required time for the Motion. In the MTM-1 method, these predetermined times are documented on data cards. The times are average values that skilled workers can execute without fatigue over a long period of time [4]. The user executes this procedure for all steps of the work process and thereby determines the overall time.

Advantages and Disadvantages. The MTM-1 method has several advantages: Firstly, it is very detailed, as it breaks the processes down into basic motions. By determining influencing factors, it also yields logical starting points for improvement processes. Additionally, it allows for a comparison with actual process times.

However, there are two disadvantages: The high level of detail prevents a comparable, practicable recording of actual times with existing methods [5]. Furthermore, the method requires expert knowledge, and induces relatively high effort. Those facts are preventing the application in small and medium-sized companies.

2.2 Analysing Working Time by Identifying Wastes

One widely used approach to analyse manual working time is the division of processes into value-adding and non-value-adding processes. In particular, Lean Management with its methods for avoiding non-value-adding processes has made its way into many companies [6]. A common classification divides work processes into *value adding processes* (e.g. welding or assembly processes), *non-value adding but necessary processes* (e.g. the preparation of welding edges before welding), and *non-value adding processes* (e.g. waiting, searching) [6, 7].

Liker defines seven types of waste to help users identify non-value adding process steps and categorize typical problems [7]:

- high inventory,
- unnecessary movement,
- rejects and rework,
- waiting times,
- overproduction,
- unnecessary transport, and
- wrong technology/errors during processing.

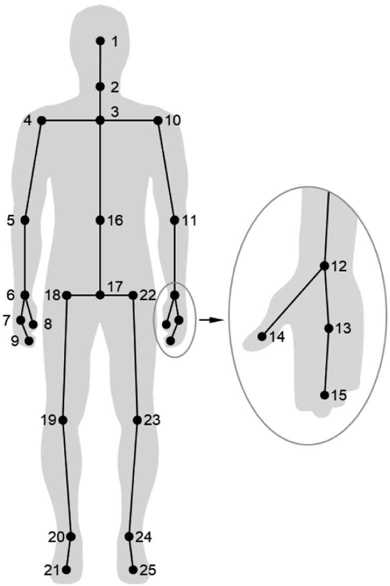
These definitions to identify non-value adding activities is applicable for all manual work processes. In addition, the method is easy to understand. The method is usually used in the evaluation of actual times. However, in general it is also suitable for analysing the times generated by the MTM-1 method.

3 Motion Capture with 3D Cameras

Motion Capturing technologies can detect and track human motions, which can then be processed digitally [8, 9]. There are different technological approaches, such as optical, electromechanical, electromagnetic or acoustic [9]. Popular examples are 3D cameras, such as the Microsoft Kinect or the Intel RealSense, and Motion Capturing suits like the XSens suit or the AXS suit.

In case of 3D cameras, which belong to the optical systems, a transmitter emits an optical signal, usually infrared waves. Observed objects then reflect it and a receiver record them [9]. The Microsoft Kinect, which is used in this presentation, belongs to the non-marker-based 3D cameras. This means that it utilizes the natural reflection of the body. This way, the observed person does not have to wear any potentially disturbing devices.

The Kinect determines the distance between itself and reflecting objects via the time-of-flight of the infrared wave [10], creating a partial 3D model of the observed area. The Kinect then compares this model with a database to detect people and their postures. If it detects persons, it determines the position of 25 body joints per person (see Fig. 3) [11]. Additionally, it is capable of recording a regular video of the process.



No.	Joint name
1	Head
2	Neck
3	Shoulder, Mid
4	Right Shoulder
5	Right Elbow
6	Right Wrist
7	Right Hand
8	Right Thumb
9	Right Hand Tip
10	Left Should
11	Left Elbow
12	Left Wrist
13	Left Hand
14	Left Thumb
15	Left Hand Tip
16	Spine, Middle
17	Spine, Base
18	Right Hip
19	Right Knee
20	Right Ankle
21	Right Foot
22	Left Hip
23	Left Knee
24	Left Ankle
25	Left Foot

Fig. 3. Body joints tracked by the Kinect

Though other Motion Capture technologies work in different ways, most of them yield similar results. The AXS suit for example uses Inertial Measurement Units to record angles and accelerations and then calculates the position of the observed person and its posture [12]. This way, it also determines the positions of certain body joints.

4 Experimental Set-Up

To demonstrate the approach, an exemplary assembly task was recorded. In this task, the worker assembled gearboxes for 20 min. The gearboxes consist of two housing parts and 16 further individual parts that the worker assembled at three different workstations. Figure 4a shows the layout of the workstation. The worker assembled eight gearboxes in 14 min and 48 s.

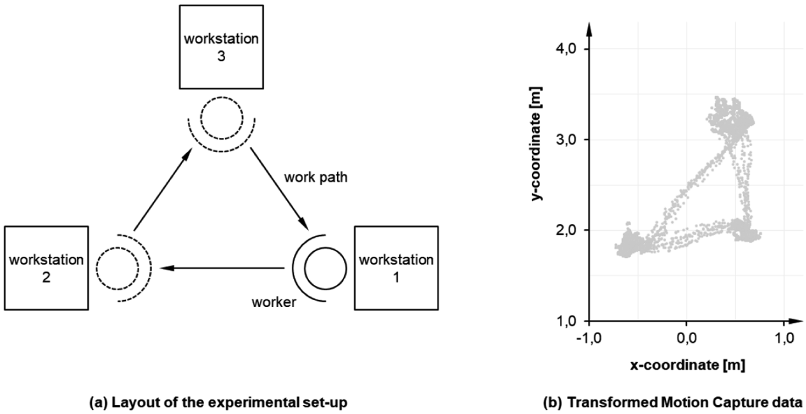


Fig. 4. Experimental set-up – layout and motion capture data

5 Pre-processing the Motion Capture Data

5.1 Data Sorting and Transformation

The usability of the recorded data to determine reliable times heavily depends on data quality. Objects may obscure the line of sight between the camera and the worker. Then, either it does not detect the joints or it calculates faulty coordinates. Furthermore, even with an unobstructed view, it does not always interpret the 3D data correctly. Therefore, the first step is to sort out incorrectly generated data. This can happen automatically, for example via detecting extreme or inconsistent values, or manually by examining all points using the video recording [5].

The next step is the data transformation, which transforms the coordinates into a coordinate system that can be assessed in a standardized and understandable way. Depending on the observed work system, the user might have to tilt the Kinect to record the worker during all work steps. For example, when material boxes or shelves prevent the positioning directly in front of the worker. The Kinect angle is then used to transform the data into a coordinate system, where x and y span the horizontal plane and z points upwards. Additionally, the user can turn the coordinate system to match it to existing layouts (e.g. Fig 4a). The result can be visualized in a top view diagram (see Fig. 4b). Here every point represents the coordinates of one recording during the process.

5.2 Calculation of Motion Parameters

The next step to prepare the Body Motion recognition is the calculation of relevant Motion Parameters. These Parameters describe important variables such as joint velocity to identify Motions later on. For example, the identification of the Motion 'Walk' requires the Parameters: horizontal distance covered of the joint 'Spine, Base' between two recordings and its horizontal velocity. Formulas 1–3 show exemplary Parameters and their calculation. Other relevant Parameters are calculated analogically.

$$d_{j,a,b} = \sqrt{(x_{j,b} - x_{j,a})^2 + (y_{j,b} - y_{j,a})^2 + (z_{j,b} - z_{j,a})^2}. \tag{1}$$

$$d_{h,j,a,b} = \sqrt{(x_{j,b} - x_{j,a})^2 + (y_{j,b} - y_{j,a})^2}. \tag{2}$$

$$v_{h,j,a} = \frac{d_{h,j,a-1,a}}{t_a - t_{a-1}}. \tag{3}$$

$d_{j,a,b}$: distance of joint j between recording a and b [m]
$d_{h,j,a,b}$: horizontal distance of joint j between recording a and b [m]
$v_{h,j,a}$: horizontal velocity of joint j at recording a [m/s]
$x_{j,a}$: x-coordinate of joint j at recording a [m]
$y_{j,a}$: y-coordinate of joint j at recording a [m]
$z_{j,a}$: z-coordinate of joint j at recording a [m]
t_a: time of recording a [s].

6 Recognition of Body Motions

6.1 Definition of Worker States

The first step in recognising Body Motions is the definition of possible states of the worker. The purpose is to be capable to assign a state type to each timeframe of the recording. The definition follows the basic principle that the worker either performs a Motion or does not move. Figure 5 shows the worker states of the developed approach. To ensure a reliable recognition, it only covers Motions that can be derived by assessing the Motion Capture data of the trunk and the head. These body parts have a relatively high tracking quality when using 3D cameras like the Kinect [5].

The Motions (see Fig. 5) were defined in accordance to MTM-1. This includes the Motions 'Bend', 'Sit', 'Walk', Sidestep', and 'Turn Body'. One difference is the combination of the Motions 'Stoop' and 'Kneel'. The body postures in these movements is very similar and therefore difficult to distinguish. Additionally, all 'Arise' Motions have been integrated into their respective Motions, since they follow the same sequence of postures, with the distinction of the reversed sequence. Furthermore, 'Foot Motion' and 'Leg Motion' are not included due to the relatively low tracking quality of the legs.

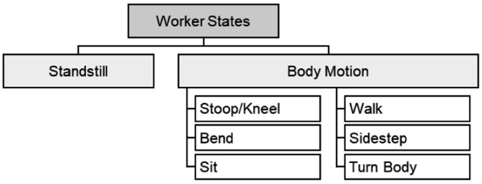


Fig. 5. Recognised worker states

6.2 General Approach to Recognise Worker States

The Recognition of Worker States is based on the differentiation between Standstills and Body Motions. Firstly, it identifies Motions by finding characteristic Parameter Sequences in the Motion Capturing data [5]. For example, a temporarily increased velocity of the hip is characteristic for the Body Motion 'Walk'. Afterwards, it logically infers the remaining time segments as Standstills. Figure 6 visualizes this approach.

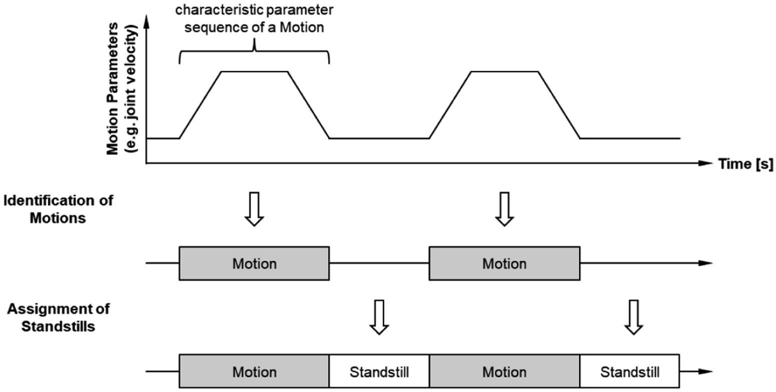


Fig. 6. General approach of the state recognition

Identification of Motions. The identification through characteristic parameter sequences follows three steps (see Fig. 7) [5]:

1. *Motion Localization*: The Localization scans the data for characteristic values or curves that represent specific Motion Types. For example, an increased speed of the joint ‘Spine, Base’ characterises the ‘Walk’ Motion. The Localization therefore scans for sequences in which this velocity has a high value.
2. *Determination of start and end point*: In addition to the localized time sequences, the complete Motion usually includes more than the localized recordings. For example, the velocity at the beginning of a Motion only gradually increases to the value that the Localization uses.
3. *Motion Verification*: Verification of the Motion completes the Identification process. It checks whether a Motion has actually taken place during the specific time segment. In the case of ‘Walk’, it is done by checking the distance travelled between the start and the end point of the Motion. If the distance falls below a minimum value, the Motion has been falsely identified.

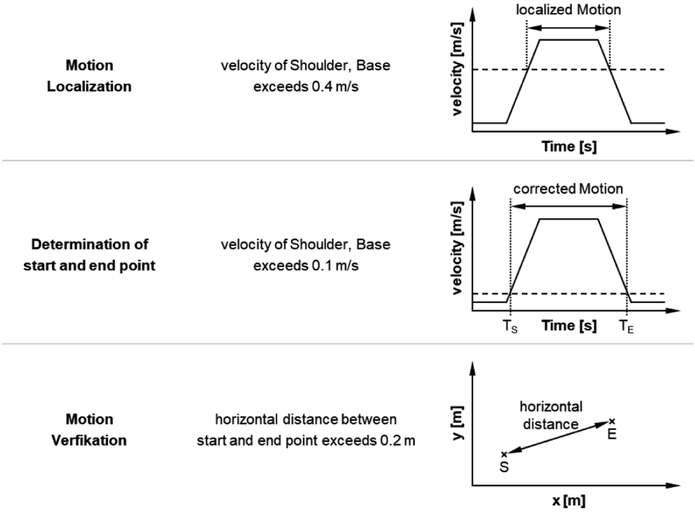


Fig. 7. Motion identification for the Motion ‘Walk’

Following the Motion identification, the remaining time sequences are marked as Standstill times.

6.3 Identification of Body Motions

The previous explanations describe the identification process for the Body Motion ‘Walk’. Table 1 shows an overview of the complete Identification Parameters for all Body Motions. One can for instance see that the identification of the Motion ‘Sidestep’ uses the same Parameters as the Motion ‘Walk’. This is because the movements are very similar and only differ by the Motion direction.

Table 1. Body motion identification parameters

Body motion	Identification parameters		
	Motion localization	Determination of start and end point	Motion verification
Stoop/Kneel	Vertical velocity of head	Vertical distance of head from start to end	Vertical angle of thighs, vertical angle of lower legs
Bend			
Sit			
Walk	Horizontal velocity of spine, base	Horizontal distance of spine, base from start to end	Horizontal angle between motion direction and hip
Sidestep			
Turn body	Angular speed of Shoulders	Change of horizontal angle of Shoulders from start to end	

7 Analysis of the Recognized Body Motions

7.1 Determination of MTM-1 Times

The Analysis firstly aims to assign MTM-1 Motions to the recognized Motions and to determine the influencing factors. Thus, the first step is matching MTM-1 Motions to the recognized ones. Figure 8 provides an overview of the corresponding Motions.

One can see that during the state 'Standstill' the worker can perform the Motions 'Foot Motion' and 'Leg Motion' as well as Basic Motions and Visual Functions. These

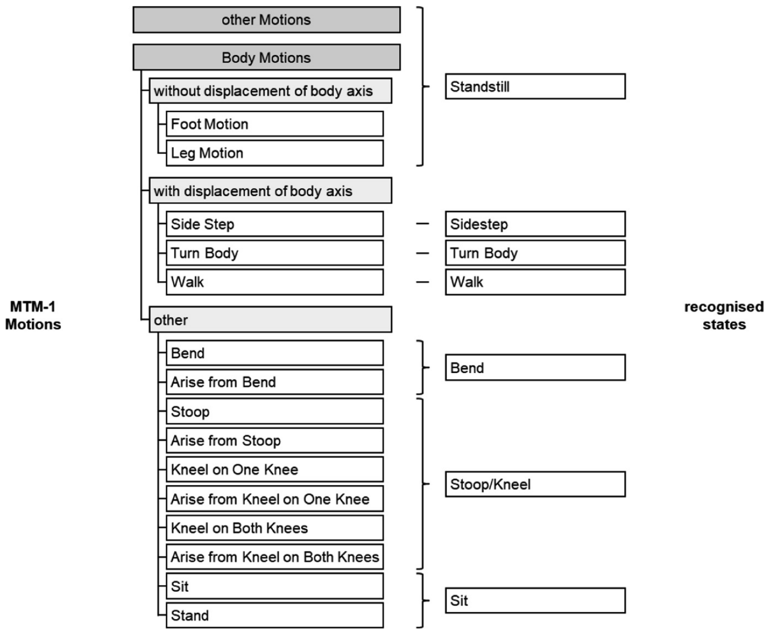


Fig. 8. Comparison of MTM-1 motions and recognised worker states

Motions are not in the scope of the Body Motion recognition. Furthermore, the Motion ‘Stoop/Kneel’ corresponds to several MTM-1 Motions. The distinction between these Motions is not included here, but can be integrated if the data quality of the legs is high. In this case, the correct MTM-1 Motion was chosen manually. Although the recognition does not define different upwards and downwards Motions, it still recognizes the direction during the identification by observing the sign of the covered vertical distance (see Table 1).

The next step is the determination of the influencing factors for the identified MTM-1 Motions. Table 2 shows the influencing factors and specifies, whether they can be determined automatically by analysing the Motion Capture Data.

For the MTM-1 Motion ‘Sidestep’, the factor Motion Length can be assessed by calculating the travelled distance of the hip. The Motion Case on the other hand indicates if the person moved one or both legs. This information can be derived if the data quality is high. In this example, it was chosen manually.

This applies accordingly to the Motion Case of ‘Turn Body’ and the Number of Paces of ‘Walk’ as well. In the case of Walk however, it is possible to determine the number of paces by assessing the Motion Length [4].

One can see that it is possible to derive MTM-1 times with relatively small effort for Body Motions. This effort would be even smaller, if a high tracking quality could be guaranteed.

Table 2. Deducible influencing factors of the MTM-1 motions

MTM-1 motion	Influencing factors	Deducible from motion capture data
Sidestep	Motion length	Yes
	Motion case	Yes (if data quality is high)
Turn body	Motion case	Yes (if data quality is high)
Walk	Number of paces	Yes (if data quality is high)
	Type of walk	No
Bend, Arise from Bend, Stoop, Arise from Stoop, Kneel on One Knee, Arise from Kneel on One Knee, Kneel on Both Knees, Arise from Kneel on Both Knees, Sit, Stand	None	

7.2 Identifying Productivity Potential

After determining the MTM-1 Motions and their influencing factors, the analysis uses this information to identify productivity improvement potential. It differentiates between two kinds of potential: *speed loss* and *layout waste*.

It determines *speed loss* by comparing the actual time of the detected Motion with the MTM-1 time (see Formula 4). A positive value reflects a work speed lower than

MTM-1, while a negative value represents a higher speed. The relative speed of a Motion is relatively comprehensible and describes the ratio between the MTM-1 time and the actual time (see Formula 5). For example, a relative speed of 50% means that the actual time is twice as long as the MTM-1 time. Thus, it is similar to the performance defined by the MTM-1 method [4]. Formulas 6 and 7 demonstrate how the overall speed loss and relative speed of a process are calculated. Table 3 shows the calculated speed losses for all process steps in the shown example.

Layout wastes result from unfavourable layout design. They appear when the worker travels a longer distance than is theoretically necessary. Thus, mainly 'Walk' Motions produce layout wastes. There are multiple ways to determine a necessary minimum distance [5]. This Analysis follows the assumption that all distances can be reduced to the distance of the shortest 'Walk' Motion. In the recorded example, the 'Walk' Motion between workstation three and one had the shortest distance (see Fig. 4b). The worker used three Paces to travel the said distance. Thus, the analysis assumes that the worker can execute all 'Walk' Motions in three Paces. The layout waste of each 'Walk' Motion then is the difference between the MTM-1 time of that Motion and the MTM-1 for walking three Paces (see Table 3). The overall layout waste can be calculated analogically using Formula 7.

$$sl_n = t_n - t_{MTM,n}. \tag{4}$$

$$srel_n = \frac{t_{MTM,n}}{t_n}. \tag{5}$$

$$sl_p = \sum sl_n. \tag{6}$$

$$srel_p = \frac{t_{MTM,p}}{t_p} = \frac{t_p - sl_p}{t_p}. \tag{7}$$

sl_n: speed loss of Motion n [s]
sl_p: speed loss of process p [s]
$srel_n$: relative speed of Motion n [-]
$srel_p$: relative speed of process p [-]
t_n: actual time of Motion n [s]
t_p: actual time process p [s]
$t_{MTM,n}$: MTM-1 time of Motion n [s]
$t_{MTM,p}$: MTM-1 time of process p [s]

Table 3 shows the result of the Analysis of the Body Motion. One can see that the overall speed loss is negative, which means that the worker executed the Motions faster than the MTM-1 times. A possible reason is the fact that the worker only assembled gearboxes for a short period of time, in which he might have moved faster than what is sustainable over a longer period. Furthermore, there is limited layout waste. This is because the experimental set-up used an optimised layout that was designed to reduce walking times.

Table 3. Result of the body motion recognition and analysis

Recognised body motion	Location	MTM-1 motion	Actual time [s]	MTM-1 time [s]	Speed loss [s]	Layout waste [s]
Standstill	*WS1*	-	**34.2*	-	-	-
Turn body	WS1	Turn body (TBC 1)	0.6	0.7	−0.1	0.0
Walk	WS1 → WS2	Walk (W3P)	1.8	1.6	0.2	0.0
Sit (downwards)	WS2	Sit (SIT)	0.9	1.2	−0.3	0.0
Standstill	*WS2*	-	**35.7*	-	-	-
Sit (upwards)	WS2	Stand (STD)	0.7	1.6	−0.9	0.0
Turn body	WS2	Turn body (TBC 1)	0.7	0.7	0.0	0.0
Walk	WS2 → WS3	Walk (W4P)	2.1	2.2	−0.1	0.6
Standstill	*WS3*		**29.7*	-	-	-
Turn body	WS3	Turn body (TBC 1)	0.7	0.7	0.0	0.0
Stoop/Kneel (downwards)	WS3	Stoop (S)	1.0	1.0	0.0	0.0
Stoop/Kneel (upwards)	WS3	Arise from stoop (AS)	0.8	1.1	−0.3	0.0
Turn body	WS3	Turn body (TBC 1)	0.7	0.7	0.0	0.0
Walk	WS3 → WS1	Walk (W3P)	2.0	1.6	0.4	0.0
Sum (body motions)			12.0	13.1	−1.1	0.6
Percentage [%]			100.0	109.2	−9.2	5.0

8 Summary

This article presents an approach to recognize Body Motions using 3D cameras. Furthermore, it determines corresponding MTM-1 Motions and times. It identifies speed losses and layout waste, thus uncovering improvement potential. In addition to the benefits of the classical MTM-1 method, the approach has the following advantages:

- *Assessment of actual times:* The presented method allows recording actual times and comparing them with MTM-1 times.
- *Identification of productivity potential:* By assessing speed loss and layout waste, it presents the user with logical starting points for an improvement process.
- *Semi-Automated:* By using Motion Capture data from the Microsoft Kinect, most method steps can be automated, resulting in a relatively low method effort.

Some aspects allow further development. The methodology requires a reliable data quality. The Microsoft Kinect only partly fulfils that requirement. The general approach however does not require the Kinect as the Motion Capture Technology. Thus, the used calculations and values need to be adapted to other technologies. Another possible improvement is the integration of existing ergonomic assessments like the EAWS [12].

References

1. Sumanth, D.J.: Productivity Engineering and Management: Productivity Measurement, Evaluation, Planning, and Improvement in Manufacturing and Service Organizations. McGraw-Hill, New York (1984)
2. Weber, H.K.: Rentabilität, Produktivität und Liquidität. Größen zur Beurteilung und Steuerung von Unternehmen. Gabler, Wiesbaden (1998)
3. Bokranz, R., Landau, K.: Produktivitätsmanagement von Arbeitssystemen: MTM-Handbuch. Schäffer-Poeschel, Stuttgart (2006)
4. MTM-1. Lehrgangsunterlage; A/AB. Deutsche MTM-Vereinigung e.V, Hamburg (2014)
5. Benter, M.: Analyse von Arbeitsabläufen mit 3D-Kameras. Dissertation. TUHH, Institut für Produktionsmanagement und -technik, Hamburg (2018)
6. Womack, J.P., Jones, D.T.: Lean thinking. Ballast abwerfen, Unternehmensgewinne steigern. Campus, Frankfurt am Main (2004)
7. Liker, J.K.: The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer. McGraw-Hill, New York (2004)
8. Jackèl, D., Nenreither, S., Wagner, F.: Methoden der Computeranimation. Springer, Berlin (2006)
9. Kitagawa, M., Windsor, B.: MoCap for Artists: Workflow and Techniques for Motion Capture. Elsevier/Focal Press, Amsterdam (2008)
10. Steward, J., et al.: Performance assessment and calibration of the Kinect 2.0 time-of-flight range camera for use in motion capture applications. In: Conference Papers of FIG Working Week, Sofia, Bulgarien, pp. 1–14 (2015)
11. Zhang, Z.: Microsoft Kinect sensor and its effect. IEEE Multimed. **19**, 4–10 (2012)
12. Benter, M., et al.: Automatisierung von Ergonomiebewertungen durch Bewegungserfassung am Beispiel des Ergonomic Assessment Worksheet (EAWS). In: Gesellschaft für Arbeitswissenschaft e.V.: Gestaltung der Arbeitswelt der Zukunft, Dortmund (2018)

Automatic Generation of Methods-Time Measurement Analyses for Assembly Tasks from Motion Capture Data Using Convolutional Neuronal Networks - A Proof of Concept

Jochen Deuse[1(✉)], Lukas Stankiewicz[1], Ronny Zwinkau[1], and Frank Weichert[2]

[1] Institute of Production Systems, TU Dortmund University, Leonhard-Euler-Straße 5, 44227 Dortmund, Germany
{Jochen.Deuse,Lukas.Stankiewicz, Ronny.Zwinkau}@ips.tu-dortmund.de
[2] Chair for Computer Science VII, TU Dortmund University, Otto-Hahn-Straße 16, 44227 Dortmund, Germany
Frank.Weichert@tu-dortmund.de

Abstract. This paper describes the research hypothesis that motion data can be utilized to derive MTM analyses. As a first step, manual assembly tasks are recorded with motion capture systems to generate motion data. These motion data are used as a training data set for an end-to-end deep learning architecture for motion classification. The result of this classification is the assignment of data sequences to corresponding basic motions of MTM-1. The paper also describes the prerequisites for an automatic generation of MTM analyses by considering an adaptation of the original MTM methodology to fit for an automatic approach, the acquisition of motion capture data and the automatic annotation of motion data.

Keywords: Motion capture · Motion recognition · Methods-time measurement · Neuronal networks · Geometric deep learning

1 Motivation

The global market situation leads to increasing challenges for manufacturing companies. To maintain or to regain competiveness within their competence areas, globally acting companies need to optimize production processes continuously. This in particular applies to manual work as it is often associated with high labor costs and high potential for further economic improvements [1]. Particularly in high-wage countries such as Germany, the existing potential therefore must be used extensively [2]. A high percentage of manual labor occurs in assembly tasks [1]. Assembly workstations show an increasing planning frequency due to short product life cycles as well as an increasing planning effort due to complex products and processes [3]. Thus, a planning

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 141–150, 2020.
https://doi.org/10.1007/978-3-030-20040-4_13

support for the design of manual workstations is needed to minimize the planning effort on the one hand and increase the planning quality on the other hand.

Following the vision of Industry 4.0, these challenges of cyber-physical systems and the digital transformation could be resolved through a modern concept of machine learning. We propose to use convolutional neural networks (CNNs) adapted to graphs, representing the skeletal graph model, as an alternative approach for human pose estimation in real-time to derive MTM codes. The main objective of this publication is to provide the potentials, advantages, and open questions of this concept.

1.1 Production Planning

There are several tools and approaches for production planning. One approach focuses on the application of integrated modules in enterprise resource planning (ERP) software like SAP or Sage. To gain accurate planning data, e. g. in the fields of planning of machine utilization, price calculation, capacity planning, scheduling, capacity analysis and personnel planning there is a need for extensive and current time data [4]. However, valid time data in terms of extent and level of detail are often not available in production companies [5]. Surveys concerning the application of industrial data science show that despite of an increasing data volume in companies there still is a general lack of requisite data for precisely focused analyses [6–9]. This trend is present in enterprises across all sectors and company sizes but it especially relates to time data regarding small and medium-sized companies (SME) [10]. Even though time management is of high importance for work system design, it does not have an equivalent prioritization within SME due to the high effort involved in acquiring time data.

1.2 Acquisition of Time Data

There are two commonly used methods to gather time data: Recording and processing of actual times (e. g. REFA) and assessing target times by predetermined motion time systems (e. g. Methods-Time Measurement (MTM)). For the measurement of working times according to REFA, work activities are measured with a chronograph or another timing device. In addition, a trained REFA expert has to assess the performance level of the observed employee. To derive the target times this factor needs to be allocated with the actual time. In order to achieve reliable target times, the expert needs to capture several work processes to ensure sufficient data accuracy. The effort for REFA time studies is high. It consists of the preparation including distributing the work content into work cycle elements, determining the extant and timing of the time studies, informing all participants and defining the influence factors and others [11, 12]. The actual time recording requires the person performing the studies to be present at the work process for the entire duration of all cycles.

Predetermined motion time systems follow the approach that small units of basic motions compose a motion sequence. Each basic motion is assigned to a time value, which are summed up to generate the overall target time. The time values for a motion sequence result from the combination of a basic motion and the corresponding influencing factors. In comparison to time studies according to REFA, there is no need for subjective assignments of the performance level. Furthermore, there is no need to analyze several process cycles. Beyond that, the detailed planning of the work process offers the possibility of deriving process optimizations. A well-established predetermined motion time system in Germany and other industrial states is MTM [13]. For MTM time studies, there is only the need to observe one cycle of work content. Subsequently, up to 200 times of the observation time is required to perform the analysis, depending on the MTM method level [14]. The duration of the analysis depends on the experience of the employee and requires extensive training in MTM. The same counts for REFA time studies. Both approaches, time recording in accordance to REFA and predetermined motion time systems in accordance to MTM-1 have a high expenditure of time for analysis and evaluation in common.

This leads to the point that there is a global need for a tool to accelerate and simplify time studies. The approach presented in this paper bases on recent studies and developments in the area of motion capture as well as machine learning. The following chapters will provide a proof of concept, concerning the generation of motion data by means of infrared-based motion capture systems and the automatic classification of different basic motions in assembly to derive MTM codes. Taking into account the presented aim, time analyses in accordance to MTM are more suitable than REFA methods. MTM offers the possibility to derive weaknesses in work processes more directly and concise. Moreover, the use of standard basic motions is an enabler to detect different motion sequences automatically by identifying typical characteristics of each motion sequence. Therefore, MTM serves a foundation to perform time studies automatically by identifying time data and process modules.

2 Investigation Domain and Data Generation

To reach the mentioned aim there is the necessity to collect motion data from manual assembly tasks. Therefore, Sect. 2.1 describes the setup of a work system for motion captures in a test environment of the Institute of Production Systems. Beside the setup of the workplace, further prerequisites of the actual motion capturing for the work tasks and motions are presented in Sect. 2.2 in detail.

2.1 Experimental Setup

Different replications of industrial assembly workstations need to be recreated. These workstations have to represent a certain variety of work tasks so that capturing of various motion paths and motion forms is possible. Considering the aim of an automatic identification of MTM codes with motion data based on deep learning approaches, modeling expressive and generalizable CNN architectures is a significant issue. One challenge in this context is to prevent overfitting, for example. Overfitting appears if an analysis corresponds too close to a certain data set with the consequence that additional data cannot be predicted reliably [15]. The replications must fulfill the requirements of a wide variety of work tasks to not only fit to the assembly task in a given setup but apply on any manual work tasks. The replications pursue the goal to generate motion data in industry-like work systems and assembly tasks in a laboratory environment. The design of the workstations need to relate to real workstations accurately. This includes the work content, the workplace design, operating material and equipment. By replicating real workplaces, it is possible to adjust system settings and represent various work tasks in any assembly workplaces holistically. In order to maintain the naturalness of the movements on the replication, the equipment, tools, material and environment must be as close as possible to real workplaces. Hence, the replications are equipped with professional tools and supplies as far as possible and economically reasonable. These replications represent an assembly environment that gears towards current and future-oriented industrial standards. The aim is to cover all basic motions according to MTM-1 and their corresponding forms in the smallest possible number of different workplaces or modifications of a workplace. In a first setup for initial motion captures, a standard worktable system has been installed (s. Fig. 1). For cost reasons, the provision of parts for the workpiece carriers was implemented by means of cardboard engineering and wooden constructions, which do not have any impact on the test person's motion sequences. The use of flexibly applicable equipment allows for installing an almost unrestricted universal work system, capable of capturing all necessary motion patterns.

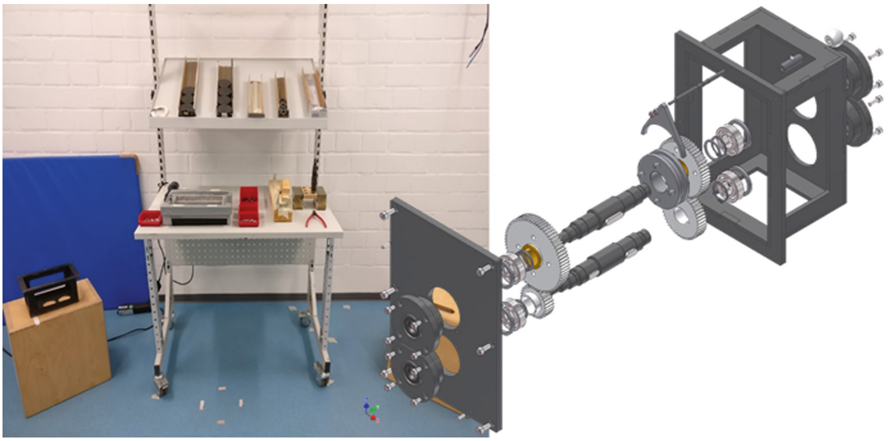


Fig. 1. Workplace setup for motion recordings of an assembly task (left) and the assembly object (right)

The assembly object (s. Fig. 1 – right) is a two-stage gearbox, which is used within the scope of various seminars and lectures at the Institute for Production Systems. This assembly object meets the need for tracking assembly motion patterns very well as there is a wide range of required assembly tasks. The assembly includes joining operations with and without tools, screwing operations, handling of light and heavy parts, low and high tolerances for marker points and the handling of mixed lying and separated material.

2.2 Generation of Motion Data

An essential step to derive MTM analyses of a given work task is to generate motion data that fit several prerequisites. Taking into account the goal of a detailed detection and annotation of transaction data using machine learning methods, it is necessary to obtain high-quality data for learning and for evaluating the classification results. In order to capture motion data extensively, a sequence of assembly task sequences representing a majority of possible motions needed to be defined. The workplace setup shown in Sect. 2.1 is based on various motion sequences and highly flexible. Thus, the setup effort for different types of systems can be reduced to a minimum, even though a universal workstation to cover various motion patterns and motion tasks is more complex.

The interpretation of captured motions depends on the classification task and the classes to be identified. In the given case, these classes are specified by MTM. The chosen approach follows the classification according to the basic movements as reported by MTM-1. Thus, a high method level is mapped. This offers the possibility, based on the results of this work, to derive classifications for lower method levels such as MTM-UAS or MTM-MEK as a combination of different MTM-1 codes. The MTM-1 data card thus forms the basis for deriving the motion sequences to be identified and classified by a machine learning algorithm. Restricted to the upper limbs, the basic motions are *reach*, *grasp*, *move*, *position* and *release*. For each of these motion sequences, there are different characteristics that influence the appearance of the motion to be distinguished which are shown in Table 1.

Table 1. Basic motions for the upper limbs and their influencing factors

Basic motion	Influencing factors
Reach	Motion length
	Motion case
	Motion path
Grasp	Type of gripping
Move	Motion length
	Motion case
	Motion path
Position	Class of fit
	Case of symmetry
	Handling
Release	Type of release

All basic motions and influencing factors mentioned in Table 1 need to be mapped in the given workplace setup to ensure the data set, to cover all information needed for the automatic allocation of MTM codes.

In addition to the influencing factors given by MTM, there are further factors to be considered like the direction of a basic motion, which is unimportant for a manual MTM analysis but significant for an automated analysis. Considering all influencing factors means that each possible combination of two or more factors has to be recorded, as the dependencies of various factors are not clear beforehand. This leads to a vast amount of required motion capture data which is not feasible to record. As an example, the direction of a basic motion in motion data is the vector of two coordinates of one marker point over a given period. The number of possible vectors for motions like *reach* or *move* is infinite and therefore cannot be considered completely. Even reducing the number of directions (left, right, up, down, forward and backward) leads to numerous possibilities in combination with influencing factors (e. g. degree of control). It is not possible to consider all these combinations with motion captures economically. Moreover, the presented influencing factors solely relate to the motion itself as well as the objects involved. Equally significant are the influences on motions that the executing person is responsible for. These influencing factors include the person's anthropometry, agility, handedness, constitution, physique, sex, skill, experience and routine for a given task. In order to record motion captures, the group of test persons must be selected in a balanced manner according to these criteria.

With regard to the influencing factors, a simulation of human motions within a workplace can derive the high number of data sets necessary to achieve a significant set of motion data for training the algorithm. This way it has to be ensured that the extremes for all basic motions and influencing factors, e. g. minimum and maximum reaching distances, are captured. These captures provide the framework for a simulation to interpolate further motion sets in between. Thus, the execution of motion captures is limited to a manageable scope.

An optical, marker-bound, infrared-based motion capture system provides high quality data and therefore serves as a benchmark for the acquisition of training data sets. The used motion capture system is Qualisys with twelve cameras that capture motions with a framerate of 125 Hz. With this configuration, it is capable to capture even very fast or subtle motions. Under real conditions in an industrial environment, such a setup is difficult to implement because of existing limitations of the workplace. Optical systems with several cameras need to be installed. Masking and dazzling surfaces must be avoided in order to generate high data quality. Therefore, after training the machine-learning algorithm, other motion capture systems based on inertial sensors (e.g. Xsens) or markerless optical motion capture system (e.g. Kinect) may be utilized as well. A main prerequisite is to provide the data in a given form with coordinates according to the given marker structure in Fig. 2. The data sets consist of different rows with X-, Y- and Z-coordinates of every marker point. Each row is linked with a timestamp and a label for the valid MTM code. To ensure the algorithm to work properly on different motion capture systems there needs to be a standardized nomination of the marker points, same units and the same coordinate system. In this case, a world coordinate system is used which is automatically specified by Qualisys software.

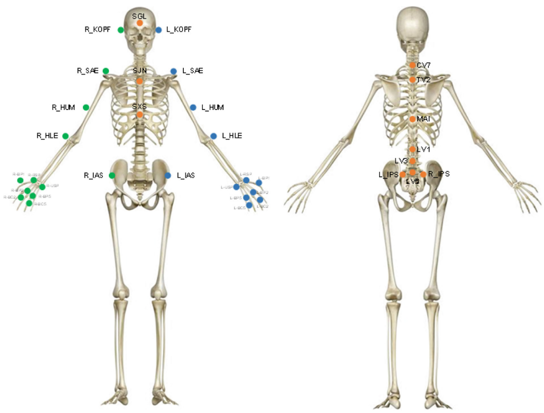


Fig. 2. Tracked marker points for basic motions of the upper limbs

Initial motion captures have been taken already and the data sets have been labeled according to MTM-1 with the help of video recordings of the assembly processes. The evaluation of the video data with regards to an automatic annotation of MTM-1 codes leads to the conclusion that the methodology of analysis differs from original MTM analyses. The analysis of video or motion data is restricted to the individual operative work method, whereas MTM is designed for the assessment of a generic work method. This results in a different view of motions and influencing factors. Therefore, the MTM methodology needs to be adapted which is shown in Sect. 3.

3 Adaptation of MTM Methodology

A major difference between an automated MTM analysis and the original manual analysis is the ability of a human to assess the context of the work task, whereas an algorithm can only draw conclusions from the available motion data. For an automatic generation of MTM-1 codes from motion data, those subjective assessments of the original classification procedure according to MTM must be substituted by objectively ascertainable criteria. The environment, especially for non-optical motion tracking systems, cannot be tracked. The generation of various MTM codes requires to asses subjectively ascertainable environmental factors, though. Therefore, for these MTM codes there is the necessity to derive characteristics not directly linked with the motion but the environment within the information of motion data and its components. These components consist of coordinates linked with a body segment and a time stamp. Thus, other motions beside the explicitly involved body segments (e. g. head movements to readjust the field of view to indicate the given degree of control for a motion) need to be identified and associated to characterize a certain motion and its linked influence factors. This means that motion patterns of the whole body need to be analyzed and connected in high detail. By observing multiple assembly tasks, the characteristics for each class need to be pointed out. The observations require experienced assembly workers as well as various assembly tasks in order to highlight universal features.

Considering that a motion capture cannot detect a given work method and hardly can differentiate in advance which body segment to be dominating for MTM analyses, parallel motions of several body segments need to be recorded. Thus, potentially there are various MTM codes for each time stamp to be derived. Because of the characteristics identification and the parallel MTM code derivation, the MTM data card has to be customized or extended for this purpose.

4 Classification Using Graph CNN

Based on the executed motion captures and the adaptation of the MTM methodology for the determination of the MTM codes, a classification of the data into the classes according to the MTM-1 subdivision is performed. Therefore, the mentioned adaption of the MTM methodology with a customized MTM data card is essential to hand over characteristics to be taken into account by the algorithm. Thus, a high quality of the forecasts can be expected. The classification operates in two stages. The first stage serves to identify the basic motions whereas the second stage is responsible for the influencing factors for each basic motion. The determination of MTM codes requires the detection of motions of the body and hands, possibly also additional technical aids (e.g. screwdriver). These different entities form a complex actuator network are representable by graph based data structures. Therefore, the objective is to adapt the generalization of CNNs on non-Euclidean domains (e.g. graphs) to the classification of graphs with respect to MTM codes. Recent works in this field were brought together under the umbrella term geometric deep learning [16] and came up with a set of methods that aim to expand the convolution operator in deep neural networks to handle irregularly structured input data. Fey et al. [17] introduced SplineCNN with continuous kernel functions based on B-splines, which make it possible to directly learn from the geometric structure of graphs. A new graph convolutional neuronal network (graph CNN)-supported concept for the tracking of MTM-1 sequences is to be developed, taking into account current results in geometric deep learning research.

Corresponding graph-based CNNs, for which initial preliminary work has already been done, represent a promising starting point and also guarantee fast inference about the use of GPU optimization techniques [18]. Recent research results represent a very promising starting point and also ensure quick information about the use of GPU optimization techniques [17]. During the analysis and development, it will be examined whether the tracking of common graph representations by means of recurrent graph CNNs exceeds the results of classical methods such as the extended Kalman Filter. Further it will be examined whether one of the two approaches or a combination of both is better suited to solve the concrete problem. Quality criteria among others are the usability in real time as well as the sufficient accuracy for the classification of MTM-1 codes. A challenge in the identification and classification of MTM codes from gestures lies in the detection of more complex motion patterns, which can be described by submotions. Therefore, a hierarchical classifier based on Graph-CNNs and coarsening layers for encoder-decoder architectures seems promising, from which coarse and fine as well as environmental and secondary features shall be extracted.

5 Conclusion and Outlook

This paper shows that there is a strong demand to support the planning processes of manual work systems to increase productivity and improve global competitiveness, especially for high-wage countries. In particular, time management provides the opportunity to generate knowledge about processes and therefore is an important tool to organize and structure internal processes. The high cost and time intensity to generate time analyses lead to incomplete or missing time management and time data. It is shown that there is the possibility to identify MTM sequences automatically and to derive complete MTM-1 analyses. The given methodology needs to be adapted so that the subjective assessment of the working environment and the processes is substituted by objective characteristics. The paper states that there are possibilities to take different body segments into account to define characteristics that lead to a clear allocation of coordinates of body segments over time to MTM motion sequences. Thus, the use of motion capture systems leads to the fact that the basis for automated determinations of MTM motion sequences is given. Recent studies have already shown that the state of the art in motion capture technologies is accurate enough to detect even subtle motions [19–21]. A promising approach to process the given motion data is provided by graph CNN. By means of geometric deep learning approaches, it is possible to identify different motion patterns independently of a strictly restricted and specified area of activities.

Further studies will complete the presented results by extending the motion capture of work tasks and therefore creating a wider basis of motion data for training the neuronal network. This leads to guarantee a high quality of forecasts, which are essential to use time data for purposes like workforce planning or defining fee structures.

References

1. Lotter, B., Wiendahl, H.-P. (eds.): Montage in der industriellen Produktion: Ein Handbuch für die Praxis. Springer, Heidelberg (2012)
2. Meier, K.-J.: Montage in Hochlohnländern. In: wt - Werkstattstechnik online, no. 4, pp. 232–235 (2005)
3. Manns, M., Otto, M., Mauer, M.: Measuring motion capture data quality for data driven human motion synthesis. In: 48th CIRP Conference on Manufacturing Systems (CMS), pp. 945–950 (2015)
4. Grauer, M., Metz, D., Karadgi, S., Schäfer, W.: Auf dem Weg zum Echtzeitunternehmen. Prod. Manag. **15**(1), 17–19 (2010)
5. Sackermann, R.: Eine wissensbasierte Methode zur Zeitermittlung in der Einzel- und Kleinserienfertigung. Zugl.: Technical University, Shaker, Aachen, Dortmund (2009)
6. Bauer, N., Stankiewicz, L., Jastrow, M., Horn, D., Teubner, J., Kersting, K., et al.: Industrial data science: developing a qualification concept for machine learning in industrial production. In: European Conference on Data Analysis (ECDA) (2018)
7. Bitkom: Potenziale und Einsatz von Big Data: Ergebnisse einer repräsentativen Befragung von Unternehmen in Deutschland, Bitkom, Berlin (2014)
8. Rexer, K., Gearan, P., Allen, H.: Data science survey. Rexer Analytics (2015)

9. Stankiewicz, L., Bauer, N., Horn, D., Jastrow, M., Eickelmann, M.: Maschinelles Lernen in der Industrie: Umfrage im Rahmen des Forschungsvorhabens. Industrial Data Science (InDaS), Institute of Production Systems (2018)
10. Lindl, M., Haferkorn, F., Wächtler, A., Kersebohm, P., Heil, M.: Stammdaten und Stammdatenmanagement: Grundlagen, Herausforderungen und Lösungsansätze für kleine und mittlere Unternehmen, Praxisbroschüre (2015)
11. REFA - Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensent wicklung: Industrial Engineering: Standardmethoden zur Produktivitätssteigerung und Prozessoptimierung, 2 Auflage, Hanser, München (2015)
12. REFA - Verband für Arbeitsstudien und Betriebsorganisation: Ausgewählte Methoden des Arbeitsstudiums, 2 Aufl., 6–10 Tsd, Hanser, München (1994)
13. Fricke, W.: Arbeits- und Zeitwirtschaft verstehen: Von der Zeitstudie bis zur Abtaktung, 1 Auflage. Books on Demand, Norderstedt (2016)
14. Deuse, J., Busch, F.: Zeitwirtschaft in der Montage. In: Lotter, B., Wiendahl, H.-P. (eds.) Montage in der industriellen Produktion: Ein Handbuch für die Praxis, 2nd edn, pp. 79–107. Springer, Heidelberg (2012)
15. Freitas, A.A., Lavington, S.H.: Mining Very Large Databases with Parallel Processing. Springer, Boston (2000)
16. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond euclidean data. IEEE Signal Process. Mag. **34**(4), 18–42 (2017)
17. Fey, M., Lenssen, J.E., Weichert, F., Müller, H.: SplineCNN: fast geometric deep learning with continuous b-spline kernels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 869–877 (2018)
18. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural Message Passing for Quantum Chemistry, 2017, http://arxiv.org/pdf/1704.01212v2
19. Heetmann, M., Portoff, T., Elei, T., Deuse, J., Stankiewicz, L., Kuhlenkötter, B., et al.: Individualisierte sozio-technische Arbeitsassistenz für die Produktion (INDIVA): Schlussbericht (2016)
20. Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst. **104**(2–3), 90–126 (2006)
21. Alfred, R., Lim, Y., Ibrahim, A.A.A., Anthony, P. (eds.): Computational Science and Technology: 5th ICCST 2018, Kota Kinabalu, Malaysia, 29-30 August 2018. Springer, Singapore (2019)

Human-Machine Interaction Applications

Preliminary Development of an Integrated Mobility, Lethality, and Survivability Soldier Performance Testing Platform

Stephanie A. T. Brown(✉) and K. Blake Mitchell

Combat Capabilities Development Command Soldier Center,
Natick, MA 01760, USA
stephanie.a.brown88.civ@mail.mil

Abstract. There is a desire for an integrated tool to measure the impact of military clothing and equipment on mobility, lethality and survivability. This study was a first step to develop such a test platform. Twenty Soldiers executed the test in three levels of encumbrance. Mobility was measured via obstacle completion timing. Lethality tasks included static and dynamic shooting engagements, with traditional marksmanship measures and cognitive decision-making. Quantification of body exposure and exposure to threat time comprised survivability measures. Preliminary results indicated that as encumbrance increased, mobility, lethality and survivability were altered. Obstacle completion times increased, marksmanship precision and vertical stability during the static elements improved, and shooting efficiency and threat elimination during the dynamic elements decreased. By expanding on this methodology, we create additional capabilities for the US Army and our international partners. Lessons learned from this study will allow for improvements to the test platform as it is developed.

Keywords: Clothing and Individual Equipment (CIE) · Human-systems integration · Human factors · Military · Combat obstacle course · Lethality · Mobility · Survivability · Marksmanship

1 Introduction

In 2017, the US Army announced their modernization priorities as a means of maintaining their military strength. Six specific areas were targeted for focus improvement and development, with the first five being specific technologies or end products. The sixth was "Soldier Lethality" or Soldier's ability to shoot, move, communicate, protect and sustain by improving human performance and decision making [1]. In an effort to support this priority area for those trying to make clothing and individual equipment (CIE) acquisition and development decisions, there is a desire for an integrated or holistic objective tool to measure Soldier performance, specifically mobility, lethality and survivability incorporating underlying measures of human factors, biomechanics and cognition.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 153–164, 2020.
https://doi.org/10.1007/978-3-030-20040-4_14

Defense research organizations from Australia, Canada, Singapore and the United States use the Load Effects Assessment Program (LEAP), a military mission obstacle course originally developed by the US Marine Corps, to assess the impact of CIE on dismounted warfighter performance and specifically, mobility. The standardized LEAP test platform includes a 10-station obstacle course, in addition to static simulated rifle firing, vertical jump and weight transfer activities. Previous studies investigating completion times for military task-oriented obstacles (i.e., jumping, running, crawling, climbing) and obstacles courses have been able to differentiate CIE designs and configurations [2–4]. In the last 5 years, a number of studies have been completed using similar versions of the LEAP obstacle course. That testing has shown that course performance is affected by differences in CIE [5–7]. The LEAP course has also shown to be sensitive to changes in gross weight [8] and in percent body weight carried [9].

Additionally, previous marksmanship performance research had shown sensitivity to CIE encumbrance when using live fire [10–13]. Multiple studies have shown that weapons simulator/training systems are predictive of live-fire qualification scores [14–18]. Moreover, simulated marksmanship has also shown to be sensitive to CIE encumbrance level comparisons [19, 20] and allows testers to assess compatibility and performance degradations in an easy, quick, low cost, and safe manner [21]. The simulated marksmanship tasks have used static and on-the-move shooting at single and multiple targets (varied heights and locations) to demonstrate these differences in performance when wearing CIE products [22]. They have also integrated basic cognitive decision-making elements (go-no-go tasks), but with only high-level mobility differences seen across CIE encumbrance levels [23].

By integrating these mobility and marksmanship methodologies, with additional developmental methodologies in the areas of cognitive decision-making and survivability metrics, the LEAP Mobility, Lethality and Survivability (LEAP-MLS) was conceived. This research provides an initial proof of concept of this developmental test platform. This study is a first step in developing a methodology that incorporates objective measures of performance and is sensitive to changes in Soldier-system equipment, thus helping to inform the Soldier performance trade space during product development and acquisition, using a single, standardized and controlled event.

2 Methods

2.1 Study Participants

Twenty active duty Soldier volunteers executed the course in three levels of CIE encumbrance, in a repeated measures design of experiments. All participants were active duty Army personnel, with a majority (all but three) having an infantry military occupational specialty. The test participants had a mean age of 24.85 ± 5.11 years. Their years in service ranged from less than a year to over thirteen, with a mean of 4.5 ± 3.48 years. Five had combat deployment experience. The test participants had a mean weight of 189.15 ± 27.27 lbs and mean height of 69.68 ± 1.96 inches.

2.2 Test Configurations

All twenty participants executed the course in three configuration that represented three levels of CIE encumbrance: unloaded, minimal-encumbrance, and maximal-encumbrance. The baseline or natural unloaded body included the participant, their duty uniform (e.g., Army Combat Uniform, duty boots, undergarments), a helmet (i.e., Advanced Combat Helmet) (worn for safety), and a surrogate (training aid) M4 weapon (with sensor and optics). The minimal-encumbrance configuration included all the components of the Unloaded in addition to body armor. The body armor was one of two fielded or soon to be fielded systems based on fit and availability, to include their identified baseline kit for the rifleman duty position. This included magazines, grenades, water, and a first aid kit, weighing approximately 23 kg. The maximal-encumbrance configuration included the components of the minimal configuration kit and additional items that are associated with the Grenadier duty position, such as additional 40 mm grenade, with a weight of approximately 25–30 kg. Although this kit is not much heavier than the minimal-encumbrance, it has the maximum amount of bulk around the torso of the body that a dismounted Soldier could experience when wearing body armor (approximately 17 cm as measured at waist circumference [24]).

2.3 Test Procedures

A repeated measures test design was used for this evaluation, where the participants ran through the scenario one time in each test configuration (three times total, randomized and counterbalanced). The participants were instructed to transverse the course as quickly as possible, while maintaining tactical discipline in their movements and accuracy during threat engagements. Practice of movements was provided during training, focusing on the difficult obstacles and incorporated shooting tasks. When testing commenced, the participant was dressed with the proper equipment and zeroed two weapon simulators in the prone position, one at a 5 m distance (target set to simulate 75-m) for use during the obstacle course portion of the test. The other was zeroed at 6.67-m (target set to simulate 200 m) for use during the static shooting task.

The scenario begins with the static shooting tasks. Participants were instructed to shoot three trials of five shots at a 100-m simulated distance (15 shots) and a 200-m simulated distance (15 shots), with a priority on accuracy over speed. This was completed for two firing position (i.e., standing unsupported and prone unsupported, order randomized). Upon completion of the static shooting, the participants commenced the integrated LEAP-MLS obstacle course (Fig. 1). The course contains standard obstacles for assessment of mobility, and fourteen dynamic (on-the-move) shooting engagements for assessment of lethality, including cognitive decision making (go/no go) and threat discrimination measures. At five locations, videos were captured to quantify bodily exposure and time of exposure to threats during engagements (survivability metrics). Finally, participants completed a post-static marksmanship session (firing position order same as pre-LEAP).

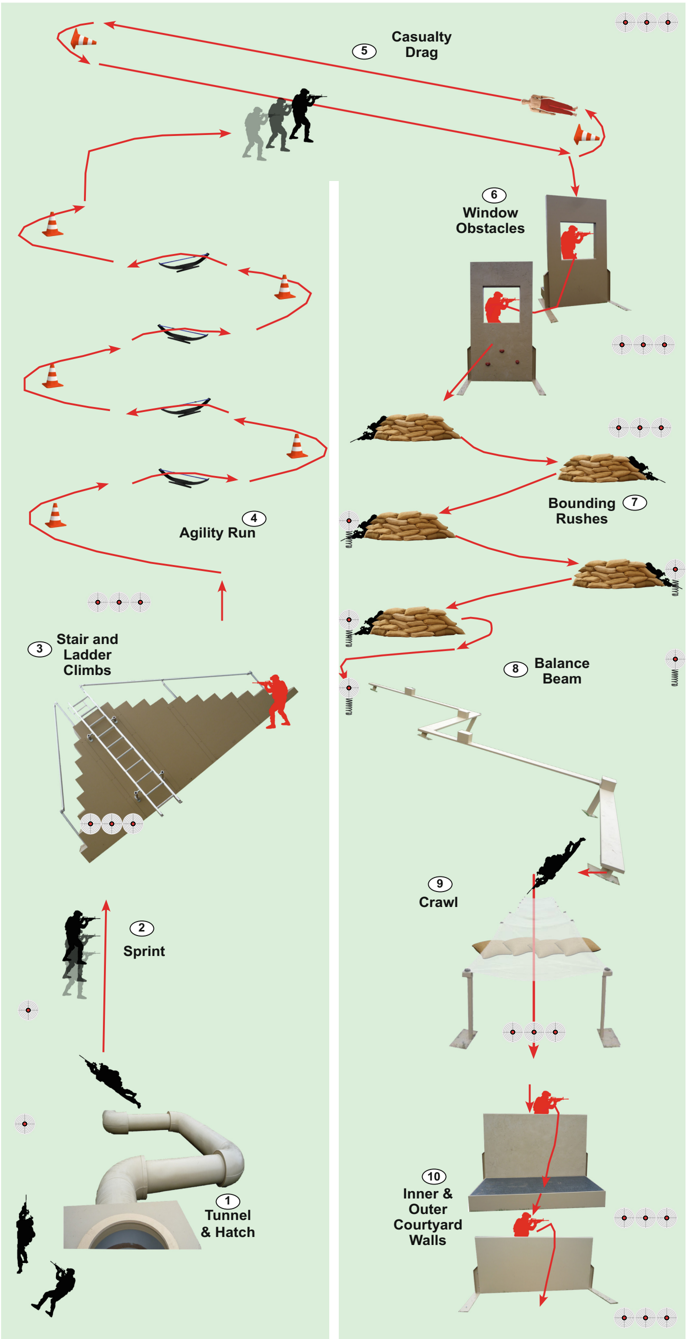


Fig. 1. Diagram of the LEAP-MLS test platform layout, starting in the lower left with the static shooting. The soldiers depicted in red indicate the survivability assessment points.

2.4 Test Apparatus

The LEAP-MLS course was set up according to the diagram above (Fig. 1). Each component and associated equipment will be discussed in further detail below.

Mobility Tasks. The standard LEAP obstacle course consists of a tunnel/hatch, sprint, ladder and stair climb, agility run, casualty drag, windows, bounding rush, balance beam, low/supine/high crawl, and walls. These obstacles replicate activities Soldiers would execute in a dismounted combat deployment.

Lethality Tasks. All of the lethality tasks utilized the FN Expert marksmanship simulator system in combination with a M4 training aide with an integrated CO2 recoil system (LaserShot, Stafford, TX). The FN Expert Weapon Simulator consists of an optical unit and software package loaded on a computer. The optical unit was mounted on the Picatinny rail system of the M4 training aid, and emits an eye-safe infrared (IR) light emitting diode (LED) beam. Paper ring targets, incorporating FN Expert specific diamond grade reflector rings, were used in the marksmanship activities throughout the scenario. The integrated lethality marksmanship tasks were set up throughout the course. The three major types of shooting incorporated into this event include shooting on the move (SOM), target discrimination tasks (TDT), and go/no go (GNG) reaction time tasks.

Static Tasks. The targets used in the static event were scaled to represent 100-meters (m) and 200-meters when placed at an actual distance of 6.67-m, and were placed at a height of 1-meters and 1.57-m on the same vertical plane.

SOM Tasks. The targets utilized for dynamic portion of the course were all scaled to represent 75-m at an actual distance of 5-m. Most targets were mounted at 1-m height. For those targets engaged while the participant was moving (sprint and agility run), a shooting line approximately 5- to 10-m from the target was provided. Targets requiring shooting in the prone posture (end of tunnel and crawl) were placed .5-m high.

TDT Tasks. A manual target tracker was used to help determine which TDT targets were engaged and number of actual shots executed during the post processing of the marksmanship data. The TDT utilized a camouflage patterned target set, with two camouflage patterns designated as enemy (or threat) and friendly (or non-threat), and a gradation of the two for partial threats and partial non-threats. The test had eight sets of targets, with orders selected in a randomized and partially counterbalanced manner for order exposure across the course. Each set had one 100% threat camouflage, one 100% non-threat camouflage, and one of the eight various mixed partial threat/non-threat patterns.

GNG Tasks. The GNG tasks included six pop-up targets placed at a height of 0.5-m and a 5-m distance from the bounding rush sand bags, but scaled to represent 75-m. The same camouflage pattered targets were utilized for this task, but in isolation. Two of the six targets used the two 100% threat or non-threat camouflage patterns, two displayed the 75% partial threat and non-threat patterns, and one showed the 51% partial threat, with randomized exposure order.

Survivability Tasks. Survivability was assessed after the stairs, at the windows, and at the walls. Each of these points had a TDT engagement. Video footage captured from the threat's perspective (i.e., camera attached to the target facing outward) was used to capture the participant's bodily exposure and exposure time prior to threat engagement.

2.5 Data Measures

Mobility Measures. Mobility was measured by timing data during completion of the various obstacles. Overall completion time (time to complete all eleven obstacles) was the primary metric. In addition, time to complete each individual obstacle was also utilized in order to determine which types of movements were compromised when adding various equipment items to the Soldier's load. Inertial measurement unit data was also collected and will be utilized for a more in depth analysis of movement quality.

Lethality Measures. The static shooting task provides information on lethality outcomes in a low stress environment where shooting accuracy is prioritized over speed (untimed), whereas the integrated dynamic marksmanship tasks focused on balancing speed of completion with shooting accuracy. Traditional lethality measures of accuracy (distance from shot to target center), precision (shot group dispersion), probability of hit (percentage of hits), and probability of lethal hit (percentage of hits at center of mass) were utilized across all shooting tasks. In addition, measures based on the weapon handling, stability and time spent aiming during the period prior to engagement were also utilized for analysis across all shooting tasks (i.e., aiming time, trigger control, vertical and horizontal barrel stability, and barrel rotation).

Additional marksmanship mobility measures were assessed only during the dynamic task, to include target acquisition time (move, detect and position), engagement time (aim, shoot, adjust, shoot, etc.), total trial time, and shooting efficiency (number of hits per second). Cognitive measures were assessed for the TDT and GNG tasks utilizing signal detection theory [25, 26]. Reaction time from signal presentation to threat engagement can also be calculated for the GNG task.

Survivability Measures. Survivability was assessed via threat elimination, bodily exposure amount, and exposure time. Threat elimination was calculated based on all target engagements during the dynamic portion of the session. Each target was designated a threat, partial threat, or non-threat based on the camouflage presented. For every target that was presented and engaged correctly, the threat was graded as correctly eliminated (i.e., 1). When a target was not engaged (i.e., no shots were fired), then threat was graded as not eliminated (i.e., 0). The probability of threat elimination was then calculated based on total correct sets of engagements divided by total target sets displayed. Additional refined measure of threat elimination entails identifying critical areas on the body/target that would incapacitate the threat upon engagement and only counting the threat as eliminated when the shots fired have hit those identified zones.

Bodily exposure *amount* was measured by the amount of body not concealed behind the obstacle structure. Video footage from the vantage point of the threat was analyzed to capture the amount of the Soldier's body exposed in pixels up until engagement. Additional refined measure of bodily exposure amount will include critical body parts exposed as opposed to any body part.

Bodily exposure *time* was measured in two manners. First, total exposure time from first exposure to threat through engagement (time of first shot) was captured. Second, the time from last exposure to engagement was captured. The reason for two measures was due to the various manners in which Soldiers can approach and engage threats while maintaining concealment. Some Soldiers look, assess, and engage immediately, while others look, conceal their bodies while assessing, and then re-expose themselves in order to engage the threat.

3 Results

The results reported here are preliminary findings and will only review the high-level measures across each area of the assessment. The statistical analysis was conducted to primarily investigate sensitivity to configuration (unloaded, min-encumbrance, max-encumbrance), while also investigating shooting posture (standing, prone) and fatigue (pre, post), as appropriate depending on the dependent variable. Within-subjects repeated measures analysis of variance (ANOVAs) were utilized to analyze each dependent variable. Tests of multiple comparisons were conducted using the Tukey Honestly Significant Differences (HSD). Confidence intervals were set at 95% (alpha = .05). Further detailed analysis is still required for a comprehensive understanding of this tool.

3.1 Mobility

Repeated measure ANOVA tests were conducted for each dependent mobility timing variable to compare performance across the three configurations. For this report, only the overall course completion time results will be presented. Overall course completion time was significantly different across configuration, $F(2, 26.82) = 23.7$, $p < .0001$. Post hoc analysis indicated that the unloaded configuration ($M = 322.4$ s, $SD = 39.9$) was significantly faster than both the min-encumbrance ($M = 409.3$ s, $SD = 68.9$) or max-encumbrance ($M = 406.9$ s, $SD = 60.2$) configuration as seen in Fig. 2. However, the two encumbrance conditions had comparable completion times.

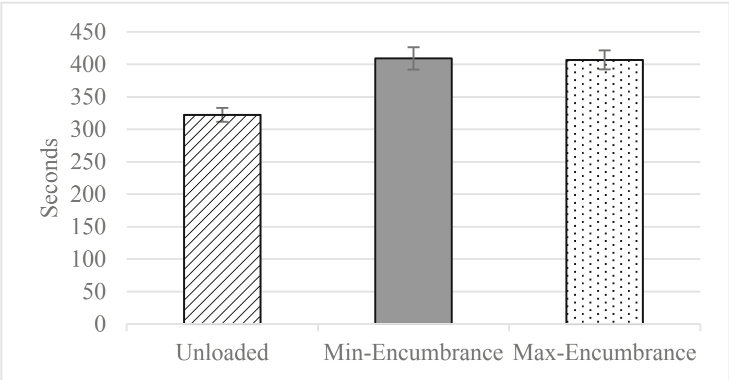


Fig. 2. Mobility as measured by overall course completion times across CIE (error bars represent Standard Error)

3.2 Lethality

Static Marksmanship. All static marksmanship measures were evaluated in this analysis, and only dependent variables that revealed a main effect of configuration will be reported here. A main effect of configuration was found for Precision, $F(2, 38.02) = 4.52$, $p = .017$, and Vertical Stability, $F(2, 38) = 4.24$, $p = .022$. Post hoc analysis indicates that the unloaded configuration (*M* = 161.5 mm, *SD* = 139.9) produced less precise shot groups as compared to the min-encumbrance (*M* = 146.8 mm, *SD* = 128.3) or max-encumbrance (*M* = 144.6 mm, *SD* = 128.0) configuration. In addition, the unloaded configuration (*M* = 174.5 mm, *SD* = 179.6) had significantly more vertical movement during aiming than the min-encumbrance (*M* = 152.7 mm, *SD* = 158.9) or max-encumbrance (*M* = 156.3 mm, *SD* = 159) configuration, as seen in Fig. 3.

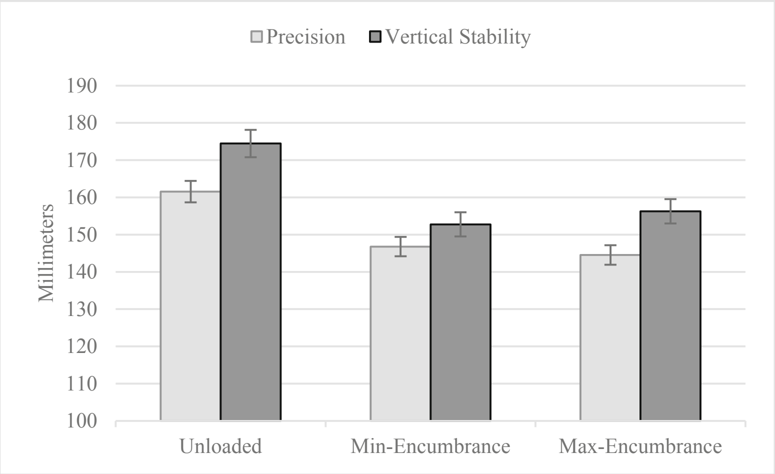


Fig. 3. Differences in Lethality as measured by mean Precision and Vertical Stability across CIE configuration (error bars represent Standard Error)

Integrated Dynamic Marksmanship. The Soldiers were instructed to traverse the course in a tactical manner, with a focus on speed while maintaining tactical movements and eliminating all threats along the route. With this in mind, shooting efficiency (hits/second) was focused on for this initial analysis. A main effect of configuration was revealed for shooting efficiency, $F(2, 38) = 4.24$, $p = .0218$. The unloaded configuration (M = .11 hits/s, SD = .03) was significantly more efficient (greater hits per second on the course) than the max-encumbrance (M = .086 hit/s, SD = .04) or min-encumbrance configuration (M = .08 hit/s, SD = .05) as seen in Fig. 4.

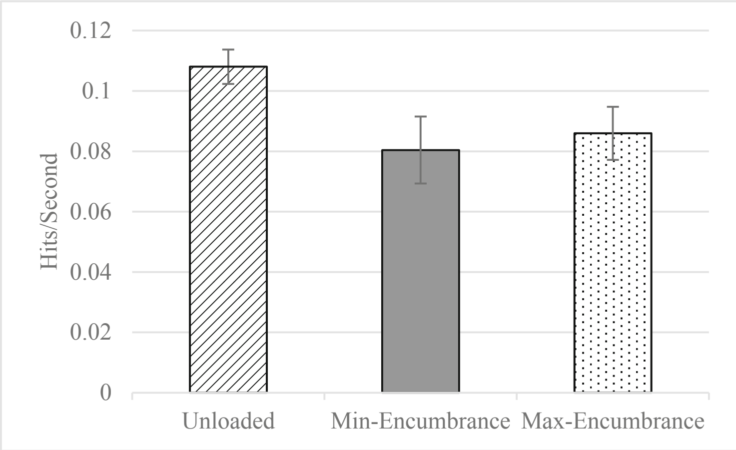


Fig. 4. Differences in Lethality as measured by mean Shooting Efficiency across configuration (error bars represent Standard Error).

3.3 Survivability

These initial results of survivability will only report the basic threat elimination measure based on correct threat engagement. The other measures of survivability (i.e., bodily exposure and exposure time) are still undergoing post-processing and analysis.

Analysis of threat elimination revealed a main effect of configuration, $F(2,34) = 3.98$, $p = .028$. When wearing the max-encumbrance configuration (M = .66, SD = .16), Soldiers had a significantly more difficult time identifying and eliminating all of the threats that were displayed as compared to the min-encumbrance condition (M = .77, SD = .11) as seen in Fig. 5.

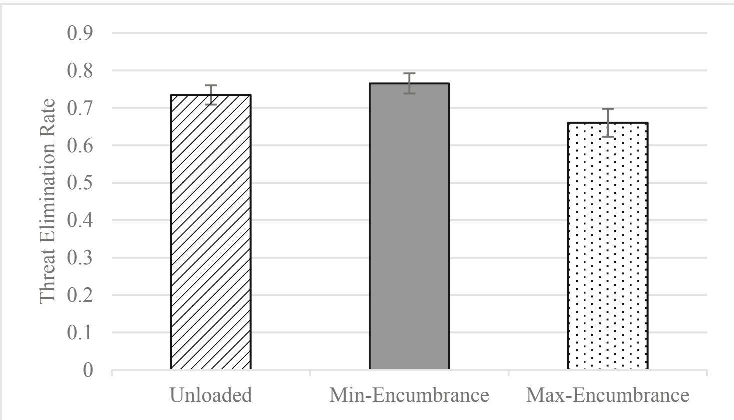


Fig. 5. Differences in Survivability as measured by threat elimination across configuration (error bars represent Standard Error)

4 Discussion

This research is the first proof of concept of an integrated mobility, lethality and survivability test platform for CIE related product evaluation. Although other methods have successfully looked at each individual element in isolation, this test platform provides an opportunity to consider the tradeoffs between these areas in a single assessment. Other research had begun to quantify behavior adaptations (taking cover) in the training environment when facing high or low fidelity simulated threats [27], but this is the first methodology for CIE testing that identifies and quantifies survivability (a critical aspect of lethality) through measurements of threat elimination and bodily exposure to threats when taking cover and engaging targets. Further research is warranted in order to add the appropriate protection level of worn equipment to the survivability scoring model within this test platform. The initial high-level results in the three categories of mobility, lethality and survivability have provided indications that this novel test platform is a good enhancement for integrated performance testing during CIE related product evaluations.

This initial research has provided great lessons learned for further development of the integrated test platform. Some high level benefits include the flexibility of the simulator system, the novel measures of survivability, and the integrated look at metrics for CIE assessment. Additionally, the platform is relatively inexpensive, is easy to set up and is mobile. The initial findings prove that the traditional LEAP system can be enhanced with supplemental operationally relevant tasks that focus on lethality without losing the system's original design intentions of being a comprehensive and sensitive mobility test for CIE product assessments. Further detailed analysis will provide additional information on the areas of enhancement and their value for future CIE product assessments. Future analysis efforts will also work towards a combined weighted index measure for each component area (i.e., mobility, lethality, and survivability) for ease of use and standardization of performance interpretation during CIE product tradeoff assessments.

Some limitations experienced during this initial research include some equipment-based issues. The minimal recoil and muzzle rise of the weapon surrogate as well as the lack of external environmental stressors (i.e., gunfire noise, wind adjustments, etc.) experienced during live-fire can influence the psychological physical performance during the shooting engagements. Also, some of our equipment (i.e., eye tracking glasses and cameras) had unanticipated technical difficulties during testing. Eye tracking could not be used with the close-combat optical sight on the weapon due to the focal point shift of the pupils to the sights rather than the target. Iron sights were attempted during the training period, but the shooting tasks were too difficult without the sights, and performance was falsely degraded due to the weapon configuration rather than the CIE configuration being evaluated. The difficulty level of the cognitive measures (camouflage-based threat images in the TDT and GNG tasks) and their location within the LEAP obstacles should be assessed further in order to determine the appropriate size and display required to capture the decision-making processes with minimal potential learning effects and acuity issues. Additionally, the eye tracking devices used was not field grade and overheated after one set of runs (approximately 20 participant completions).

Multiple military research and acquisition groups, within the US and internationally, use LEAP as a standardized assessment of mobility. By expanding on this test platform to create this alternative methodology, we create additional capabilities for the US Army and our international partners. Lessons learned from this study will allow for improvements to the test platform as it is developed.

Acknowledgments. This research was funded by Combat Capabilities Development Command Soldier Center (CCDC Soldier Center) research program 16-159 and Program Executive Office (PEO) Soldier Project Director Soldier System Integration (PD SS&I). The authors would like to acknowledge the data collection and support team that assisted in the execution of this study from Aberdeen Test Center and CCDC Soldier Center's Biomechanics and Human Factors Teams. Most importantly, the authors would like to acknowledge the Soldiers who participated in this study.

References

1. Department of the US Army. https://admin.govexec.com/media/untitled.pdf
2. Brainerd, S.T., Bruno, R.S.: Human factors evaluation of a prototype load-carrying system (Tech. Memo 15-85). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory (1985)
3. LaFiandra, M., Lynch, S., Frykman, P., Harman, E., Ramos, H., Mello, R.: A comparison of two commercial off the shelf backpacks to the Modular Lightweight Load Carrying Equipment (MOLLE) in biomechanics, metabolic cost and performance (Tech. Rep. T03-15). U.S. Army Research Institute of Environmental Medicine, Natick, MA (2003)
4. Brewster, F.W.: Technical memorandum: observation report for the load effects assessment program – army (LEAP-A) pilot evaluation, 1–12 December 2014. Maneuver Battle Lab, Fort Benning, GA (2014)
5. Tack, D., Kelly, A., Richter, M., Bray-Miner, J.: Preliminary Results of MC-LEAP Testing of U.S. Marine Combat Load Order Configurations. U.S Marine Corps Systems Command, Quantico, VA (2012)
6. Bray-Miners, J., Kelly, A.: CAN-LEAP summary of results—Fall 2012 experimentation series (Tech Memo). Defence Research and Development Canada, Toronto, CA (2013)
7. Dutton, B., Stryker, T.: Soldier load benchmark evaluation. Battle Lab Project Number 0338. Fort Benning, GA (2015)
8. Batty, J.M., Coyne, M.E., DeSimone, L.L., Mitchell, K.B., Bensel, C.K.: Evaluation of weight effects on a soldier physical readiness test course. In: Proceedings of the 2016 American Biomechanics Society. Raleigh, NC (2016)
9. Mitchell, K.B., Batty, J.M., Coyne, M.E., DeSimone, L.L., Choi, H.J., Gregorczyk, K.N., Bensel, C.K.: Impact of weight on military mission oriented obstacle course performance. Poster presentation at the 7th International Conference on Applied Human Factors and Ergonomics, Orlando, FL (2016)
10. Bensel, C.K.: Soldier performance and functionality: Impact of chemical protective clothing. Mil. Psychol. **9**(4), 287–302 (1997)
11. Carbone, P.D., Carlton, S.D., Stierli, M., Orr, R.M.: The impact of load carriage on the marksmanship of the tactical police officer: a pilot study. J. Aust. Strength Cond. **22**(2), 50–57 (2014)

12. Johnson, R.F., Kobrick, J.L.: Effects of wearing chemical protective clothing on rifle marksmanship and on sensory and psychomotor tasks. Mil. Psychol. **9**(4), 301–314 (1997)
13. Johnson, R.F., McMenemy, D.J., Dauphinee, D.T.: Rifle marksmanship with three types of combat clothing. In: Proceedings of the Human Factors Society 34th Annual Meeting, pp. 1529–1532 (1990)
14. Brown, S.A.T, McNamara, J.A., Choi, H.J., Mitchell, K.B.: Assessment of a marksmanship simulator as a tool for clothing and individual equipment evaluation. In: Proceedings of the Human Factors Society 60th Annual Meeting, pp. 1424–1428 (2016)
15. Crowley, J.C., Hallmark, B.W., Shanley, M.G., Sollinger, J.M.: Changing the Army's Weapon Training Strategies to Meet Operational Requirements More Efficiently and Effectively. Rand Arroyo Center, Santa Monica (2014)
16. Hagman, J.D., Smith, M.D.: Weapon zeroing with the laser marksmanship training system (LMTS). Research report no. 1744. U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA (1999)
17. Schendel, J.D., Heller, F.H., Finley, D.L., Hawley, J.K.: Use of weaponeer marksmanship trainer in predicting M16A1 rifle qualification performance. Hum. Factors **27**(3), 313–325 (1985)
18. Torre, J.P., Maxey, J.L., Piper, S.: Live fire and simulator marksmanship performance with the M16A1 rifle. Study 1. A validation of the artificial intelligence direct fire weapons research test bed, vol. 1. Main Report. Advanced Technology Inc, Orlando, FL (1987)
19. McNamara, J., Choi, H.J., Brown, S.A.T., Mitchell, K.: Evaluating the effects of clothing and individual equipment on marksmanship performance using a novel five target methodology. In: Proceedings of the Human Factors Society 60th Annual Meeting, pp. 2043–2047 (2016)
20. Choi, H.J., Mitchell, K.B., Garlie, T., McNamara, J., Hennessy, E., Carson, J.: Effects of body armor fit on marksmanship performance. In: Advances in Physical Ergonomics and Human Factors, pp. 341–354 (2016)
21. Brown, S.A.T, McNamara, J., Mitchell, K.B.: Dynamic marksmanship: a novel methodology to evaluate the effects of clothing and individual equipment on mission performance. In: Proceedings of the Human Factors Society 61th Annual Meeting, pp. 2020–2024 (2017)
22. Hasselquist, L., Eddy, M.D., Mitchell, K.B., Brown, S.A., McNamara, J., Hancock, C.L., Caruso, C.: Assessing the impact of clothing and individual equipment (CIE) on soldier physical, biomechanical, and cognitive performance part 1: test methodology (No. NATICK/TR-18/004). Army Natick Soldier Research Development and Engineering Center, Natick, MA (2018)
23. Brown, S.T., McNamara, J.A., Mitchell, K.B., Eddy, M.D.: Development of a building clearing methodology for the assessment of soldier. J. Sci. Med. Sport **20**, S67 (2017)
24. Mitchell, K.B., Choi, H.J., Garlie, T.N.: Anthropometry and range of motion of the encumbered soldier (No. NATICK/TR-17/010). Army Natick Soldier Research Development and Engineering Center, Natick, MA (2017)
25. Green, D.M., Swets, J.A.: Signal Detection Theory and Psychophysics. Wiley, New York (1996)
26. Stanislaw, H., Todorov, N.: Calculation of signal detection theory measures. Behav. Res. Methods, instrum. Comput. **31**(1), 137–149 (1999)
27. Taverniers, J., De Boeck, P.: Force-on-Force handgun practice: an intra-individual exploration of stress effects, biomarker regulation, and behavioral changes. Hum. Factors **56**(2), 403–413 (2014)

Detecting Human Factors that Induce Errors in Movement Patterns for the Development of a Web-Based Telerehabilitation Platform

Wilmer Esparza Y[1,2](✉), Arián Aladro-Gonzalvo[3], Jorge-Luis Pérez-Medina[1], Karina Beatríz Jimenes[1], and Jonathan Baldeon[1]

[1] Intelligent and Interactive Systems Lab, Universidad de Las Américas, Quito 170125, Ecuador
{wilmer.esparza, jorge.perez.medina, karina.jimenes, jonathan.baldeon}@udla.edu.ec

[2] School of Physical Therapy, Universidad de Las Américas, Quito 170125, Ecuador

[3] Pontifical University Catholic of Ecuador, Av. 12 de Octubre 1076 y Roca, Quito, Ecuador
araladro@puce.edu.ec

Abstract. Clinical patient evaluations are important in determining motor performance. Commonly, these evaluations are performed by physiotherapists and include the pain assessment, range of motion, muscle strength and functionality. Recent studies have shown to be effective in quantifying the patients and the healthy subjects' motor performance. Several telerehabilitation assessments may be conducted with or without devices in order to aid in diagnosis, to determine treatment goals and to evaluate treatment outcomes. On the other hand, the clinical assessment is thought to be highly variable affecting the reliability. This work presents several reflections about the differences between the assessments performed by physiotherapists and the evaluation carried out by a web-based telerehabilitation platform. We analyze the human factors that could have affected the assessments performed by physiotherapists.

Keywords: Web platform · Telerehabilitation · Human errors detection · Movement assessment

1 Introduction

The validation of movements is a clinical challenge considering that each person moves differently and presents some variability in the patterns of movement and the evaluation of an individual. This complexity is maintained when the process is made by computer systems such as telerehabilitation platforms. This work presents a brief review of human movement assessments, movement assessment devices and human errors assessment.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 165–176, 2020.
https://doi.org/10.1007/978-3-030-20040-4_15

This manuscript allowed us to determine some of the relevant aspects that could lead to the error in the evaluation of the movement observed in our laboratory. Among these, we indicate those related to the physiotherapist, the subject/patient and the environment. Finally, we present some examples of reliability and validity measurements.

2 Human Movement Assessments

Human movement offers a way of interaction with the world and is essential to consider the ability of a person to participate in society. The interest in knowing its characteristics goes back to the history of humanity. The evaluation of the human movement has had several stages related to the development of technology and the understanding of the movement as a whole. Thus, we went from monoplane isolated evaluations of a joint to integrated evaluations of multiple body segments in different planes within a kinetic chain [1–3]. It also went from simple observations (clinical and imaging) to automated computer analysis systems of movement (Optotrak, Vicon, …). Currently, the movement is considered one of the pillars in the maintenance of health and the quality of life of people [4, 5], so that their optimal recovery is very important.

From a clinical point of view, the physiotherapist is the professional responsible for evaluating the movement [6]. Among the parameters that are generally evaluated are: pain, range of motion (ROM), strength, coordination, posture and functionality. Several studies have documented the values considered normal in healthy people and in patients. However, these values are the result of the averages obtained in a group of subjects and may not necessarily be the objective to be achieved [7]. This is because each person moves differently and a variability degree in the movement patterns is "normal" at the same time [8, 9]. Due to the multi-factorial nature and intrinsic variability of human movement, the ideal or "normal" concept of moving probably does not exist. Another aspect to take into account is that the evaluations does not consider the multifactorial pathological conditions or the functional recovery capacity of each patient, focusing only on specific parts related to explaining the clinical phenomenon and/or the function [10]. Finally, certain evaluations are carried out under strict (analytical) protocols, while others are carried out in a natural (spontaneous) (functional) manner.

In current clinical practice, movement is evaluated in a personalized manner and a specific individualized plan is developed by the physiotherapist to maximize functional capacity. Lately, the development of technology let us use devices that allow the evaluation of movement in real time and even to develop telerehabilitation "TR" programs that evaluate and control treatment sessions [11, 12]. The use of new technologies should not marginalize this individualized evaluation, so the development of personalized real-time motion assessment platforms is a challenge. Starting from the difficulty that the assessment itself (multi-factorial nature, the variability of the movement, analytical or functional), it must also look for low-cost movement capture systems to be accessible to a large part of the population.

For the development of evaluation programs through TR platforms it is necessary to establish the limits that will be considered "normal". Two possibilities have been reported, one considerate as reference the correct movement executed by physiotherapist

[11], and the other the average of a series of tests performed freely by healthy people [13]. These two studies were conducted using a low-cost camera (Kinect) to evaluate hip movement. In the first study the ROM, the unipodal balance and a squat movement executed with the lower left and right limb were evaluated. The second study evaluated ROM, coordination and compensation. The results of the two studies showed a good consistency (90%) comparing, either the therapist's performance or the subjects with the movement capture system.

Although there is still no consensus about the parameters that should be evaluated through TR platforms and regardless of whether the evaluation is analytical or functional, it is important to quantify the coordination and the compensation of the movement. The compensatory presence of movements is associated with movement deficits. The coordination evaluation is important since the coordinative strategies of the movement can be modified at any time, by execution of the movement itself (fatigue, error in the execution) or by pain [14–18]. These changes in the coordinative strategies of the movement may persist, creating erroneous patterns of movement and subsequently interfering with the ability to participate in other physical activities [19, 20]. Patients often use compensatory strategies to facilitate movements [21]. Compensatory movements limit the functional improvements of the therapy of the affected limb and can cause pain. This is the reason why a TR system should capture, limit and quantify the compensations. Compensatory movements assimilated by patients create altered movement patterns that could limit the proper use of the injured limb in the activities of daily life [22, 23].

The development of a series of platforms and devices for clinical evaluation is promising, as long as it is monitored by a specialist for the patients' treatment (physiotherapist). Eliminates the presence of the patient in a center or a laboratory, presents the results immediately and is a reliable and valid clinical tool.

3 Movement Assessment Devices

Advances in computer software, biosensors and communication technology allow health professionals to manage several applications based on TR. Applications are feasible to assess and to treat musculoskeletal disorders. They have started to be employed as a complement to the clinical practice in physiotherapy [24, 25]. Physiotherapist services are integrated to the regular clinical practice. Therefore, they are based on valid and reliable objective evaluations [26]. Validity and reliability are essential clinical components to achieve an objective evaluation, to monitor progress and to guide an appropriate implementation of the intervention [27]. Nevertheless, poor validity and reliability have negative influences on the outcomes that could be misleading to determine the effectiveness of the results of the TR based intervention [28].

A prerequisite in the current clinical practice is testing the new technology by comparing it with the conventional face-to-face evaluation. If the difference in the comparison is small and clinically insignificant, the TR can be used as an alternative modality [29]. Thus, several studies have been carried out to evaluate the validity and reliability of TR platforms quantifying parameters such as ROM; the posture; the strength, the resistance and control; the pain, the swelling and scarring; the march and the balance. The results of these studies show that physiotherapy evaluations through the TR are technically feasible. They can measure parameters such as pain, swelling, ROM, muscle strength, balance and gait. In addition, the functional results show good concurrent general validity [30].

A recent experience in our laboratory confirms the validity and reliability of a TR platform to assess ROM, coordination and compensation. The capture of the movement was made by a Kinect camera [11, 13, 31]. Seven subjects with four different rehabilitation movements (hip abduction, hip and knee flexion, hip extension, and frontward-sideway-backward step sequence) were consider for our study. They followed a script unknown to the therapists and different from one participant to another. The only similarity between the scripts was the fact that they were composed of six accurate executions, two movements with an incorrect ROM, and three movements with compensation errors. The movements were executed eleven times each (one of them was used as reference). The trials were assessed in terms of ROM and compensatory movements. The results showed that the algorithm matched at 88% with the assessment made by the physiotherapists. Also, there was some difference of congruence between the human assessment and the machine assessment according to the type of exercise. The accuracy rate can reach 92.5% for the frontward-sideway-backward sequence and hip abduction [13]. However, this study showed greater variability in the error detection by physiotherapists evaluating hip movements at the same time that a movement analysis system. In the next section, the factors that probably have exercised the influence in the variability related with the error detection in the movement evaluation (see the next section).

Based on the recognition that each person moves differently and that there is certain variability in movement patterns, the evaluation of an individual is a clinical challenge. Traditionally, there is great variability in the initial evaluation of a patient, but registering a great variability in the final evaluation has been seen as undesirable [32]. Generally, high variability in movement and evaluation is observed for the performance of functional tasks such as activities of daily living, occupational and sports skills. In contrast, a low variability is observed in the execution of particular isolated movements. In either case, too much variability can be indicative of an alteration in the movement [8].

The ranges of variability in measurements are different between healthy subjects and patients. However, these ranges have a certain consistency among the populations. Too much variability between consecutive repetitions within the same environment in healthy people may indicate a less optimal interaction between the therapist, the patient and the environment [9, 33]. In the case of patients, the ranges of variability may be affected by pain, the deficits of movement and fatigue [34–37]. The variability of human movement could be subject to the conditions and the environment in which the movement is made. Thus, a spontaneous movement will have more variability than a controlled movement. In the same way, an unstable environment would increase the variability. In addition, in the clinical evaluation process, the patient is often asked to perform some repetitions of the same movement. No repetition will be exactly the same as the previous one, so it is advisable not to make clinical interpretations based on a single repetition of a specific task [6–38].

In this context, the validity and reliability of the clinical evaluation requires the expertise of the therapist, the ability to understand the movement to be performed by the patient and the control of the environment. Regardless the methodology used during the assessment, the clinical interpretation according to the individual in a multidimensional context remains essential [39]. Funded on this clinical interpretation, a certain preferred movement patterns can then be considered as biomechanically more or less advantageous for a particular person, at a given time.

4 Assessment Human Errors

According to a conceptual model of human performance, the patterns or sequences of movements are governed by generalized motor programs (GMPs). Each of them has at least an unchanging temporal organization of certain motor parameters (e.g. distance, speed, …) [40]. However, a GMP can vary depending on the subject's selection of these motor parameters [41]. From here, that observing a movement pattern is extremely complex due to the interaction of several factors related to the person himself (limitations of the organization), the task that is being carried out (task limitations), and the environment or context in which it is done (environment, restrictions). In turn, the movement coordination strategies and the resulting movement patterns are influenced by multiple dynamic and interactive factors. For these reasons, it is known that a large number of measurement errors is derived from the evaluation of movement patterns.

It must be understood that the change in a variable, in this case typical of human movement, can be attributed to three sources of variance: the true variance, the error variance and the foreign variance. These sources show the degree of internal validity of a measurement. In the case of experimental studies, internal validity refers to the researcher's certainty that the changes observed in the dependent variable are due to the independent variant effect [42].

The true variance refers to the change that occurs in a variable as a consequence of an intervention [42]. Let us suppose that a physiotherapist is interested in knowing if an intervention based on TR is more effective than a conventional intervention, to expand the ROM in patients with a hip arthroplasty. Patients are randomly assigned to two groups: A1 (intervention TR) and A2 (conventional intervention). It is defined at random which group receives A1 and which A2. If after treatment, the average of A1 is greater than the average of A2, this difference in intervention A1 can be considered as the true or systematic variance of the ROM.

On the other hand, the error variance refers to the inconsistent variation in the values of the measurement of the study variable. According to Hyllegard et al. [43], the error variance is a variation that occurs naturally and inexplicably and that is always present in the measurements. As mentioned above, when analyzing free-running daily life movements, the variation in the subject is very high because there is a high interaction with the environment that shows very changing situations. This error variance can be reduced if this situation is minimized and the subject's actions are restricted in time and space.

Even when the researcher thinks that all possible variables have been controlled in the best way, there will always be an error variance. In practical terms, the variance of error is considered as the variation between the scores of the subjects. For Kerlinger and Lee [42], this variance can be reduced by using samples that are more homogeneous with each other, and using measurement instruments that have high reliability.

Finally, the foreign variance is related to other sources of external variation that have not been included in the study and which should usually be known and controlled. These sources of foreign variance can be given in the first place with regard to the errors of the instrument. For example, a system of TR that adds 20° to each measurement of the ROM is a source of strange variance. This instrument of measurement improperly calibrated, would systematically add 20° to the measurements of the movement that is carried out. So, the readings of this measurement would always be 20° higher than they really are. If the sources of strange variance are present in all measurement methods, then the results will possibly overestimate or consistently underestimate the true values of the variable of interest [43].

Second, sources of foreign variance can also be found with respect to human errors. The measurements will not be reliable, in terms of consistency and precision, when the assessments are performed by an evaluator who has problems of interaction with digital technology, who has poor clinical experience or knowledge about the motion kinematic characteristics. In this way, the assessment errors are specific to the human or the interaction with the system. Regarding the human errors that can occur in a conventional clinical assessment, a preliminary study conducted by our laboratory found great error detection variability by physiotherapists when a real-time assessment is carried out, executing movements spontaneously and without script. On the one hand, this could be explained through the three aforementioned components, therapist, subject/patient and environment. At the subject/patient level, they did not receive enough information about the movement to be executed. As for the physiotherapists, they did not have a well-structured formal training on what they had to evaluate. Finally, neither the physiotherapists nor the participants knew the environment in which the experiment would be developed (laboratory with cameras recording the movement in real-time). Although the motor system performs actions highly stereotyped, there is evidence available showing that information lack affects decision making [44]. This explains the possibility of making mistakes during the execution or assessment of a motor action. Unawareness of the environment in which a particular action takes place also alters motor performance [45, 46]. In addition, it is well known that human behavior is susceptible to being modified when we are in front of a camera. In some cases, it helps to retrieve corrective information in memory, but in others it is presumed to increase the psychophysiological response of the subject during an assessment [47].

On the other hand, the TR platform assessment was carried out in real-time. Real-time assessment implies that all users (health professionals and patients) participate can exchange information instantly. This simultaneous interaction could induce errors by distraction during the assessment. Assuming there is a doubt in the score, the movement could be repeated, but as we pointed out earlier, no movement is similar to another.

There are other aspects related to the execution of the movement itself, such as if the movement was spontaneous or controlled and how many body segments participate in it. With respect to spontaneous movement, it presents more variability than controlled movements. In the case of the body segments evaluated, our study analyzed the trunk compensatory movements and extremities, implying a double task for the therapist. In this case the evaluator had to observe the movement of the involved limb and the compensation that this movement could induce. A dual-task assessment could induce errors, especially when the physiotherapist has no expertise in the assessment.

After revealed several aspects that could induce errors in this study, clearly the control of all these aspects would reduce errors in the measurements. Video recordings are necessaries in new studies in order to verify the TR system validity. These recordings would allow to evaluate the same movement several times, this would probably help to determine what was the cause of the error and also allow the physiotherapists to perform a movement re-evaluation. In the case of assessment errors as a result the interaction system-human, some psychological mechanisms that lead to errors have been documented. These are the elements related to the cognitive function (attention, memory, knowledge of the process), the characteristics of the person and their interaction [48]. Other means include human error modes, the interaction means between humans and automation. In addition, it has been shown that usability problems, related to the design problems of the interface, are closely associated with the appearance of specific types of errors [49]. Therefore, knowing these mechanisms is fundamental for the analysis of human errors and human reliability in the human-system interface. Hence, there is a growing interest in the effects of computer-based digital technology on the reliability of the evaluator/operator. In either case, reliability is a critical point of any assessment. Reliability refers to the consistency or repeatability of measurements [50], and is clearly an essential knowledge to help the physiotherapist decide if there is a particular capacity. given time.

5 Reliability and Validity

The measurement reliability can also be verified by the intra-evaluator and inter-evaluator variability. In these cases, the objective is to know how the evaluations of the same evaluator and between the evaluators, are consistent. For to improve the measurement consistency, it is necessary to design protocols with standardized measurement procedures previously selected by evaluators. In addition, an inferential statistical check is required to know the error level measurement. The statistic used with a certain frequency is the intraclass correlation coefficient (ICC). The ICC indicates the measurement stability and provides information about the distortion (error) level that has the measurements of the same variable. The correlation coefficient scores vary from 0 to 1, where 0 is an indication that the results of the comparative measurements are unstable and relatively unpredictable (inconsistent), while a value of 1 or near may suggest that the measurements are stable and relatively predictable. Reliability is generally population-specific, so caution is recommended when making comparisons between studies.

The validity is another essential psychometric property in evaluations. This validity is related to the measurement accuracy [42]. A frequent way to report the assessment accuracy is to evaluate the errors that occur in a measurement. This is obtained by comparing the errors of the instrument/method with another standard or reference instrument. The expected result is that the difference in the errors between the instruments is the minimum possible or is within tolerable theoretical or clinical ranges.

A method that provides useful and easily interpretable information about measurement errors is the determination of mean absolute percentage errors (MAPE). This average error is obtained comparing an instrument/method with another instrument recognized by the scientific community as the gold standard for its robust psychometric properties. At the same time the MAPEs can be contrasted with the reference values according to the standard of the Consumer Technology Association (CTA) of the American National Standards Institute (ANSI) (Physical Activity Monitoring for Fitness Wearables: Step Counting, 2016) [51]. For example, in the case of portable devices monitoring physical activity, Xie et al. [52] have reported that these devices have a fairly high measurement accuracy with respect to the steps number and the distance traveled, with a MAPE of approximately 0.10%, which is within the limits of the value suggested by the CTA (<10%).

In the case of TR technology, these limits of tolerable errors are not yet available. Although validity and reliability are generally perceived as desirable, there is no exact definition of the validity level and consistency required to achieve clinical acceptability. However, statistically significant levels of reliability may not translate into clinically acceptable levels [50].

According to our clinical experience, during the assessments that involve the interaction between two or more evaluators, some errors may arise in the data collection. Between frequently errors are the reading and word, auditory, registry, as well as errors in identifying the parameters to evaluate. The reading and word error can be committed when the evaluator measures and transmits the values incorrectly to the second evaluator. On the other hand, the auditory error is committed by the latter even when registering the values by writing the numbers clearly. The registration error is attributed to the evaluator who measures when he has not correctly performed the evaluation. Finally, the errors in identifying the parameters to be evaluated are commented when a priori certain points have not been defined, such as the performance of an analytical or global evaluation, focusing on aspects of the movement control (timing of movement), the compensations, fatigue, coordination, fluency, speed, etc.

The ICC analysis can also be useful to evaluate the validity of an assessment tool. The ICC can be accompanied by Bland-Altman graphical and technical error statistics methods. Let's see an example of how a validity analysis could be proposed with these integrated methods. Imagine that you want to test the validity of an observational assessment method for shoulder movements (MOASM). The validity of the MOASM was determined by intraclass correlation coefficient (ICC), Bland-Altman graphical and technical error statistics methods. The ICC was calculated between the ROM using the Kinect® method and that obtained using the MOASM. These ICC values were later complemented with Bland-Altman graphical statistics. A one-sample t-test was used to determinate the grade of agreement between the difference generated by 2 methods of measurement and the acceptable referential value the 0.5 grade. The limits of agreement were sufficiently narrow to allow for the bedside method (i.e. MOASM) to replace the reference method when encompass 95% of all measured values. A multiple regression analysis using the Enter method was performed to determine the moderating effect of the mean of bedside and reference methods over the difference generated by these 2 methods. The technical error of the reading made with both methods was also calculated as follow, square root of $[1/2n \times \sum(\text{Xi} - \text{yi})2]$, where Xi and yi are the

ROM values obtained with each method in each subject, «i» represent each subject (i = 1, 2, 3… n) and «n» is the total number of subjects. Finally, to assess the differences between movements (non-spontaneously "scripted" and spontaneously "unscripted" of technical error, paired-sample t student test was used. The level of statistical significance for data was defined at $\alpha = 0.05$, and the 95% confidence intervals (CI 95%) are presented [29, 53].

6 Conclusions

The review of the literature and our experience in the development of a series the exercises adapted for a TR platform shows that errors assessments can occur at the level of the therapist, the subject/patient and the environment. Other factors related to the execution of the movement itself (type: spontaneous or controlled, number of body segments involved) could also induce errors in the assessment. Among the aspects that could be controlled to limit measurement errors are: (1) informing subjects about the task they should do; (2) providing a well-structured training to physiotherapists on the evaluation to be performed; (3) recognizing and becoming familiar with the environment where the evaluation will take place; (4) limiting the simultaneous interactions between the subjects; (5) avoiding the evaluation of a dual-task (especially to the inexperienced physiotherapists); and (6) making video recordings to evaluate spontaneous movements in real time.

Acknowledgements. The work is supported by Universidad de Las Américas (UDLA) under Grant No.: SIS.JPM.18.01.

References

1. Dingenen, B., Malfait, B., Vanrenterghem, J., et al.: Can two-dimensional measured peak sagittal plane excursions during drop vertical jumps help identify three-dimensional measured joint moments? Knee **22**(2), 73–79 (2015)
2. Dingenen, B., Malfait, B., Vanrenterghem, J., et al.: The reliability and validity of the measurement of lateral trunk motion in two-dimensional video analysis during unipodal functional screening tests in elite female athletes. Phys. Ther. Sport **15**(2), 117–123 (2014)
3. Kibler, W.B., Wilkes, T., Sciascia, A.: Mechanics and pathomechanics in the overhead athlete. Clin. Sports Med. **32**(4), 637–651 (2013)
4. Bauman, A., Merom, D., Bull, F.C., et al.: Updating the evidence for physical activity: summative reviews of the epidemiological evidence, prevalence, and interventions to promote "active aging". Gerontologist **56**(Suppl 2), S268–S280 (2016)
5. Rhodes, R.E., Janssen, I., Bredin, S.S.D., et al.: Physical activity: health impact, prevalence, correlates and interventions. Psychol. Health **32**(8), 942–975 (2017)
6. American Physical Therapy Association. Vision statement for the physical therapy profession and guiding principles to achieve the vision (2013). http://www.apta.org/Vision/. Accessed 25 Jan 2019
7. Greenhalgh, T., Howick, J., Maskrey, N.: Evidence Based Medicine Renaissance G. Evidence based medicine: a movement in crisis? BMJ **348**, g3725 (2014)

8. Hamill, J., Palmer, C., Van Emmerik, R.E.: Coordinative variability and overuse injury. Sports Med. Arthrosc. Rehabil. Ther. Technol. **4**(1), 45 (2012)
9. Kiely, J.: The robust running ape: unraveling the deep underpinnings of coordinated human running proficiency. Front. Psychol. **8**, 892 (2017)
10. Bittencourt, N.F., Meeuwisse, W.H., Mendonca, L.D., et al.: Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition-narrative review and new concept. Br. J. Sports Med. **50**(21), 1309–1314 (2016)
11. Anton, D., Nelson, M., Russell, T., et al.: Validation of a Kinect-based telerehabilitation system with total hip replacement patients. J. Telemed. Telecare **22**(3), 192–197 (2016)
12. Rybarczyk, Y., Deters, J.K., Cointe, C., Esparza, D.: Smart web-based platform to support physical rehabilitation. Sens. (Basel, Switzerland) **18**(5), 1344 (2018)
13. Rybarczyk, Y., Luis Pérez Medina, J., Leconte, L., Jimenes, K., González, M., Esparza, D.: Implementation and assessment of an intelligent motor tele-rehabilitation platform. Electronics **8**(1), 58 (2019)
14. Santamaria, L.J., Webster, K.E.: The effect of fatigue on lower-limb biomechanics during single-limb landings: a systematic review. J. Orthop. Sports Phys. Ther. **40**(8), 464–473 (2010)
15. Savage, R.J., Lay, B.S., Wills, J.A., et al.: Prolonged running increases knee moments in sidestepping and cutting manoeuvres in sport. J. Sci. Med. Sport **21**(5), 508–512 (2018)
16. Gokeler, A., Benjaminse, A., van Eck, C.F., et al.: Return of normal gait as an outcome measurement in acl reconstructed patients. A systematic review. Int. J. Sports Phys. Ther. **8**(4), 441–451 (2013)
17. Oiestad, B.E., Holm, I., Engebretsen, L., et al.: The association between radiographic knee osteoarthritis and knee symptoms, function and quality of life 10-15 years after anterior cruciate ligament reconstruction. Br. J. Sports Med. **45**(7), 583–588 (2011)
18. Risberg, M.A., Oiestad, B.E., Gunderson, R.: Changes in knee osteoarthritis, symptoms, and function after anterior cruciate ligament reconstruction: a 20-year prospective follow-up study. Am. J. Sports Med. **44**(5), 1215–1224 (2016)
19. Myer, G.D., Faigenbaum, A.D., Foss, K.B., et al.: Injury initiates unfavourable weight gain and obesity markers in youth. Br. J. Sports Med. **48**(20), 1477–1481 (2014)
20. Rathleff, M.S., Rathleff, C.R., Olesen, J.L., et al.: Is knee pain during adolescence a self-limiting condition? Prognosis of patellofemoral pain and other types of knee pain. Am. J. Sports Med. **44**(5), 1165–1171 (2016)
21. Rainville, J.B., Hartigan, C., Wright, A.: The effect of compensation involvement on the reporting of pain and disability by patients referred for rehabilitation of chronic low back pain. Spine **22**, 2016–2024 (1997)
22. Da Gama, A., Chaves, T., Figueiredo, L., et al.: Guidance and movement correction based on therapeutic movements for motor rehabilitation support systems. In: Proceedings of the 14th Symposium on Virtual and Augmented Reality, Rio de Janeiro, Brazil (2012)
23. Brokaw, E.B., Lum, P.S., Cooper, R.A., et al.: Using the Kinect to limit abnormal kinematics and compensation strategies during therapy with end effector robots. In: Proceedings of the 2013 IEEE International Conference on Rehabilitation Robotics, Seattle, WA (2013)
24. Eriksson, L., Lindström, B., Gard, G., et al.: Physiotherapy at a distance: a controlled study of rehabilitation at home after a shoulder joint operation. J. Telemed. Telecare **15**, 215–220 (2009)
25. Russell, T.G.: Physical rehabilitation using telemedicine. J. Telemed. Telecare **13**, 217–220 (2007)
26. Russell, T.: Goniometry via the internet. Aust. J. Physiother. **53**, 136 (2007)
27. Durfee, W.K., Savard, L., Weinstein, S.: Technical feasibility of teleassessments for rehabilitation. IEEE Trans. Neural Syst. Rehabil. Eng. **15**, 23–29 (2007)

28. Kimberlin, C.L., Winterstein, A.G.: Validity and reliability of measurement instruments used in research. Am. J. Health Syst. Pharm. **65**, 2276–2284 (2008)
29. Bland, J.M., Altman, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **1**, 307–310 (1986)
30. Mani, S., Sharma, S., Omar, B., et al.: Validity and reliability of internet-based physiotherapy assessment for musculoskeletal disorders: a systematic review. J. Telemed. Telecare **23**(3), 379–391 (2017)
31. Clark, R.A., Mentiplay, B.F., Hough, E., et al.: Three-dimensional cameras and skeleton pose tracking for physical function assessment: a review of uses, validity, current developments and Kinect alternatives. Gait Posture **68**, 193–200 (2019)
32. Ericsson, K.A., Lehmann, A.C.: Expert and exceptional performance: evidence of maximal adaptation to task constraints. Ann. Rev. Psychol. **47**, 273–305 (1996)
33. Harbourne, R.T., Stergiou, N.: Movement variability and the use of nonlinear tools: principles to guide physical therapist practice. Phys. Ther. **89**(3), 267–282 (2009)
34. Chiu, S.L., Chou, L.S.: Effect of walking speed on inter-joint coordination differs between young and elderly adults. J. Biomech. **45**(2), 275–280 (2012)
35. Grooms, D., Appelbaum, G., Onate, J.: Neuroplasticity following anterior cruciate ligament injury: a framework for visual-motor training approaches in rehabilitation. J. Orthop. Sports Phys. Ther. **45**(5), 381–393 (2015)
36. Hodges, P.W., Smeets, R.J.: Interaction between pain, movement, and physical activity: short-term benefits, long-term consequences, and targets for treatment. Clin. J. Pain **31**(2), 97–107 (2015)
37. Wikstrom, E.A., Hubbard-Turner, T., McKeon, P.O.: Understanding and treating lateral ankle sprains and their consequences: a constraints-based approach. Sports Med. **43**(6), 385–393 (2013)
38. Corkery, M.B., O'Rourke, B., Viola, S., et al.: An exploratory examination of the association between altered lumbar motor control, joint mobility and low back pain in athletes. Asian J. Sports Med. **5**(4), e24283 (2014)
39. Dingenen, B., Blandford, L., Comerford, M., et al.: The assessment of movement health in clinical practice: a multidimensional perspective. Phys. Ther. Sport **32**, 282–292 (2018)
40. Schmidt, R.A., Lee, T.D.: Motor Learning and Performance: From Principles to Application, 5th edn. Human Kinetics, Champaing (2014)
41. Schmidt, R.A., Lee, T.D.: Motor Control and Learning: A Behavioral Emphasis, 5th edn. Human Kinetics, Portland (2011)
42. Kerlinger, F.N., Lee, H.B.: Foundations of Behavioral Research, 4th edn. Harcourt College Publishers, San Diego (2000)
43. Hyllegard, R., Mood, D.P., Morrow Jr., J.R.: Interpreting Research in Sports and Exercise Science. Mosby-Year Book Inc, St. Louis (1996)
44. Gaffin-Cahn, E., Hudson, T.E., Landy, M.S.: Did I do that? Detecting a perturbation to visual feedback in a reaching task. J. Vis. **19**(1), 5 (2019)
45. Dunovan, K., Verstynen, T.: Errors in action timing and inhibition facilitate learning by tuning distinct mechanisms in the underlying decision process. J. Neurosci. **39**(12), 2251–2264 (2019)
46. Kuling, I.A., de Brouwer, A.J., Smeets, J.B., et al.: Correcting for natural visuo-proprioceptive matching errors based on reward as opposed to error feedback does not lead to higher retention. In: Experimental Brain Research, vol. 1 (2018)
47. Kirschbaum, C., Pirke, K.M., Hellhammer, D.H.: The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology **28**, 76–81 (1993)

48. Li, P.C., Zhang, L., Dai, L.C., et al.: A new organization-oriented technique of human error analysis in digital NPPs: model and classification framework. Ann. Nucl. Energy **120**, 48–61 (2018)
49. Kushniruk, A.W., Triola, M.M., Borycki, E.M.: Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int. J. Med. Inf. **74**(7–8), 519–526 (2005)
50. Bruton, A., Conway, J.H., Holgate, S.T.: Reliability: what is it, and how is it measured? Physiotherapy **86**(2), 94–99 (2000)
51. Physical activity monitoring for fitness wearables: step counting (2016). https://webstore.ansi.org/Standards/ANSI/CTA20562016ANSI. Accessed 25 Jan 2019
52. Xie, J., Wen, D., Liang, L., et al.: Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR mHealth uHealth **6**(4), e94 (2018)
53. Aladro-Gonzalvo, A.R., Esparza-Yánez, D., Tricás-Moreno, J.M., et al.: Validation of a force platform clinical for the assessment of vertical jump height. J. Hum. Sport Exerc. **12**(2), 367–379 (2017)

RehabVisual: An Application to Stimulate Visuomotor Skills in Preterm Babies with Developmental Alterations

Catarina Santos[1], Ana Ferreia[2,3], Claúdia Quaresma[1,4](✉), and Carla Quintão[1,4]

[1] Departamento de Física, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[2] Serviço Medicina Física e Reabilitação, Hospital de D. Estefânia, Centro Hospitalar de Lisboa Central, 1169-045 Lisbon, Portugal
[3] Departamento de Saúde, Escola Superior de Saúde, Instituto Politécnico de Beja, 7800-111 Beja, Portugal
[4] LIBPhys - UNL, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
q.claudia@fct.unl.pt

Abstract. RehabVisual allows for the monitoring of the entire clinical interaction with a patient, integrating evaluation and intervention, through the reaction to a set of visual stimulation videos. These functionalities allow for a standardization of the evaluations; the monitorization of the baby's abilities over time, as well as the analysis of their evolution; and the creation of a more systematic, protocolized visual stimulation interaction. The use of the platform was tested, through the application of a variety of usability tests, which were performed by occupational therapists at D. Estefânia Hospital and Rehabilitation Medicine Center of Alcoitão. The efficiency of the tool, in a population of six infants younger than 18 months of age, displaying clear developmental alterations due to preterm birth, was also evaluated during occupational therapy. Throughout the therapy program, the occupational therapist selected, for each baby, specific stimuli that were in line with that baby's behaviour in previous sessions. After using the platform for a while, all feedback from therapists was highly positive. In addition, the possibility to use it throughout the whole paediatrics service was thoroughly appreciated.

Keywords: Rehabilitation · Visuomotor skills · Pediatrics

1 Introduction

A good view is essential for the overall development of the baby, so it is necessary to have attention from birth especially to the risk groups as is the case of premature infants [1]. Complications associated with premature infants are related to the immaturity of organ systems for life in an extrauterine environment, and the lower the gestational age the greater the risk of the baby developing complications [2]. In these cases, it is necessary to perform more ophthalmologic examinations early in the first year of life [1].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 177–184, 2020.
https://doi.org/10.1007/978-3-030-20040-4_16

Visual assessment comprises two strands the evaluation of visual function that allows evaluating the functioning of the eyes and the visual system and the evaluation of the functional vision that evaluates the way the baby uses the vision in performing tasks, the level of focus of objects, visual attention and visuomotor coordination [3, 4].

Early visual stimulation is very important for infants with developmental alterations since it allows to improve the fixation, follow-up and oculomotor coordination capacity. This recovery is justified by the fact that during the first year of life some maturation processes still occur. This period is called cerebral plasticity and allows the adaptation and modification of the brain according to the stimuli present [3].

The RehabVisual platform intends to make the stimulation process of visuomotor competences more Individualized and specific to the needs of each baby. Additionally, it offers an evaluation component that allows evaluating the initial state of visual reaction as well as its evolution during the treatment [5].

The objective of this article is to describe the process of applying the Rehabvisual platform to a population of six premature infants aged up to 18 months with developmental alterations. It is also intended to describe the usability tests applied to the occupational therapists, which allowed verifying whether the solution developed is in accordance with their needs.

2 Materials and Methods

2.1 Platform

The Rehabvisual platform intends to monitor the entire clinical process of the baby, both in the evaluation components as well as in the intervention programs that allow to complement the therapy sessions.

It was designed taking into consideration five types of users: Administrator, physicians, technicians, occupational therapists and caregivers, who are the people accompanying babies in appointments and therapy sessions [5].

After registering the baby on the platform, it is possible to add ophthalmologic and behavioral assessments by physical/technicians ophthalmologists.

At the start of occupational therapies, the platform includes an initial assessment that allows analyzing the baby's initial condition and it's suitable for monitoring the development of the baby throughout the therapy sessions. Rehabvisual encompasses an intervention program consisting on a wide range of videos with different levels of complexity, making it adapted to the needs of each baby. At the end of each treatment session, the therapist makes an evaluation that allows the recording of the visualized videos and where the follow up and fixation capacities of the baby are analyzed.

The parameters included in the various assessments were defined by the professionals in order to allow register of the evaluated parameters.

The assessments made to the functional vision and Intervention Programme are based on the behavioral analysis of the baby when a stimulus is presented, these changes can be translated on indicators such as looking, smiling or balancing. In these assessments, an evaluation scale was adopted, consisting of the following parameters:

1. Never - (0%)
2. Rarely - (25%)
3. Occasionally - (50%)
4. Often - (75%)
5. Always - (100%).

This scale allows you to make the Faster and standardized evaluation process for the entire service.

In all assessments there is also the possibility of inserting comments that allow the user to add relevant information to the baby's condition.

All the assessments can be consulted by all health professionals registered on the platform, which allows greater communication between ophthalmologists and occupational therapists. This process allows you to make a more specific treatment of the baby and a better monitorization of its development.

2.2 Protocol of Visual Stimulation

The Visual stimulation videos are intended for the intervention program implemented in the therapy sessions and the functional evaluation performed by occupational therapists. The functional evaluation aims to analyze the baby in a general way, in all existing categories of the intervention program, identifying which are the areas of the intervention program in which the baby needs greater follow-up.

The objectives of the stimulation program developed are: to define the largest possible number of choices, in order to make it more complete and thus achieving a more specific and individualized treatment for each baby; contain videos with different levels of complexity in order to encourage the cognitive development of the baby [1]. The developed protocols are based on the following parameters [1, 5]:

- Figures: Taking into account the established age range, the geometric shapes defined were quadrangular, triangular and circular and a mixed pattern was also, which is the junction of the three defined geometric figures.
- Dimension: The dimensions chosen were those referring to Teller cards, which have a size of 12×12 cm. It was thus considered an initial size of 12 cm in the figures, and as the the video progresses the dimension of the figures decreases to 8 cm and finally to 4 cm.
- Color: In this age range the considered colors are red (RGB: 196/0/0), yellow (RGB: 235/230/0), blue (RGB: 5/57/205), green (RGB: 0/153/0) and black and white.
- Movement: Static and moving figures were considered. Regarding the static figures, videos of simple figures (in which only a single picture is presented) and videos with standard were defined. Regarding the movement, the horizontal, vertical, diagonal and circular movement was considered.
- Contrast: In order to increase the contrast in the figure, black and white backgrounds were considered.

Regarding the intervention program, the videos are divided into three categories:

Simple videos: They consist of a set of static figures positioned in the center of the screen (Fig. 1). There are 20 hypotheses of stimulation videos in this category, since there are 4 different types of figures (circles, triangles, squares and mixed pattern) and 5 different colors (red, yellow, blue, green and black and white).

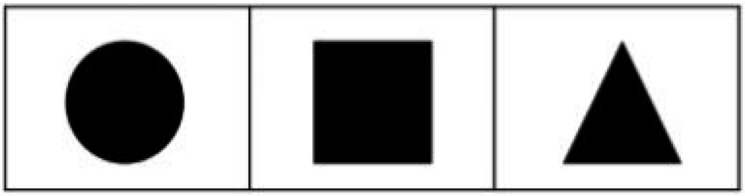


Fig. 1. Example of simple stimulation videos.

In order to increase the degree of complexity of stimuli, stimuli with patterns were included. These standards are based on the three geometrical figures considered and also the mixed pattern.

Three levels of complexity were defined: medium, high and very high (Fig. 2). With increasing complexity there is an increase in the number of figures and a decrease in the spacing between them.

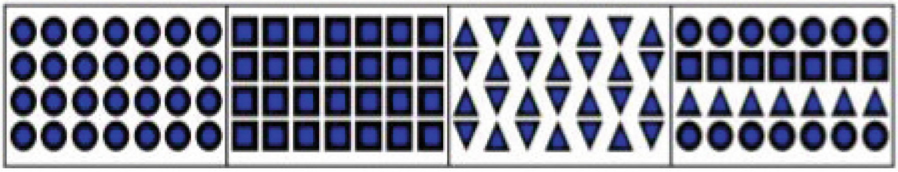


Fig. 2. Example of pattern stimulation videos.

There are 60 hypotheses of stimulation videos in this category, and each level of complexity includes 20 videos that correspond to different types of figures (circles, triangles, squares and mixed pattern) and 5 different colors (red, yellow, blue, green and black and white).

Stimulation through movement figures, allows to develop the relative perception of objects and their spatial disposition. It was for this reason that a set of stimuli that add movement (horizontal, vertical, diagonal e circular) to the simple geometric figures, circular, quadrangular, triangular and mixed pattern (Fig. 3) were defined. Three distinct levels of difficulty were defined, through the change of movement speed of the figures: slow, medium and fast speed.

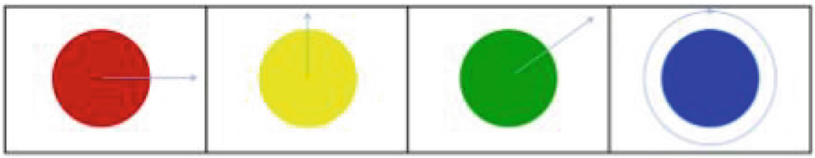


Fig. 3. Example of movement stimulation videos.

2.3 Platform Validation

Usability Tests. The System Usability test was applied to verify if the application developed was in compliance with the needs defined by the users. This questionnaire consists of 10 questions, which are punctuated by participants through a scale of 1 to 5 according to their level of agreement:

1. I think I'd like to use this system often.
2. The system is unnecessarily complex.
3. I think the system is easy to use.
4. I think I need help from a technician to be able to use the system.
5. I thought the various functions of this system were well integrated.
6. I thought there were a lot of inconsistencies in the system.
7. I imagine most people can learn to use the system quickly.
8. I found the system very complicated to use.
9. I felt very confident using the system.
10. I need to learn many things before using the system.

The choice of this questionnaire was made due to the fact that is simple, versatile and contains a small number of questions, this last factor is extremely important since the participants did not have much time available to respond to the usability tests.

Participants. The test was performed by nine occupational therapists, four from the physical Medicine and Rehabilitation Service of Hospital D. Estefânia and the remaining five from the Center for Medicine and Rehabilitation of Alcoitão. Only two of the HDE service participants had already been in contact with the platform during therapy sessions with the baby test group with development changes and less than 18 months, having already suggested some suggestions for changes during the development phase of the platform.

Test Protocol. The test session is of an individual character and is initiated with a brief introduction of the platform, then it is requested the user to perform the tasks indicated in the protocol. The protocol includes the various tasks performed by occupational therapists since the patient's registration, until the insertion of functional assessments and intervention sessions, where it is also requested to consult and edit previously submitted assessments.

2.4 Use of the Platform in the Therapy Sessions

The platform was also included in the occupational therapy sessions of six babies up to 18 months of age with developmental changes resulting from preterm birth, performed by two occupational therapists included in the validation of the platform (Table 1). Informed consent was requested to all parents, and after their authorization, the use of the platform was initiated in the occupational therapy sessions.

Table 1. Description of participants sample.

Participants	Age (months)	Prematurity	Session number
1	8/6	31 weeks	8
2	14/10	26 weeks + 1 day	9
3	9/6	29 weeks + 6 days	5
4	6/2	24 weeks + 4 days	5
5	7/4	26 weeks + 4 days	4
6	7/4	26 weeks + 4 days	4

The platform was complemented by the therapy sessions already attended at the physical Medicine and Rehabilitation Service of Hospital D. Estefânia, and the number of sessions in which it was applied, which depended on the number of weekly therapy sessions of each baby.

During the visualization of the video the child was seated on the therapist's lap.

3 Results

3.1 Usability Test

Knowing that in the SUS questionnaire half of the statements are of a positive nature and the remaining negative ones, it is necessary to convert them into a single result, as well: in the questions associated with the odd number (positive questions) the answer is subtracted by the value of 1 and in the questions associated with the even number (negative questions) the answer is subtracted by the value of 5. The values are then summed, and this result is multiplied by the value 2.5 in order to generate the SUS score, which is between 0 and 100 [6]. The average SUS score is 68, which corresponds to the 50th percentile. This score is affected by the complexity of the system and the tasks that the user has to perform [7].

A color map was used in Table 2 in order to more easily identify the positive responses so in the case of the assertions of positive character the values 4 and 5 are marked green, the 3 is identified in yellow and 1 and 2 in orange. In the case of negative responses, the representation is assigned inversely, as such the green color is assigned to the scores of 1 and 2, the yellow color is assigned to 3 and the orange color to scores of 4 and 5 [6].

From the analysis of the Table 2 it is concluded that the assessment made by users is generally positive, and only two users (participant 1 and 3) rated the platform negatively (SUS value less than 68). It should be noted that only participant 3 shows a very negative result (SUS value less than 51). Of all the participants with positive results, two classify the platform as good (SUS value higher than 68 and less than 80.3) and the remainder as excellent (SUS value exceeding 80.3) [8]. The two participants with a SUS rating of less than 68, during the test both stated that they do not often use the computer which may have compromised the answers given in the questionnaire.

Table 2. SUS results.

Participant/affirmation	1	2	3	4	5	6	7	8	9	10	SUS results
Participant 1	4	4	3	4	5	1	3	3	3	3	57,5
Participant 2	4	1	5	1	5	2	5	1	4	1	92,5
Participant 3	2	3	3	3	3	2	2	3	2	4	42,5
Participant 4	5	1	5	2	5	2	4	1	4	2	87,5
Participant 5	4	2	4	2	4	2	5	1	4	4	75
Participant 6	4	2	4	2	5	1	4	1	4	1	85
Participant 7	3	1	5	2	4	3	4	1	3	4	70
Participant 8	3	2	5	2	4	1	5	1	4	2	82,5
Participant 9	5	1	5	4	5	1	4	1	4	1	87,5
SUS mean score											75,6
Confidence interval (95%)											12,63

3.2 Therapy Sessions

Through the observation of the cases studied, it is concluded that infants in the age group considered, can see, on average, three videos per session before they begin showing signs of fatigue.

Throughout the sessions, all infants progressed at the level of difficulty of the videos. The choice of stimuli to be used in each session was performed taking into account the analysis of the occupational therapist of the baby's behavior in the previous session and during the visualized stimuli throughout the session. This process allows to have a personalized treatment for the needs of each baby, that is, the choice of stimuli is made taking into account the evaluation of the acquisition of competencies by babies and not according to their diagnosis.

It was also verified that since within each category of videos (simple, with pattern and movement) there is a vast set of choices that allows within the same level of difficulty, the therapist can define a treatment with different videos which decreases signs of habituation that can lead to lack of interest.

It is also concluded that the intervention protocol allows the creation of a common methodology for the entire service and assessments of the sessions allow, in a fast and easy way, to have access to the protocol used and the behavior of infants in each of the sessions (which facilitates the analysis of the development of the baby), and allows a treatment adapted to the patients responses, being the premise in common to all sessions to include videos with less complexity followed by more demanding videos in order to stimulate the baby and not to show signs of habituation.

Regarding the effectiveness of the therapy, with a reduced number of therapy sessions it is not possible to draw conclusions from the development of the babies. However, it was observed that the ability to fix never presents a lower assessment of the ability to pursue and that throughout the sessions most infants were able to experience all the categories of videos. A test with a larger number of sessions would also allow infants to experience all subcategories included in the category of patterns (medium, high and very high complexity) and in the category of movements (horizontal, vertical, diagonal), which not possible to do with all participants.

4 Conclusion

After the application of RehabVisual, it is concluded that it allows to create an individualized and specific treatment to the needs of each baby, due to the wide range of stimuli present for use in therapies and also by the fact that in a single platform the health professional has access to the clinical record and records of all assessments carried out by the child, which also allows monitoring over time. It also allows for greater cooperation between the ophthalmology service and the physical medicine and rehabilitation service, because the assessments introduced can be viewed by the other authorized users.

It provides standardized evaluations that optimizes fill time, improves interpretation of results and provides a common method for the entire service.

At the level of therapy sessions, the tool is quite versatile since new stimuli can be easily introduced more adapted to the needs of the infants and their ages, making the treatment more individualized and specific to the population in question.

References

1. Direcção de Serviços de Cuidados de Saúde/Comissão de Coordenação do Programa Nacional Para a Saúde Da Visão: Boas Práticas em Oftalmologia - Elementos Clínicos de Avaliação e Referenciação. Direcção-Geral da Saúde, Lisboa, Portugal (2008)
2. Institute of Medicine Committee on Understanding Premature Birth and Assuring Healthy Outcomes: Mortality and acute complications in preterm infants. In: Behrman, R.E., Butler, A.S. (eds.) Preterm Birth: Causes, Consequences, and Prevention, cap. 10. National Academies Press, Washington (DC), USA. https://www.ncbi.nlm.nih.gov/books/NBK11385/
3. Alimovic, S.: The assessment and rehabilitation of vision in infants. Paediatr. Croat. Suppl. **56**, 218–226 (2012)
4. Rossi, L.D.F., Vasconcelos, G.C., Saliba, G.R., Brandão, A.O., Amorim, R.H.C.: Avaliação da visão funcional em crianças: revisão da literatura. Oftalmologia **37**, 1–9 (2013)
5. Machado, R., Ferreira, A., Quintão C., Quaresma C.: Rehabvisual: development of an application to stimulate visuomotor skills. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018), BIODEVICES, Funchal, Madeira, Portugal, 173–178 (2018). https://doi.org/10.5220/0006597001730178
6. McLellan, S., Muddimer, A., Peres, S.C.: The effect of experience on system usability scale ratings. J. Usabil. Stud. **7**(2), 56–67 (2012). http://uxpajournal.org/the-effect-ofexperienceon-system-usability-scale-ratings/.f
7. Klug, B.: An overview of the system usability scale in library website and system usability testing. Weave: J. Libr. User Exp. **1**(6) (2017). https://quod.lib.umich.edu/w/weave/12535642.0001.602?view=text;rgn=main
8. UX research, Measuring and interpreting system usability scale (sus). https://uiuxtrend.com/measuring-system-usabilityscalesus/

Prescriptive and Descriptive Similarity of Team Contexts

Daichi Mitsuhashi[1](✉), Taro Kanno[1], Satoru Inoue[2], Daisuke Karikawa[3], Kohei Nonose[4], and Kazuo Furuta[1]

[1] The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
skyblue.m.daichi@gmail.com,
kanno@sys.t.u-tokyo.ac.jp,
furuta@rerc.t.u-tokyo.ac.jp

[2] Electronic Navigation Research Institute, 7-42-23 Jindaiji-higashi-machi, Chofu, Tokyo 182-0012, Japan
s-inoue@mpat.go.jp

[3] Tohoku University, 6-6-11 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi 980-8572, Japan
daisuke.karikawa@tohoku.ac.jp

[4] Central Research Institute of Electric Power Industry, 2-6-1 Nagasaka, Yokosuka-shi, Kanagawa 240-0196, Japan
nonose@criepi.denken.or.jp

Abstract. This paper presents the T^2P model, which is a model of team contexts modified and extended from the previous model. T^2P stands for Team, Task, and Place—the major elements of team contexts. This model describes and characterizes a team context with the attributes of these elements and the relationship among them. This paper also presents the criteria to assess the similarity between different team contexts and shows a few experimental results to support the validity of these criteria and the descriptive power of the model.

Keywords: Meso-cognition · Team context · Similarity measurement · Experimental design

1 Introduction

Human behavior has a great impact on the resilience of socio-technical systems; however, it has both positive and negative contributions. This makes it necessary to better understand practitioners' naturalistic cognitive behavior to make socio-technical systems safer and more resilient [1]. There are, however, many constraints and difficulties in conducting macro-cognitive studies [2]. For example, it is difficult for both academia and industry to find collaborators for macro-cognitive studies. Even if we do manage to find a collaborator, it is still difficult to obtain sufficient data through field observations, or to repeatedly design and conduct experiments considering various real-world conditions, and recruit several participants from practitioners. Another problem in macro-cognitive studies is the reliability of the knowledge obtained from a specific

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 185–193, 2020.
https://doi.org/10.1007/978-3-030-20040-4_17

domain and context, because human behavior is heavily context dependent. Therefore, it is difficult to generalize and extend the knowledge to other domains.

The meso-cognitive approach is a way to cope with these problems [2]. In this approach, we aim to design and use laboratory experiments that are similar to the actual working environments in essence, but are easy to conduct with non-practitioners. This approach is expected to help obtain useful data and knowledge from laboratory experiments relatively easily and intends to compensate for the limited and less reliable data obtained from macro-cognitive studies. In order to design similar experimental contexts, it is necessary to measure and evaluate the degree of similarity between different contexts. Such a measurement is also expected to provide a basis and criteria for the generalization of the knowledge obtained in different domains and contexts. For that purpose, a framework to formalize team context was proposed in our previous study [2, 3]. However, this framework is still conceptual and has not provided detailed criteria and measurement of context similarities.

This paper presents the T^2P model that is modified and extended from the previous framework and provides the criteria to assess the similarity between different team contexts. It also shows a few experimental results to support the validity of these criteria and the descriptive power of the model.

In the next section, we explain the T^2P model and propose the “prescriptive” similarity defined by the context features from the T^2P. In Sect. 3, we explain three experimental contexts with different prescriptive similarities. We conducted laboratory experiments with these contexts and gathered various cognitive and behavioral data. The analysis of the data and the comparison with the prescriptive similarities are explained in Sect. 4 followed by conclusions in Sect. 5.

2 T^2P Model and Prescriptive Similarity

The previous context-modeling framework consisted of seven major elements: human, task, device, resource, expertise, authority, and place [2, 3]. The structure of context is described by these elements and the relationships among them. In this paper, we introduce several extensions and elaborations to the previous model. Then, various criteria to assess context similarity are defined, which refers to prescriptive similarity in this paper.

2.1 T^2P Model

The seven major elements in the previous model were listed based on an organizational structure model. Thus, it is not necessarily well suited for describing a team context. We, therefore, redesigned and simplified the context elements by deleting “device,” “resource,” “expertise,” and “authority” from the major elements for the following reasons. Device and resource are both used for tasks and thus should be placed as task elements. Expertise and authority should be defined and described as human-tasks or human-human relations, not as elements. As a result, Team, Task, and Place remain as the major elements of the team contexts model.

We developed sub-elements of each major element. Team consists of members. Members correspond with humans in the previous framework. Task consists of a target system, devices, and resources. We define Task as "operation or interventions on a target system to achieve its desired state" [3]. A target system is defined for each specific context, such as airplanes for air traffic control, a patient for a surgical operation, etc. A target system usually has many variables, such as the position, velocity, and altitude of an airplane. Device and resource are physical tangible materials. Resource indicates consumable materials. Place consists of areas and sections, which are divisions of working space for its use.

Table 1 shows model elements and examples of their attributes. Table 2 shows examples of the relationships between model elements that can be defined by the T^2P model.

Table 1. Major elements, sub-elements, and attributes of the T^2P model

Major elements	Sub-elements		Attributes
Team	Member		Age, sex, etc.
Task	Target	Variable	Capability of quantification, objective, priority, desired value, perceived sense, possibility of false perception, active change, non-routine events, etc.
	Device		–
	Resource		Limitation
Place	Area		Size, etc.

Table 2. Relationships between sub-elements in the T^2P model

		Team	Task			Place
		Member	Target/variable	Device	Resource	Area
Team	Member	Acquaintance, direct communication, indirect communication, chain of command	Assignment, capability			Existence
Task	Target/variable		Correlation	Assignment	Assignment	Existence
	Device			–	–	
	Resource				–	
Place	Area					Distance, time to access

2.2 Prescriptive Similarity

Based on the T^2P model, we considered how to assess the similarity between different contexts. Table 3 shows 36 criteria selected to evaluate the degree of the context similarity. Those criteria related to tasks are the extension of the task components proposed by Liu et al. [4]. Some criteria are defined and can be measured by using the attribute values of the model elements. Others can be measured by nominal, ordinal, interval, and proportional scales. For example, team size is defined as the number of members in a team. Goal clarity can be quantified by, for example, the Likert scale. While similarity in cognitive and behavioral data obtained in the experiment refers to the descriptive similarity, the prescribed criteria, based on the model, refer to the prescriptive similarity in this study.

Table 3. Criteria to evaluate the degree of context similarity

Related major elements		Criteria
Team		Size, existence of leader, communicability, acquaintance, depth of hierarchy, commandability
Task	Goal	Clarity, quantity, conflict, change
	Input	Clarity, quantity, diversity, inaccuracy, active change, redundancy, non-routine events
	Operation	Clarity, quantity of paths, quantity of actions/steps, repetitiveness, types of actions, activeness, risk, spatial accuracy, temporal accuracy
	Time	Duration, waiting time
	Device	Quantity
	Resource	Quantity, limitation
	Role	Assignment
Place		Indoor/outdoor, size, quantity of areas, flexibility

3 Experiment

We prepared three experimental tasks: the marshmallow challenge (Task M), the breakfast task (Task B), and the marshmallow challenge with orders (Task MO). Task MO has prescriptive similarities with both the other two tasks. Task M is a well-known team exercise to build a free-standing tower using dry spaghetti, tapes, and strings with a marshmallow on top. Task B is to cook light meals as ordered according to the recipe provided. There are four kinds of menus: coffee, steamed cake, biscuit with marshmallow, and dessert mousse. The marshmallow challenge with orders is to build ordered, free-standing towers using dried noodles, tapes, and strings. The height of the towers is 5 cm, 10 cm, 20 cm, and 30 cm.

Using these three experimental tasks and contexts, we conducted experiments with 10 three-person teams and gathered cognitive and behavioral data for comparison under different contexts (Fig. 1).

Fig. 1. The marshmallow challenge, the breakfast task, and the marshmallow challenge with orders

3.1 Formalization of Experimental Contexts

We formalized three experimental contexts employing the T^2P model. As an example of the formalization, attributes of target systems and Member-Target assignment relationships are shown in Table 4 and Fig. 2, respectively.

Table 4. Formalization of Target system

Formalization
Task = {$Target_1$, $Target_2$, …}
Target = {$Variable_1$, $Variable_2$, …}
Variable = {capability of quantification, objective, priority, desired value, perceived sense, possibility of false perception, active change, non-routine events}
- capability of quantification = quantity, quality
- objective = objective, peripheral
- priority of objective variable = 1, 2, 3, …, NA
- desired value = w, w/o
- sense to percept = {touch, taste, hear, see, smell}
- possibility of false perception = w, w/o
- active change = w, w/o
- non-routine events = w, w/o
Instances
The marshmallow challenge = {Tower}
- Tower = {height, free-standing, marshmallow}
height = {quantity, objective, 2, w/o, (see), w, w/o, w/o}
free-standing = {quantity, objective, 1, w, (see, touch), w, w, w/o}
marshmallow = {quantity, objective, 1, w, (see), w/o, w/o, w/o}
The breakfast task = {Customer, Food}
- Customer = {offered items}
offered items = {quantity, objective, 1, w, (see), w/o, w/o, w/o}
- Food = {progress}
progress = {quality, peripheral, NA, w, (see), w/o, w/o, w/o}
The marshmallow challenge with orders = {Customer, Tower}
- Customer = {offered items}
offered items = {quantity, objective, 1, w, (see), w/o, w/o, w/o}
- Tower = {height, free-standing}
height = {quantity, peripheral, NA, w, (see), w, w/o, w/o}
free-standing = {quantity, peripheral, NA, w, (see, touch), w, w, w/o}

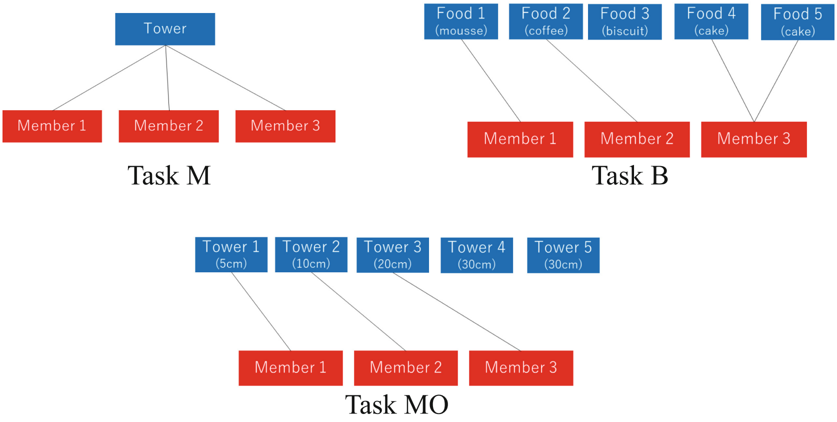


Fig. 2. Member-target assignment structures

The criteria are quantified and calculated mainly based on the formalized contexts. Table 5 shows several examples of the prescriptive similarity criteria.

Table 5. Examples of prescriptive similarity criteria

Criteria related to Team
Quantity = |{Member}| - M = B = MO = 3
Criteria related to Task
Goal-Clarity = $\frac{\lvert\{\text{Variable} \mid \text{capability of quantification=quantity,objective=objective}\}\rvert}{\lvert\{\text{Variable} \mid \text{objective=objective}\}\rvert}$ - M = 0.67, B = 1.00, MO = 1.00
Goal-Conflict: with or without negative correlations between objective variables - M = w, B = w/o, MO = w/o
Input-Active change = |{Variable | active change = w}| - M = 1, B = 0, MO = 3
Operation-Clarity: with or without manuals - M = w/o, B = w, MO = w/o
Role-Assignment: The average degree of target node in member-target assignment network - M = 3, B = 1, MO = 1
Criteria related to Place
Quantity = |{Area}| - M = 1 or 3, B = 7, MO = 4

3.2 Experimental Method

Three experimental tasks were conducted in a random order by 10 three-person teams. The duration of M, B, and MO are 18 min, 10 min, and 10 min, respectively. All the participants were college students.

To gather cognitive data, a questionnaire survey was conducted after the completion of each task. The questionnaire consisted of three sections: personal workload, team workload, and task characteristics. We used NASA-TLX [5] and Teamwork Workload Scale [6] to measure personal workload and team workload. NASA-TLX consists of six question items: mental demand, physical demand, temporal demand, performance, effort, and frustration. The Teamwork Workload Scale consists of eight question items: team leadership, team orientation, performance monitoring, backup behavior, adaptation, trust, shared mental model, and communication. The question items on the task characteristics corresponded to the selected prescriptive criteria: Goal-Conflict, Goal-Change, Operation-Clarity, Input-Active change, Device-Quantity, and Place-Quantity. The behavioral data gathered in this experiment included team performance and the number of moves in a session.

4 Result and Discussion

The purpose of analyzing the experimental results is to confirm whether the prescriptive similarity can predict the descriptive similarity, assuming that if two contexts are similar in essence, the cognition and behavior under those contexts will exhibit similarly. In other words, we analyzed correlations between the similarity criteria and cognitive and behavioral data for the validation of the similarity criteria based on the T^2P model.

4.1 Correlations Between Prescriptive and Descriptive Similarity

The analysis is still in progress, therefore some examples of the correlation found between prescriptive and descriptive similarity so far is presented in this section.

Number of Moves. As shown in Fig. 3, there were correlations between the number of moves, from one place to another, in the experiments and the number of areas, which is the prescriptive criteria regarding Place (see Table 5). This result suggests that the number of areas in the meso-cognitive context should be designed equal to or similar to that in the real working environments.

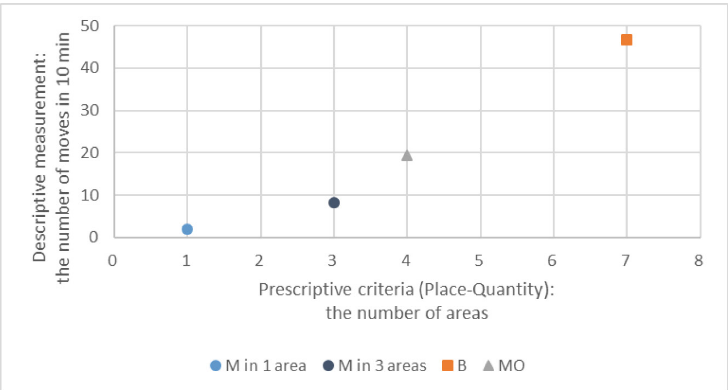


Fig. 3. The number of moves

Team Leadership. As shown in Fig. 4, there were correlations between the team leadership, which rates how proactively they need to interact with other team members, and the number of members allowed to work for one target. This result suggests that role assignment is one of the important factors in designing a meso-cognitive context.

Goal Conflict. As shown in Fig. 5, the percentage of participants who perceived goal conflict during the experiment was the highest in Task M. Among the three tasks in the experiment, only Task M had the prescriptive goal conflict. This implies that it is possible to predict the occurrence of actual goal conflicts in the meso-context by the formalization and design of the context.

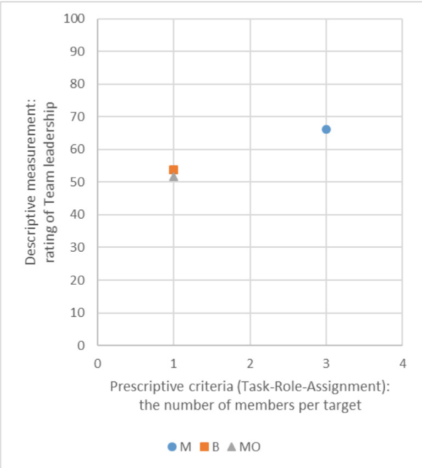


Fig. 4. Team leadership

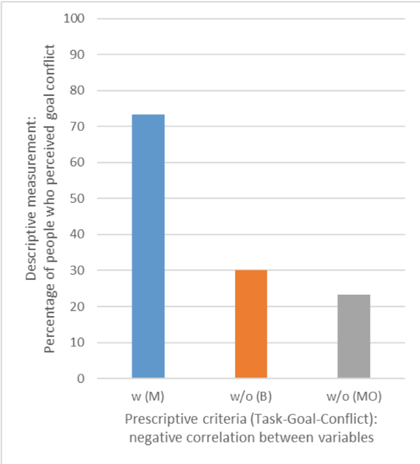


Fig. 5. Goal-conflict

4.2 Descriptive Similarity that Is Difficult to Prescribe

Some of the descriptive similarities were difficult to predict by prescriptive similarity. For example, the rating of operation clarity, which rates the ambiguity of operation, was expected to be much lower for Task B, as cooking manuals were provided in Task B. However, the actual rating in Task B did not show a big difference from that of the other tasks (Task M; $p = .73$, Task MO; $p = .63$). It is implied that other conditions and variables have an impact on the degree of ambiguity of operation. Such unexpected results also provide important information and insight into the design of similar contexts.

5 Conclusion

We developed a team context model named T^2P and presented a criteria to measure the prescriptive similarity of team contexts. An experiment for the validation of the proposed model and criteria were conducted. It was found that there were correlations between some prescriptive and descriptive similarities, suggesting that the T^2P model and the prescriptive similarity criteria have a certain validity for meso-cognitive context design. The results of the analysis also imply that not only can prescriptive similarity be utilized for designing a meso-cognitive context but the same can be done for descriptive similarity as well. In order to further design similar contexts, it is considered appropriate to modify the criteria for the prescriptive similarity by referring to descriptive similarities observed in experiments.

Acknowledgments.. This work was partly supported by JSPS KAKENHI Grant Number 16H03131.

References

1. Klein, G., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., Hollnagel, E.: Macrocognition. IEEE Intell. Syst. **18**(3), 81–85 (2003)
2. Kanno, T., Inoue, S., Karikawa, D., Chao, D.: Modeling framework of team contexts for the design of laboratory experiments and team training. In: Proceedings of the International Conference on Applied Human Factors and Ergonomics, pp. 155–161 (2017)
3. Kanno, T., Mitsuhashi, D., Inoue, S., Karikawa, D., Nonose, K.: Formalization and quantification of team contexts for meso-cognitive studies. In: International Conference on Human Systems Engineering and Design: Future Trends and Applications, pp. 838–844 (2018)
4. Liu, P., Li, Z.: Task complexity: a review and conceptualization framework. Int. J. Ind. Ergon. **42**, 553–568 (2012)
5. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Human Mental Workload, North-Holland, Elsevier Science Publishers B. V., pp. 139–183 (1988)
6. Nonose, K., Yoda, Y., Kanno, T., Furuta, K.: An exploratory study: a measurement of workload associated with teamwork. Cogn. Technol. Work **18**(2), 351–360 (2016)

Towards a Mobile Phone Pupillometer

Ricardo Vigário[1,2(✉)], Mariana Santos[1], Raúl Vences[1], and Carla Quintão[1,2]

[1] Departamento de Física da Faculdade de Ciências e Tecnologia da, Universidade Nova de Lisboa (FCT-UNL), Monte da Caparica, 2892-516 Caparica, Portugal
{r.vigario, cmquintao}@fct.unl.pt, {mnm.santos, r.vences}@campus.fct.unl.pt
[2] Laboratório de Instrumentação, Engenharia Biomédica e Física da Radiação (LIBPhys-UNL), Monte da Caparica, 2892-516 Caparica, Portugal

Abstract. It is said that the eyes are the windows of the soul. Although rather poetic, such a statement becomes clearly more relevant if we acknowledge that dynamic changes in pupillary dimensions convey a great amount of information, regarding the general psychological and neurophysiological condition of the observed person. Most commonly used pupillometers are rather expensive, and may require highly controlled experimental setups to be used. Those can limit greatly the applicability of the devices in practise. Based on a basic smartphone and a cardboard support, originally proposed for virtual reality applications, we developed a portable pupillometer, which can be used in natural, as well as controlled stimuli conditions. The proposed device fits the category of monocular video pupillometers, meaning that it continuously measures one of the user's eyes, while the other may receive pre-determined visual stimulation. To help validating the use of the pupillometer, we measured the evolution of pupillary dimensions during a standard "ice bucket" hand test. The results followed quite accurately the behaviour reported in literature, with increases of pupillary diameters in the order of 15% to 20%, as a result of placing or removing the subject's hand. Recorded pupillary reaction response times were about 2.6 s, which suggest an interplay between the sympathetic and parasympathetic nervous system controlling activity.

Keywords: Low-cost pupillometer · Autonomous nervous system · Attention · Portable · Neuroscience · Psychological assessments

1 Introduction

Pupillary reactivity to internal or external stimuli contains a great amount of information regarding a person's physical and psychological wellbeing, as well as how (s)he reacts to such stimuli. That reactivity is one of the first clinical tests, when evaluating the integrity of the autonomic nervous system (ANS, [1]). There, the subject's eyes are made to oscillate between a normal light exposure condition and that resulting from direct light stimulation, while the medical professional observes the concomitant changes in pupillary dimensions [2, 3].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 194–206, 2020.
https://doi.org/10.1007/978-3-030-20040-4_18

Another common example of the workings of the ANS is one's reaction to a frightening or embarrassing event. Our heart start racing, the palms of our hands sweat, while our mouths become dry. Those are some of the typical hallmarks of the activity of the sympathetic nervous system, which is often associated to a "fight or flight" reaction to a personal attack. Conversely, the parasympathetic system has an antagonistic function, facilitating vegetative, nonemergency responses [1]. Activity in those two ANS subsystems leads also to an increase or decrease in pupillary dimensions, respectively.

Whether for clinical purposes, or to assess a person's reaction to a given stimulus, say, while observing and formulating a particular decision over a movie or a publicity spot, pupillometry has been considered a valuable tool to evaluate how human beings interact with their environment [3], as well as to support psychological assessments and basic health monitoring [2, 4, 5].

Existing pupillometers allow for the study of variations in pupillary dimensions, over time, at a basal level or in response to a number of a variety of different stimuli. With that goal in mind, the device proposed in the current manuscript can be seen to fit the category of monocular video pupillometers, as it continuously measures one of the user's eyes, while the other is fre to be stimulated visually, by the device's own screen.

Commercially available pupillometers are usually rather expensive, with their costs ranging from several tens of dollars to thousands. The lesser expensive ones are often incapable of evaluating pupillary dynamics, whereas the more expensive tend to have rather restrictive experimental conditions for their deployment [6]. The main specifications set for the proposed device were a low-cost and a high degree of portability. In this way, we expect to augment the range of experiments in which such a pupillometer may be used in practice.

The system proposed in this manuscript was developed as a M.Sc. thesis [6], and follows the eye tracker device presented in [7]. Not unlike the one suggested in [8], the proposed device is built around a smartphone application, for visual stimulation and recording, and a computer, for post-processing the collected data. It differs from the one in [8] since it is not restricted to the light reflex test application. Here, through an Android application, the phone continuously records the image of the pupil, with its front camera, while its screen may be used to present specific types of stimuli. The mobile phone is kept at a fixed relative position to the user's eyes. The recorded information, which is transferred to a computer, can then be processed through a Matlab interface with a pupillary segmentation algorithm.

To validate the results obtained by the developed pupillometer, a cold stress test was carried out with 12 subjects. One of the main purposes of this study, routinely used in the field of neuroscience, is to understand the functioning of the autonomic nervous system during a cold stimulation [9].

2 Materials and Methods

The device follows on the one proposed in Gamas' M.Sc. thesis [10], which was later validated by Vences [11]. The stimulus management and delivery, as well as the capture and storage of all ocular activity were performed by a mobile phone, through a dedicated app, which was developed using Android Studio 2.3. All processing of the acquired images and videos was performed offline, using Matlab[1].

2.1 The Experimental Setup

The Phone. The image recording was done using a *Huawei P8 Lite,* mobile phone, from 2017, with Android 6.0 operating system. It has a frontal camera with 8 M pixel resolution and a wide angular objective, since it is mostly used for short range image capturing. It captures video at a maximum of 30 frames per second and a 1080p resolution. The phone has also a 5.2″ *Full HD* display, with a resolution of 1920 × 1080p.
Because the used phone differed from the one in [7, 10], several parameters needed to be adjusted in the processing software. Those pertained mostly to differences in the focal length and resolution of the cameras, and their relation to the sizes of the recorded pixels, crucial for a correct estimate of the pupillary area.

The Support. We used Virtoba's V2 mobile phone support ([12], shown in Fig. 1), designed following the specifications in the original Google Cardboard [13], for 3D visualization and VR applications.

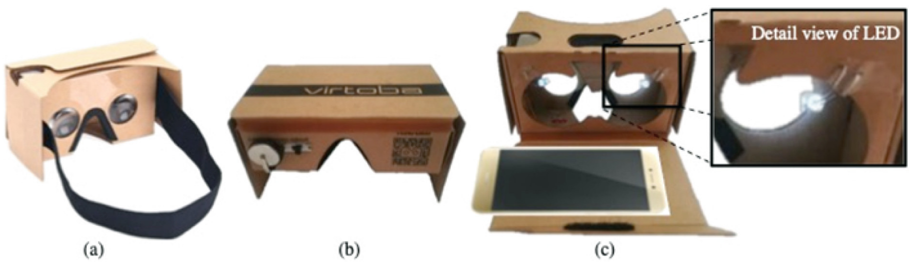


Fig. 1. The proposed pupillometer. (a) Back view, corresponding to the direction where the subject's eyes and nose contact the support. The other frames show the bottom (b) and front (c) views of the cardboard support. A zoom, on the right, shows the mounted LED lights.

The lenses (Fig. 1a) guarantee that the eyes will appear correctly focused in the recorded videos, even at a very short distance to the camera. The strap, on the other hand, insures a stable relative position between the eyes and the phone. The eye tracker in [8] relied on a frontal flash, in the phone, to secure sufficient lighting conditions for the eyes. To render the current pupillometer less dependant on such particular phone feature, we added two LED lights (see Fig. 1c, and the zoomed detail on the right). In the middle frame we see a bottom view of the cardboard, where the power supply to the LED and switch are clearly visible. As was the case with the eye tracker, also the proposed pupillometer imposes a small offset between the phone and the edge of the

[1] MATLAB Release 2012b The MathWorks, Inc., Natick, Massachusetts, United States.

support. The main reason for that offset is to avoid the cardboard surrounding the circular opening. In addition, it reduces also the eccentricity of the camera, when compared to the recorded eye.

The choice of the LED's power, as well as its position in the cardboard was carefully considered, so that the disruptive effect of its reflections over the eyes was reduced to a minimum. Those artefacts are particularly relevant when the subject looks in the general direction of the light. In the current study, subjects were asked to stare ahead, without much wandering. In addition, any such artefact was dealt with in post-processing.

Stimuli and Calibration. The complete system was designed with combined eye tracker and pupillometer studies in mind. Hence, visual stimulation is allowed. Yet, the current study did not comprise any visual stimulus. To validate the device, we have used a cold stress test instead.

Since the dimensions of the pupil are strongly coupled to the quantity of external light, we assumed that a thorough calibration procedure should be carried out, in order to remove such spurious variations from more important types of (emotionally or sensation-oriented) stimuli. Nevertheless, and possibly because of the introduction of a constant light environment, caused by the LEDs, no such variations were observed.

Still, and so that one may have a reference baseline for any future study, we generally recommend the recording of some 10 s of video, from the eyes, prior to the experiment, while subjects are presented with a visual stimulus with a visual complexity close to the one used during explicit stimulation.

Temporal Resolution. Video recording, with the *P8*, is limited to 30 frames per second, with full HD resolution. Yet, we observed that most of the applications we had in mind did not require such a high sampling rate. In fact, as can be observed in Fig. 2, variations in pupillary dimensions vary little, when moving from 30 to 10 frames per

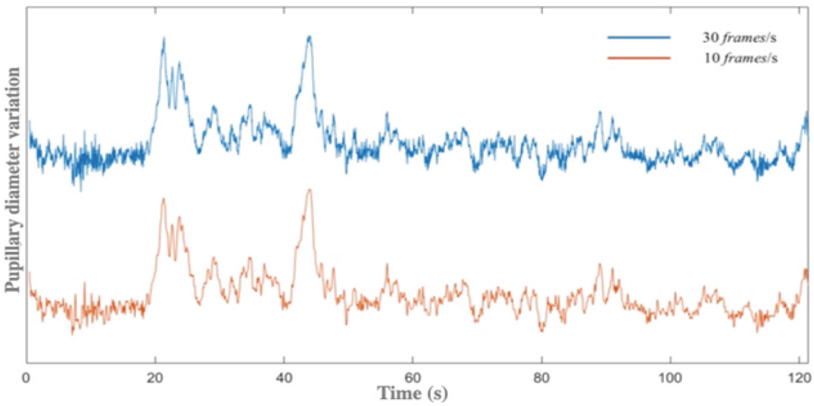


Fig. 2. Video frame rate. Two traces with the variation in pupillary diameter, as a result to inserting the hand into melting ice on a bucket. The blue line (top) corresponds to information gathered at 30 frames per second, whereas the red (bottom) line corresponds to 10 frames per second.

second. See, for example, the reaction to the insertion of one hand in cold water, which occurred at around 20 s from the beginning of the recording. Both lines show clearly the resulting change in pupillary dimensions. Hence, all data presented in this work will be based on videos with the lower frame rate.

2.2 Pre-processing

The first step proposed in [7] is to isolate the region, in each image, that corresponds to the eye. In fact, close to 75% of one image cannot be considered relevant for the intended analysis. Here, though, such step was considered optional, and most results will be presented without it.

From Colour to Saturation. Mobile phone images are often acquired in RGB (*Red, Green* and *Blue)* format. In earlier works we have reduced that data to the *Red* channel, as we considered it to convey the most contrasting information regarding the identification of the pupil. Here we use yet another representation. First, the image coding is transformed to HSV (*Hue, Saturation* and *Value*), from which we will take only the *Saturation* channel information (as depicted in Fig. 3).

In addition to the aforementioned transformation, we added an inversion to the intensity map of the single-channel image, to simplify human perception. Please note that the latter transformation does not affect, in any way, the contrast between structures in the image.

Fig. 3. Leftmost frame shows the colour image of the eye (codded in HSV). Taking apart only the *saturation* channel in that image (middle frame), the contrast between the pupil and the iris becomes very apparent. The rightmost frame depicts a final inversion of the intensity map.

As one can observe in Fig. 3, the illuminating LED causes an artefact in the image, clearly visible in the leftmost frame, within the iris, but contiguous to the pupil. Such artefact is often of a much smaller dimension than that of the pupil; hence there is little risk of confusing one with the other. Nevertheless, when the pupil is at its smallest possible opening, such confusion may still occur, and will need to be properly handled (see "Fine identification" in Sect. 2.3).

In addition, since our study did not require following a particular visual stimulus on the phone's screen, it was possible to ask for the subject to keep a visual fixation direction that minimised morphological and lighting artefacts. Such fixation was also easier to keep than less natural gazing directions.

Additional information regarding image acquisition and pre-processing can be found from [7, 10, 11], where most of the methodology was described.

2.3 Pupil Identification

The identification of the contours of the pupil is made using a two-step approach. The is a coarse identification of possible pupil candidates, and reduces the burden of a thorough analysis of the full image/frame. In the second step are included a series of procedures that deal with specific artefacts found around the targeted object.

Coarse Approximation. In this first stage, the goal is to identify, in a rough approximation, the location of the pupil. Since its overall shape, is very closely approximated by a circle, we use Matlab's *imfindcircles()* function to highlight regions in the image with such a round shape (see Fig. 4b). That function is parameterised by the range of radii for the circles, which will be considered in the search, as well as how close to a circle should the found shape be.

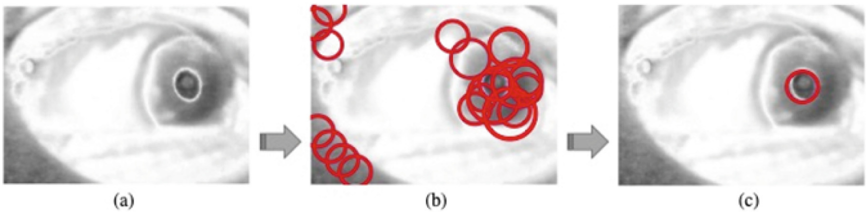


Fig. 4. (a) Original image. (b) Result of applying the regions in the image comprised of circles. (c) Circle that best matches the pupil.

Since it is a first approximation, one does not need to be too restrictive in the search for circular shapes, yet the range of radii considered is still rather narrow. Nevertheless, as can be seen in the middle frame of Fig. 4, many pupil candidates are often found. To find the best matching circle, we use the pupils found from previous frames, and compute the distance between their radii and locations with those found on the current frame. The estimated pupil with smallest distance is considered for further processing.

The only frame for which the procedure described above does not apply is the first one in the video. In the absence of a clear reference, one needs to use a different heuristics. By imposing that the initial fixating point is straight ahead from the subject, one insures that the pupil will be captured as a very close to circular shape. Hence, in the first few frames, the sensitivity to the circle should be rather high. Furthermore, the imposed fixating direction will also narrow down the searchable region.

Note that the procedure described here is rather general, and does not require the eyes to be fixating a specific direction throughout the experiment.

Fine Pupil Identification. Once the general pupillary region has been identified, its fine delineation can be done in a couple of simple procedures, all of which applied solely in the close vicinity of the estimated pupil.

The first step is illustrated in Fig. 5. Assuming the region and the pupil are centred, and the former exceeds the dimensions of the latter, the main diagonals should comprise intensities corresponding to the iris, the pupil and the iris again (clearly visible in the right-hand side of the figure).

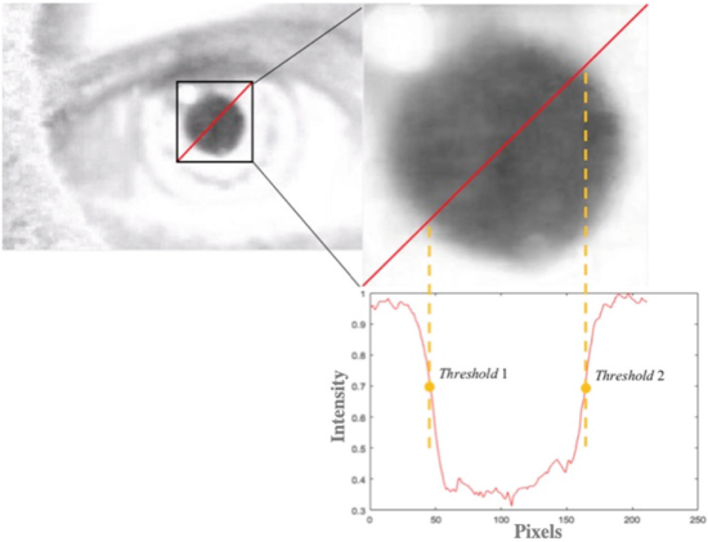


Fig. 5. Finding the pupil's limits. On the left-hand side is one frame of the video, with the coarse estimated region of the pupil highlighted. On the right-hand side that region is zoomed (top), and the intensities displayed, along the main diagonal (bottom).

Two thresholds help determining the limits of the pupil along the diagonal. Each one should be determined separately, as illustrated in Fig. 6, and discussed next. In any case, those values are found as 50% of the difference between the average of the high intensity pixels, and that of the low intensity ones.

In some situations, as mentioned earlier, the light used within the cardboard may produce artefacts, which tends to appear as a white circle, close to the border of the pupil (clearly visible on the top-left corner of the zoomed highlight in Fig. 5). To cope with those artefacts, one needs clearly to treat differently each of the borders between the pupil and the iris (see Fig. 6). No single threshold would suit this task.

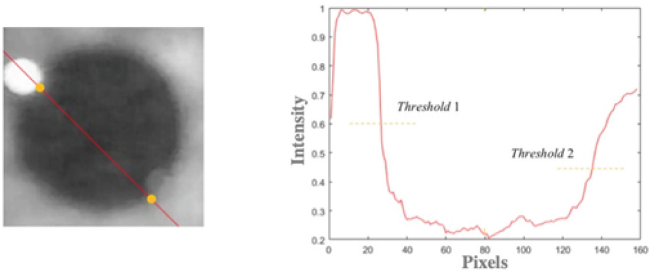


Fig. 6. Two individually determined thresholds handle potential light artefacts, caused by the LEDs.

We are not limited to finding pupillary limits only from the main diagonals. In reality, we have pursued the same procedure for 38 different lines, summarily illustrated in Fig. 7a. From those, 76 different limits were found (Fig. 7b). Not all were correctly estimated, *e.g.*, due to artefacts. A few examples are clearly visible on the top-middle frame of Fig. 7. To exclude border estimates from further processing, we started by calculating the centre of the pupil, based on all estimates. All border points that were more than 15 pixels further away than the median distance to the centre, were excluded from further processing (see Fig. 7c).

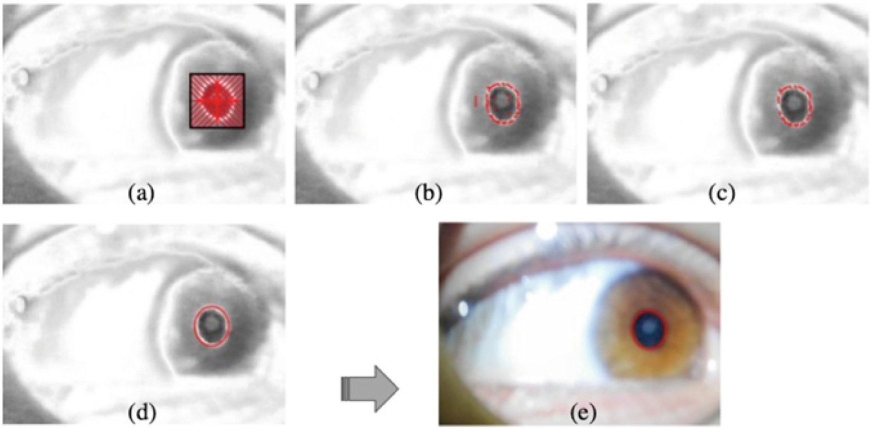


Fig. 7. Complete determination of the pupil's limits. From top to bottom, and left to right, execution of the limit searches shown in the previous figures, for a set of 38 directions, centred in the pupil; removal of wrongly estimated limits; fitting of an ellipse to the support limiting points; superposition of the best ellipse onto the original image.

The final step of processing was to find the ellipse that best matched the border points that remained from the previous analysis (Fig. 7d). For that purpose we used Matlab's *fit_ellipse()*. The main reason why we used an ellipse, as the fine estimate of the pupil's borders, is that the cardboard lenses, together with the eccentric location of the camera, when compared to the position of the eyes, tends to change the rather circular shape of the pupil into an ellipse. It is also important to note that the function used is rather robust, requiring as few as 5 points to estimate the ellipse.

Continuity and Blinks. As mentioned, we should explore the fact that the eyes do not change drastically from one position to another. We should remember that, even at a frame rate of 10 images per second, there are but 100 ms between each frame. Earlier, this reasoning was used, in the coarse estimate of the region corresponding to the pupil, to secure an easy determination of which circle corresponded to the pupil. Now we will use it further to detect, and remove, particularly difficult artefacts, such as eye blinks.

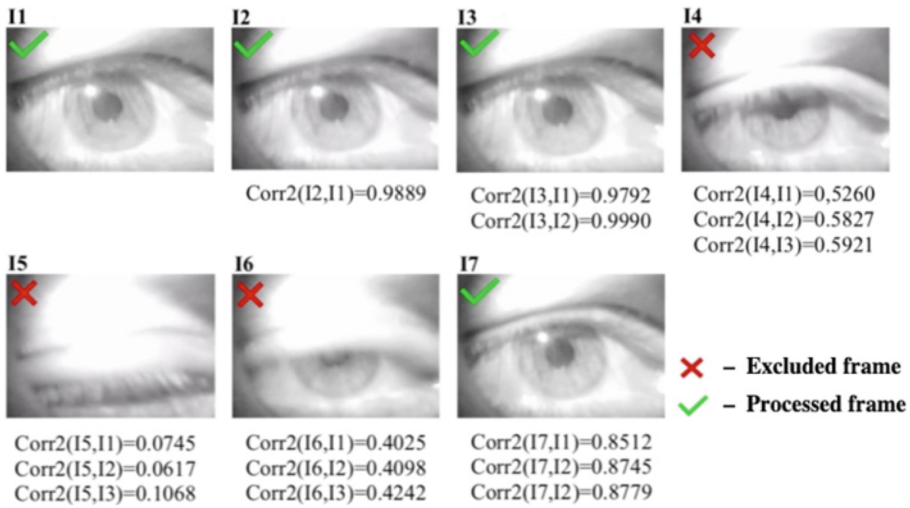


Fig. 8. Removing blinking artefacts through checks in image continuity along various video frames.

When a person blinks, the pupil estimate will most certainly fail, either because it is completely invisible (as shown in frame I5 from Fig. 8), or because it is partly occluded (as in I4). One way to prevent those miss-estimates from contaminating the collected pupillometric data is to exclude all frames that do not show a basic amount of resemblance to the remaining ones.

A simple similarity measure between frames, such as the two dimensional correlation, can identify miss-identified pupils. Assuming that the direction of gaze does not change significantly in consecutive frames, one may expect that the correlation remains rather high. If that is not the case, then we should exclude the newer frames from analysis. 7 frames of a video are shown in Fig. 8, together with the correlations between that frame and the 1st, 2nd and 3rd. When such correlation exceeds a given value (which we took as 0,7), than the frame can be considered for further analysis.

In order to keep track of true estimates of the pupil, after discarding the artefact ones, one may add a memory element to the search for continuity. In the example illustrated in Fig. 8, that memory extends to the past 7 frames.

3 Cold Stress Test

To test the efficiency of the proposed pupillometer, we have used the classic cold stress test [9]. After 20 s of baseline recording, the subjects were instructed to immerse one hand in ice water (at about 5 °C), through a verbal order. 20 s later, again verbally, subjects were asked to remove their hands from the water. Video recording was stopped 80 s after removing the hand. The last period of time was used to assess the recovery from stimulation.

3.1 Participants

12 persons participated in the experiment. They were between 22 and 25 years of age; 2 were lefthanders; and 8 had (dark) brown eyes, 2 green/brown, and 2 (light) blue/green. All were asked to place the left hand inside the bucket, regardless of their dominant hand. The right hand was used to control the cardboard button, which controlled the progress of the video collection protocol.

3.2 Processing and Feature Extraction

Once the best ellipse was found, pupillary diameters can easily be estimated via the number of pixels that build the curve. Both curves presented in Fig. 2 display such a measure for one of the 12 participants in the study.

In addition to lowering the video sampling rate to 10 frames per second, a low-pass filter was also applied to the dynamic evolution of the pupillary diameter (see Fig. 9).

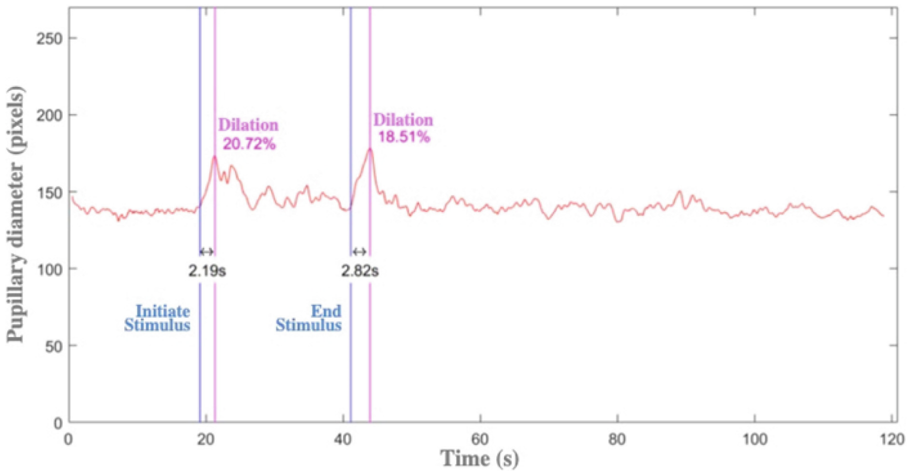


Fig. 9. Dynamics of the diameter of the pupil in a 2 min cold stress experiment. Annotations are superimposed to the graph, containing the start of cold stimulation (hand insertion in the bucket), its end, as well as two of the most relevant sets of analysis features – dilation percentage and reaction times.

Two main features were calculated directly from graphs such as the one presented above: the percentage of pupillary dilation, measured as a percentage of increase in diameter, from the baseline before the rise, and the point of maximum dilation; reaction times to a change in the experimental paradigm (insertion or removal of the hand). Both features are calculated for the subject's reaction to initiating the cold stimulus, 20 s after the beginning of the video recording, and the end of the cold stimulus, 20 s later.

In addition to the aforementioned features, a measure of the average reaction slope was also calculated, as the ratio between the percentage of dilation and the respective reaction time.

Table 1 shows a summary of the results found for all 12 subjects. Dilations occurring as a result to a reaction to the insertion of the hand in cold water were slightly more pronounced than those occurring when removing the hand.

Table 1. Pupillary dynamics, as a response to the cold stress test.

	Pupillary dilation (%)		Reaction times (s)	
	Init. stimulus (D_i)	End stimulus (D_f)	Init. stimulus (TR_i)	End stimulus (TR_f)
S_1	20.72	18.51	2.19	2.82
S_2	15.76	5.77	2.88	1.83
S_3	21.45	14.17	3.00	3.18
S_4	20.63	14.43	3.51	2.89
S_5	17.54	14.57	2.58	2.31
S_6	23.36	8.51	1.68	2.27
S_7	24.07	18.81	3.74	3.34
S_8	18.76	13.98	2.34	2.77
S_9	13.14	7.31	2.01	3.11
S_{10}	17.49	9.42	3.15	2.28
S_{11}	24.20	15.48	2.28	2.01
S_{12}	18.30	6.71	3.30	2.61
$\bar{x}$	19.62	12.31	2.72	2.62
σ	3.3	4.4	0.61	0.46

A 19% increase in pupillary diameter, albeit subtle, can convey sufficiently important information in practice. An illustrative example of such changes is given in Fig. 10.

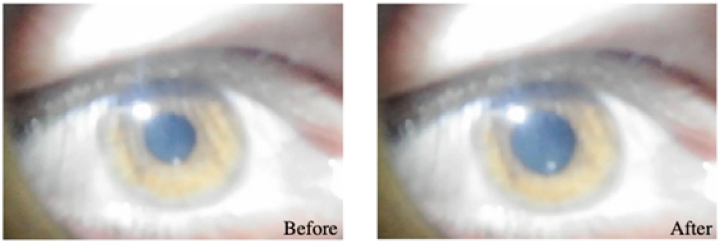


Fig. 10. Illustrative example of a 19% increase in ocular pupillary diameter.

Reaction times to changes in stimulus conditions are rather close to each other 2.72 ± 0.61 s, for the beginning of the cold stress test, and 2.62 ± 0.46 s when removing the hand from the bucket. These values can be compared to those resulting from the light-dark reflex, which are reported to be around 0.2 s; and with the ones resulting from emotional and cognitive stimulation, which vary from 1 to 3 s [14, 15].

Whereas the reflex reaction tends to be more instinctive, and related to a fight or flight induced stimulus, engaging mostly the sympathetic nervous system, our study may be closer related to a control type of reaction, which would not only require the rapid action of the sympathetic, but also the slower parasympathetic nervous system.

One final observation tells us that the significant difference between reaction slopes found for the onset and the offset of the test, $7.2 \pm 0.7\%/s$ versus $4.7 \pm 1.5\%/s$, comes mostly from the difference in dilations occurring in the first and second stimulus changes.

4 Conclusions

The goal of the current research was to build a pupillometer based on a commonly used smart phone. A basic cardboard support should insure a constant distance between the face/eyes and the phone/camera, rendering the complete system rather low-cost and widely portable. With the developed processing software, we show, in this manuscript, that the goal has clearly been reached.

In addition to its cost and guarantee of a fixed geometry of the device, it should still be stressed that the very lightweight solution guarantees an easy and portable use of the device.

To validate the proposed pupillometer, we performed a standard cold stress test with 12 subjects. The outcomes of the test were very much in line with the expected from the literature. In fact, both the insertion of the subject's hand in cold (5 °C) water, as well as the removal resulted in a clear increase in pupillary dimensions. That increase was significantly more pronounced during the insertion than the removal stage (around 20% and 12%, respectively). Pupillary reaction times, of about 2.5 s suggest that both the sympathetic and parasympathetic nervous systems are active in controlling the opening of the pupil in this kind of experiments.

In the future we expect to port all post-processing to the mobile device itself, to render the complete experience truly portable. It will also be desirable to merge the current pupillometer with the previously developed eye tracker. That step should benefit from the existing common development platform. The combination is likely to open new directions of applied research on human reactions to emotionally charged complex visual stimulation.

References

1. Kandel, E.R., Schwartz, J.H., Jessell, T.M., Breedlove, S.M., Watson, N.V.: Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience. 7th edn. Sinauer Associates (2013)
2. Wilhelm, H., Wilhelm, B.: Clinical applications of pupillography. J. Neuro-Ophthalmol. **23**(1), 42–49 (2003)
3. Bremner, F.D.: Pupillometric evaluation of the dynamics of the pupillary response to a brief light stimulus in healthy subjects. Invest. Ophthalmol. Vis. Sci. **53**, 7343–7347 (2012)

4. Fehrenbacher, D.D., Djamasbi, S.: Information systems and task demand: an exploratory pupillometry study of computerized decision making. Decis. Support Syst. **97**, 1–11 (2017)
5. Tsukahara, J.S., Engle, R.W.: The relationship between baseline pupil size and intelligence. Cogn. Psychol. **91**, 109–123 (2016)
6. Santos, M.: Desenvolvimento de um pupilómetro de baixo custo. M.Sc. thesis in Biomedical Engineering, Faculty of Sciences and Technology, University NOVA of Lisbon, November 2018
7. Vigário, R., Gamas, F., Morais, P., Quintão, C.: Development of a low-cost eye tracker – a proof of concept. In: 9th International Conference on Applied Human Factors and Ergonomics (AHFE), pp. 285–296 (2019)
8. McAnany, J.J., Smith, B.M., Garland, A., Kagen, S.L.: iPhone-based pupillometry: a novel approach for assessing the pupillary light reflex. Optom. Vis. Sci. **95**(10), 953–958 (2018)
9. Pozos, S., Danzl, D.F.: Human physiological responses to cold stress and hypothermia. In: Pandolf, K.B., Burr, R.E. (eds.) Medical Aspects of Harsh Environments. Textbooks of Military Medicine, vol. 1, pp. 351–382 (2002)
10. Gamas, F.: Desenvolvimento de um *Eye Tracker* de Baixo Custo. M.Sc. thesis in Biomedical Engineering, Faculty of Sciences and Technology, University NOVA of Lisbon, September 2017
11. Vences, R.: Validação e Calibração de um Eye Tracker de Baixo Custo. M. Sc. thesis in Biomedical Engineering, Faculty of Sciences and Technology, University NOVA of Lisbon, October 2018
12. Virtoba. http://www.virtoba.com/products/virtoba-v2. Accessed 12 Aug 2018
13. Google Cardboard for Manufacturers. http://static.googleusercontent.com/media/vr.google.com/pt-PT//cardboard/downloads/manufacturing-guidelines.pdf. Accessed 10 Aug 2018
14. Sirois, S., Brisson, J.: Pupillometry. Rev. Cogn. Sci. **5**(6), 679–692 (2014)
15. Wang, J.: Pupil dilation and eye tracking. In: Handbook of Process-Tracing Methods, pp. 1–33. P. Press (2009)

Systems Interaction Design and Automation

Automated Vehicles and Schools: An Analysis of Deployment Issues

Michael Clamann(✉)

University of North Carolina Highway Safety Research Center,
Campus Box 3430, Chapel Hill, NC 27599, USA
clamann@hsrc.unc.edu

Abstract. School zones are dynamic environments with regulations for motor vehicles and pedestrians that vary by locality. It is critical that drivers follow local rules to ensure the safety of children entering and leaving the school zone. This combination of a dynamic environment and critical safety requirements will be a challenge for fully automated vehicles (FAV), which often do not perform reliably when faced with uncertainty. However, to date there has been little research addressing the operation of FAVs in school zones. In response, this work identifies key research needs surrounding FAV deployments near schools and on school routes, with an emphasis on K-8 student safety. The research needs were informed by interviews with traffic safety researchers and school transportation professionals who identified common transportation issues related to school zones and FAV deployment. The results can be used by developers and community stakeholders to plan for future FAV deployments.

Keywords: Schools · Automated vehicles · Vulnerable road users · Children

1 Introduction

Automated vehicles (AVs) are designed to execute parts or all aspects of the driving task for prolonged periods without compromising performance due to distraction or fatigue. The impact of AVs on safety is potentially enormous given that human error accounts for an estimated 94% of vehicle crashes [1]. A key challenge in reducing these numbers includes addressing crashes involving pedestrians, who are tragically overrepresented in traffic deaths. In 2017, pedestrians made up 16% of all traffic fatalities [2], although walking accounts for just under 11% of all trips [3]. Annual U.S. pedestrian fatalities increased more than 30% between 2009 and 2016, totaling approximately 6,000 in 2016, the highest level in nearly three decades [4, 5].

Addressing these increases represents a challenge for AVs and poses a topic of ongoing research [6–10]. Automated vehicles thrive in deterministic environments; however, compared to motor vehicles, pedestrian behaviors are not particularly constrained by traffic infrastructure and regulations, which makes them unpredictable much of the time [11]. Pedestrians who make errors in judgement when choosing a crossing gap account for around half of pedestrian fatalities [12]. The unpredictable

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 209–220, 2020.
https://doi.org/10.1007/978-3-030-20040-4_19

nature of pedestrians is further complicated every weekday in school zones where large numbers of vehicles interact with children entering and departing schools, each of which may have its own unique approach to regulating traffic. If we want to achieve ubiquitous deployment of *fully* automated vehicles (FAVs), which are vehicles capable of driving anywhere at any time without the aid of a human driver, they will need to be able to navigate even the most difficult environments, which includes school zones. To date there has been little effort to assess the unique challenges of deploying FAVs in school zones to help AV manufacturers and school administrators consider new regulatory or roadway design needs. The purpose of this work was to begin to identify key issues relevant to deploying AVs near schools, with a focus on FAV operating in Kindergarten through 8th grade (K-8) school zones. This work is part of a broader effort in cooperation with the Pedestrian and Bicycle Information Center (PBIC) to develop a framework for AV deployment along school routes.

1.1 School Zones

The Manual on Uniform Traffic Control Devices (MUTCD; 13) sets standards for the messages, locations, sizes, shapes, and colors of traffic control devices in the U.S. The MUTCD defines a *school zone* as "a designated roadway segment approaching, adjacent to, and beyond school buildings or ground, or along which school related activities occur" [13]. In practice, school zones usually extend one to two blocks in each direction from a school [13]. Speed limits are often reduced in school zones during morning and afternoon hours. Special traffic control measures, including crossing signs, speed signs, and school zone pavement markings, are used to inform motorists that they should drive with special care and extra attention.

Despite the existence of these standards, traffic regulations in school zones can vary greatly. This was a consistent theme among the groups of school transportation experts interviewed for this work (see Methods section). While the MUTCD sets design standards and guidelines for signage and street markings on public roads, state and local law regulates traffic, and other groups recommend "best practices" for setting up routes, it is difficult to identify consistent behaviors that predict how students move from their arrival mode (i.e., walking, biking, school bus, private transport, etc.) to a school. For example, some schools have their own road infrastructure with bus lanes on campus property separated from parent drop off zones. Other schools are built adjacent to city streets, requiring large numbers of cars to queue on public roads to drop off and pick up children. Figure 1 (next page) illustrates this latter example, showing cars driving on the wrong side of the road and children exiting vehicles in the middle of the street with no crossing guard in sight.

Problem areas like the one depicted in Fig. 1 are being made worse by increasing numbers of children being driven to school. In 1969, half of K-8 students walked or rode bikes to school. By 2001, this number had decreased to just 15% [14]. Currently, close to half (45%) of students ride to school in a personal vehicle [15]. In some locations, these personal trips to and from schools generate as much as 20 to 30% of all

Fig. 1. Vehicles dropping off students near a school (image: Safe Routes to School)

morning rush hour traffic [16]. About 30% of parents cite traffic risks as a barrier to walking or biking [3]. This suggests, somewhat ironically, that parents who are afraid to walk their children to school because of traffic risks solve the problem by contributing to the overall traffic risk.

This increasing volume of parents dropping off and picking their children up from school also adds to the frustration of driving a vehicle through a school zone. Like rush hour traffic jams, school zones are not immune to erratic driver behaviors and road rage. Parents generally want to drop off or pick up their kids as quickly as possible and get on with their day, and like any driver they resent barriers that prevent their progress. This climate of anxious drivers operating vehicles in locations with temporary restrictions that vary from one location to another is now a potential operational design domain (ODD) for FAVs. Therefore, it is imperative to understand issues related to deploying FAVs in school zones so appropriate technology, design and regulatory approaches can be considered.

1.2 FAV Technology

Effective detection is the foundation of FAV technology. Before vehicles can interpret and respond to events and objects in their environment, they must be able to detect important road elements, including pedestrians. FAVs face substantial challenges in accurately and reliably detecting and recognizing pedestrians, who are more difficult to identify, predict, and protect in the event of a crash compared with other road users. Pedestrians have variable physical characteristics and appear in a variety of environments with different background features, obstacles, and weather conditions, making them difficult to see. A 2017 study using the California Institute of Technology Pedestrian Detection Benchmark data (a 10-h video recorded from a vehicle's perspective) suggests a tenfold improvement in this technology is needed to replicate human performance [17, 18]. Vehicle-based sensors can fail, especially when pedestrians are small (like children), too far or too close to the vehicle, or partially occluded by nearby objects [19]. This can account for detection failure at more than double the rate for adults.

Despite these shortcomings, there is little guidance for AV developers regarding school zones. Since 2016, the National Highway Traffic Safety Administration (NHTSA) has released three versions of guidelines to inform AV deployment. In these documents, NHTSA has noted the importance of safe interactions between AVs and other road users. However, the word “school” is mentioned only once: when providing advice to state departments of transportation, NHTSA suggests jurisdictions may want to restrict testing AVs in certain areas, including school zones [20]. Over time, NHTSA’s message has evolved to one supporting rapid deployment, indicating in AV 3.0 that “states should consider reviewing and potentially modifying traffic laws and regulations that may be barriers to automated vehicles” [21]. Based on the information obtained from experts on school transportation, school zones may represent one of these “barriers” And it may be time to review our approaches to regulating school transportation.

With these issues in mind, the purpose of this work was to begin to identify key issues related to deploying FAVs in school zones, with an emphasis on pedestrian safety. As nearly every student walks on school property to some extent, whether from buses, from cars, or from home, this work applies to a broad array of students. The scope of the current work includes FAVs operating in K-8 school zones. The material was informed through interviews with school transportation experts at multiple venues with the goal of identifying themes to assist FAV developers and school stakeholders in the safe deployment of FAVs near schools.

2 Methods

This work builds on a research framework developed for the *Discussion Guide for Automated and Connected Vehicles, Pedestrians and Bicyclists* [22]. The Discussion Guide was developed by an interdisciplinary group of transportation experts to highlight AV topics that affect pedestrians and bicyclists. The Discussion Guide consolidates these topics into 10 interrelated “challenge areas” that provide a basic framework for future research and policy needs. Five of the challenge areas were determined to be relevant to K-8 student safety within school zones. These areas include (see [22] for detailed descriptions of each challenge area):

- Pedestrian and bicyclist detection – the ability of AV vision systems to effectively detect, interpret, and predict movements within and near the roadway
- Communication between AVs and humans – ensuring that the formal and informal communications between pedestrians and bicyclists and AVs are at least as effective as they are with human drivers
- Determining right-of-way – preservation of local laws and social customs governing right-of-way
- Regulation of vehicle speed – maintaining safe vehicle speeds around pedestrians
- Curbside pickup and drop-offs – addressing the challenge of curbside management in the presence of other incoming and outgoing vehicles and pedestrians

Semi-structured interviews [23] were conducted with school transportation experts to broadly define current challenges related to transportation within school zones. Eighteen experts participated in this phase of the research. Participants included staff from the University of North Carolina Highway Safety Research Center (HSRC) and the Institute for Transportation Research and Education (ITRE), as well as the Transportation Research Board's (TRB) School Transportation Subcommittee. Participants were also asked to recommend relevant literature for follow-on reviews. The results of the interviews and literature review were analyzed to identify emerging themes related to K-8 student safety within school zones. The emerging themes were cross referenced against the five challenge areas identified from the Discussion Guide to determine which were relevant to AVs. The results are summarized in the following section.

3 Results and Discussion

Relevant topics were identified through the interviews and review of literature recommended by the transportation experts. Topics identified during the interviews were screened for relevance to the research scope and then compiled for a master list of topics. The master list was then further divided into two broad categories that emerged from the review. Each category was adapted from strategies that appear in the Safe Routes to School Online Guide available at http://guide.saferoutesinfo.org. The categories included:

- Engineering and infrastructure – topics related to FAV technology and/or roadway infrastructure
- Education and enforcement – topics related to human stakeholders, including training and safety-related duties assigned to people

The following sections cover findings from each area.

3.1 Engineering and Infrastructure

The engineering and infrastructure category refers to the design and operation of traffic control devices and physical measures that facilitate safe pedestrian crossings and reduced speeds within school zones. Examples include road markings, signs, and traffic signals. Part 7 of the MUTCD lists principles and standards for controlling traffic in school zones.

Speed. One of the biggest challenges in keeping children safe in school zones includes slowing down traffic. Examples of current traffic calming methods intended to reduce speeds include narrowing lanes by adding chokers and chicanes, installing speed humps and raised pedestrian crosswalks, converting intersections to roundabouts, and adding neighborhood traffic circles. While these infrastructure modifications are often effective, the most recognizable attempt at controlling speeds is with reduced speeds informed through signs and street markings.

School speed limit signs inform drivers when they are approaching a school zone and will need to slow down for school children. The MUTCD provides guidance for installing the recognizable yellow-green school area speed limit signs and "SCHOOL" stencils painted on the road's surface in school zones at a specified distance from marked school crosswalks or a certain distance from the edge of school property. Speed limits on school property vary based on state law and typically range from 15 to 25 mph.

There is a strong safety argument supporting reduced speeds in school zones. Specifically, slower moving vehicles stop in shorter distances and, in the event of a collision with a pedestrian, have a lower chance of resulting in an injury or fatality. A vehicle traveling on a level surface at 20 mph needs about 112 ft to stop in time to avoid hitting a stationary child. This increases to about 200 ft for a car traveling at 30 mph and to 300 ft for a car traveling at 40 mph [24]. Higher vehicle speeds also increase the likelihood and severity of injury [25]. As Fig. 2 shows, if a pedestrian is struck by a car traveling at 40 mph, there is an 85% likelihood that the pedestrian will be killed. This percentage drops to 45% at 30 mph and 5% at 20 mph.

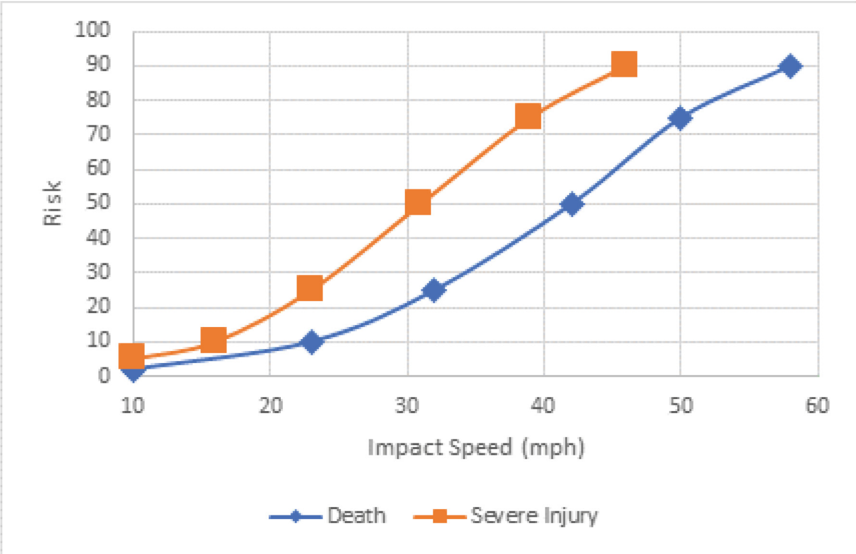


Fig. 2. Car and light truck impact speed and risk of severe injury or death (Adapted from Tefft, 2011)

Despite these figures, a large number of drivers speed in school zones [26]. This 10 mph difference has grim implications. As Fig. 2 shows, a vehicle striking a pedestrian while exceeding the speed limit by 10 mph in a school zone is more than twice as likely to injure or kill a pedestrian than if the impact occurred at the posted 25 mph limit.

Automated vehicles can regulate speeds more reliably than human drivers. If AVs are programmed to adhere to state and local laws [27], then their speed can be limited in school zones to ensure the safety of pedestrians. Whether it is computer vision systems that can detect and interpret standardized road signs or GPS map data that identifies the school zone boundaries, technology is already available that can inform vehicle automation when it approaches a school zone during arrival or departure times. This information could also be used to place a limit on a vehicle's speed.

Recommendations. (1) FAV systems must be able to detect when they enter and exit school zones and prioritize posted speed restrictions. (2) Localities should ensure roadway infrastructure is maintained to facilitate accurate detection by AV sensors.

Traffic Control Devices. Speed is not the only unsafe behavior observed in school zones. Human drivers also routinely violate rules intended to protect pedestrians crossing the street [28]. For example, in one observational study of driver behaviors at intersections, 45% of drivers failed to completely stop at an intersection with a stop sign, and seven percent of those drivers did not slow down. Even when pedestrians were waiting to cross, 36% of drivers did not come to a complete stop, and 24% of drivers did not come to a complete stop at the intersection while pedestrians were crossing [29]. Even when drivers do stop, a large percentage of drivers stop on the crosswalk, blocking the pedestrians [30]. Parents dropping off and picking up children represent one of the biggest safety problems around schools. Parents park illegally, drive through or stop in bus zones, drop off children in the street, allow them to walk between parked cars, and generally ignore established procedures. Administrators attempt to solve these issues by establishing clearly marked pick up and drop off zones and educating parents on safe procedures.

As with speed limits, FAVs are expected to closely follow the directions of traffic control devices, including staying between lane markings, stopping at stop signs, and respecting crosswalks when pedestrians are present. This assumes, however, that the signs and markings are visible to the vehicle's sensors and that they are recognizable. This means that traffic control devices need to be maintained to be consistent with the needs of the vehicle technology, and that they follow a known standard, such as the MUTCD.

This latter point has important implications for schools. In many cases, traffic on school property is regulated by school administrators. Unlike public roads, there is no universally agreed-on or enforced standard for managing traffic on school property. This means there can be tremendous variability among schools, and training an automated vehicle to navigate one school may provide little or no information about navigating another. One school may have its own access road where vehicles line up to pick up or drop off children, while others may be built alongside public roads, which then become pick up and drop off zones twice a day featuring long queues of cars lining the street in both directions while children cross in-between. School administrators use permanent and makeshift signs, traffic cones, school resource officers, plain-clothed school faculty and staff, and numerous other techniques to route traffic. Regulations are often provided directly to parents before the start of each school year, as general knowledge of traffic rules may not necessarily be enough. If the current model remains, FAVs will need to become much better at dealing with uncertainty to safely and effectively navigate through school zones during pick up and drop off times. For example, on a busy city street, a car may be expected to leave its lane to drive around a car that has stopped to let out a passenger. This same behavior would be less acceptable at a school where the custom dictates that cars wait in batches for groups of students to leave their parents' vehicles and drive away only when all the students are out of the road. Alternately, schools could broadly agree on traffic control standards on school grounds and update their infrastructure, but this would represent an unlikely upgrade to groups that often have limited budgets and schedules.

Recommendations. (3) FAVs should consistently comply with school zone traffic regulations. (4) Developers and school transportation administrators should identify low-cost solutions to support safe FAV navigation on school property.

3.2 Education and Enforcement

Education and enforcement activities involve teaching traffic safety principles to all stakeholders, incorporating strategies for interacting with new technology. These programs should involve the community because children, parents, staff, volunteers and law enforcement will all interact with different vehicle types in school zones.

Children. Between starting Kindergarten and graduating high school, a child in the U.S. will travel to and from school almost 4,700 times. Schools are a critical point for teaching children about roadway safety, whether they walk, bike, take public transportation, or ride in a private vehicle. However, compared to adults, younger children are not equipped to safely navigate busy roads on their own, and they need assistance from adults in the community.

A review of the primary causes of pedestrian crashes, broken down by age, shows how children's attention to the road changes with age [31]. When a child is under 10 years old, running into the street, running in-between parked vehicles, and playing in the road are the most common causes of crashes. Children at this age are not as skilled at monitoring peripheral areas of their visual field [32]. As children approach 10 years old, they can begin to learn the difference between safe and unsafe crossing locations and recognize the dangers of oncoming traffic [33, 34]. Between ages 10 and 14, children continue to run into the street, but they also perform unsafe maneuvers on bikes and skateboards, ignore traffic signals, and exit stopped vehicles unsafely. School children between Kindergarten and 8th grade cannot fully internalize complex traffic behaviors and the skills they have learned have not been fully developed into habits [35, 36]. Tragically, in the U.S., 245 pedestrians ages 14 and under were killed, and approximately 15,000 children in this same age group were injured while walking or bicycling in 2016 [2].

Effective traffic safety education programs target children, parents, neighbors, and other drivers in the community. Local enforcement activities should support education efforts and can help to change unsafe behaviors of drivers, bicyclists, and pedestrians. Enforcement is not necessarily limited to law enforcement personnel; students, parents, school personnel, and adult school crossing guards can help reinforce safe behaviors by communicating regulations directly to different road users. In many communities, these are ongoing efforts.

While FAVs are still under development, it is certain they will behave differently than human-driven vehicles. Well-established methods for pedestrian-to-vehicle communication, such as establishing eye contact with the driver [37], will need to be replaced with new techniques, which presently remain an emerging research area [8]. This will be more challenging with mixed fleets of human- and automation-driven vehicles where eye contact is no longer reliable. In this sense, the arrival of FAVs will necessitate a new need for new community education programs for all age groups.

Recommendations. (5) Developers should ensure pedestrian detection systems account for children. (6) Traffic safety educators should account for FAV deployment in future materials intended for children and adults.

Adults. School staff will also need to update training and procedures. Teachers and staff often monitor bus and vehicle pick up and drop off locations to manage safe arrivals and departures. This includes ensuring children enter and exit vehicles safely and, for children who walk and bike, that traffic stops periodically to let them through. Currently, parents and bus drivers are responsible for making sure children enter the vehicle, sit down, and put on their seat belts (if applicable). If children ride in FAVs, someone will still need to ensure they are safely seated in the right vehicle. This may shift responsibility to the limited curbside staff and result in delays, which will be particularly frustrating for human drivers waiting in-between FAVs.

Recommendation. (7) School administrators should plan to update local pick up and drop off procedures to account for FAV-specific regulations and capabilities.

Crossing Guards. School crossing guards will need additional training. This is a challenge because, while there are minimum qualifications for crossing guards [13], there is little consistency in who fills the role. In some schools, it is a resource officer (i.e., a trained law enforcement officer); in others, the crossing guard is a paid position. Sometimes the crossing guard is a volunteer from the community, or a faculty member assigned to the position for a limited time, or there are no crossing guards, so parents help their children cross the street. This will make universal education programs more difficult as the target audience will be different from one location to another. This is concerning, as the complexity of the crossing guard's job is likely to increase with the arrival of FAVs as they master multiple communication modes between human drivers and vehicle automation. Unlike law enforcement officers, crossing guards do not explicitly control traffic flow; rather, they are expected to "pick opportune times to create a sufficient gap" [13; p. 745]. They then walk out into one or more lanes of traffic holding an 18-in. STOP paddle and stay in the road until all the children are safe on the other side. Afterward, the crossing guard returns to the side of the road and vehicles can continue driving. While this seems straightforward to a human driver, these techniques will need to be built into the AVs vision and path planning software, accounting for variability among communities. Crossing guards will need some signal that they (and the children) are safe walking out into traffic and the AVs will need to understand when it is safe to continue driving. These interactions are not currently possible across different types of prototype AVs, and more work is needed to design and test effective means of two-way communication.

Recommendation. (8) FAV developers and school transportation administrators should develop and validate procedures for crossing guards that leverage current MUTCD standards.

Verification and Validation. The fatal crash between an Uber test vehicle and a pedestrian in Tempe, Arizona highlights the risks of testing on public roads. However, the variability among schools presents a real challenge for FAV development. Testing in school zones could put children and school staff at risk; even so, not testing in these challenging situations limits the breadth of FAV capabilities and potentially delays their deployment. Manufacturers will need to work with school officials to find a way to strike a balance between maintaining public safety by testing on closed tracks and in simulation and the robust testing that comes with the variability and uncertainty inherent to real-world testing. In the meantime, FAV manufacturers should be safely collecting data in school zones using observational methods to inform future work.

Recommendation. (9) FAV test plans must account for school zones.

4 Conclusions

While FAVs may engender numerous safety benefits that result from following traffic regulations and maintaining consistent attention to the roadway, their low capacity to cope with uncertainty complicates their ability to operate in school zones. Their deployment will undoubtedly require further innovation along with changes to local policies governing travel in school zones.

It is important to note that engineering, infrastructure, education, and enforcement issues are inherently intertwined. Further evolution of AV technology may influence changes in infrastructure which, in turn, will result in a need to update the education materials. Similarly, school infrastructure that facilitates student safety may also mandate design requirements for FAVs. In an ideal world, schools would have standardized layouts and orderly on-site drop-off and pick-up processes at well-marked locations. But the reality is chaotic. Streets get congested, other vehicles get delayed, and school buses and parents wishing to drop off children are slowed. The solution will likely be a compromise requiring open communication between local transportation programs, their affected communities, and companies developing FAVs that accounts for all the interconnected elements of the transportation system, from street, to sidewalk, to building, with considerations for stakeholders of all ages. It is likely that there will not be a one-size-fits-all solution that will apply to every jurisdiction. School transportation officials will have to find their individual approach, making use of a mix of autonomous driving models and traffic control strategies to find an ideal solution. It is important that these stakeholders are included in conversations about future deployment.

This work is part of a broader effort to assess the effects of AVs operating along school routes and within school zones. It considers FAVs operating around K-8 school zones, with an emphasis on safety. Additional work is required to assess the implications of vehicles with intermediate levels of automation in which a driver shares driving responsibilities with automation, high school students learning to drive, connected vehicle technologies that electronically link vehicles to infrastructure, and interactions with human-driven and/or automated school buses.

Acknowledgements. The Author thanks the ITRE and HSRC staff as well as the members of the TRB School Transportation Subcommittee who took the time to participate in the exploratory interviews that informed this work. This material is based upon work supported by the Federal Highway Administration and the National Highway Traffic Safety Administration under Cooperative Agreement No. DTFH61-16-H-00029. Any opinions, findings, and conclusions or recommendations expressed in this publication are solely those of the Author and do not necessarily reflect the view of the Federal Highway Administration or the National Highway Traffic Safety Administration.

References

1. Singh, S.: Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. Traffic Safety Facts Crash Stats. Report No. DOT HS 812 115. National Highway Traffic Safety Administration (2015)

2. Governors Highway Safety Association: Pedestrian Traffic Fatalities by State: 2018 Preliminary Data (2019)
3. McGuckin, N., Fucci, A.: A Summary of Travel Trends: 2017 National Household Travel Survey FHWA-PL-18-019 Washington, DC: Federal Highway Administration (2018)
4. Chang, D.: National pedestrian crash report. National Center for Statistics (2008)
5. Yanagisawa, M., Swanson, E., Najm, W.: Target crashes and safety benefits estimation method-ology for pedestrian crash avoidance/mitigation systems. National Highway Traffic Safety Administration (2014)
6. Fuest, T., Sorokin, L., Bellem, H., Bengler, K.: Taxonomy of traffic situations for the interaction between automated vehicles and human road users. In: International Conference on Applied Human Factors & Ergonomics, pp. 708–719. Springer, Cham (2017)
7. Gerónimo, D., López, A.M., Sappa, A.D., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1239–1258 (2010)
8. Merat, N., Louw, T., Madigan, R., Dziennus, M., Schieben, A.: What externally presented information do VRUs require when interacting with fully automated road transport systems in shared spaces? Accid. Anal. Prev. **118**, 244–252 (2018)
9. Rothenbücher, D., Li, J., Sirkin, D., Mok, B., Ju, W.: A field study investigating the interaction between pedestrians and driverless vehicles. In: Proceedings of the International Symposium on Robot and Human Interactive Communication. IEEE (2016)
10. Schieben, A., Wilbrink, M., Kettwich, C., Madigan, R., Louw, T., Merat, N.: Designing the interaction of automated vehicles with other traffic participants: design considerations based on human needs and expectations. Cogn. Technol. Work **21**, 69–85 (2018)
11. Lavalette, B., Tijus, C., Poitrenaud, S., Leproux, C., Bergeron, J., Thouez, J.: Pedestrian crossing decision-making: a situational and behavioral approach. Saf. Sci. **47**, 1248–1253 (2009)
12. National Highway Traffic Safety Administration. National Center for Statistics and Analysis. Traffic Safety Facts 2004: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System (2005)
13. U.S. Department of Transportation, Federal Highway Administration: Manual on Uniform Traffic Control Devices for Streets and Highways (Revision 2) (2012)
14. Hubsmith, D., Ping, R., Gutowsky, N.: Safe Routes to School: 2007 State of the states report. Fairfax: Safe Routes to School National Partnership (SRTSNP) (2007)
15. National Center for Safe Routes to School: How children get to school: School travel patterns from 1969 to 2009 (2011)
16. McDonald, N., Brown, A., Marchetti, L., Pedroso, M.: U.S. school travel 2009: an assessment of trends. Am. J. Prev. Med. **41**(2), 146–151 (2011)
17. Zhang, S., Benenson, R., Omran, M., Hosang, J., Schiele, B.: Towards reaching human performance in pedestrian detection. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 973–986 (2018)
18. Caltech Pedestrian Detection Benchmark (2017). www.vision.caltech.edu/Image_Datasets/Caltech-Pedestrians/
19. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: an evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 743–761 (2012)
20. Sandt, L., Owens, J.: A Discussion Guide for Automated and Connected Vehicles, Pedestrians, and Bicyclists Pedestrian and Bicycle Information Center, Chapel Hill, NC (2017)
21. U.S. Department of Transportation: Federal Automated Vehicles Policy (2016)
22. U.S. Department of Transportation: Automated Vehicles 3.0: Preparing for the Future of Transportation (2018)

23. Myers, M.D., Newman, M.: The qualitative interview in IS research: examining the craft. Inf. Organ. **17**(1), 2–26 (2007)
24. American Association of State Highway and Transportation Officials: Chapter 3: Elements of design. Policy on geometric design of highways and streets. Washington, DC (2001)
25. Rosen, E., Sander, U.: Pedestrian fatality risk as a function of car impact speed. Accid. Anal. Prev. **41**, 536–542 (2009)
26. National SAFE KIDS Campaign: Child Pedestrians at Risk in America: A National Survey of Speeding in School Zones (2000)
27. Essex, A., Shinkle, D., Teigen, A.: Transportation review: trends in State Speed Legislation. In: National Conference of State Legislators (2017)
28. Safe Routes to School Online Guide: Education (2015)
29. National SAFE KIDS Campaign: Kids at Crossroads: A National Survey of Physical Environment and Motorist Behavior at Intersections in School Zones (2004)
30. National SAFE KIDS Campaign: Stop Sign Violations Put Child Pedestrians at Risk: A National Survey of Motorist Behavior at Stop Signs in School Zones & Residential Areas (2003)
31. Hunter, W.W., Stutts, J.C., Pein, W.E., Cox, C.L.: Pedestrian and bicycle crash types of the early 1990's. Federal Highway Administration. Report FHWA-RD-95-193. U.S. Department of Transportation, Washington DC (1995)
32. Tapiro, H., Meir, A., Parmet, Y., Oron-Gilad, T.: Visual search strategies of child-pedestrians in road crossing tasks. In: Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society – Europe (2013)
33. Percer, J.: Child pedestrian safety education: applying learning and developmental theories to develop safe street-crossing behaviors. National Highway Traffic Safety Administration Report HS 811 190. U.S. Department of Transportation, Washington, DC (2009)
34. Meir, A., Parmet, Y., Oron-Gilad, T.: Towards understanding child pedestrians' hazard perception abilities in a mixed reality dynamic environment. Transp. Res. Part F: Traffic Psychol. Behav. **20**, 90–107 (2013)
35. Cross, D.S., Hall, M.R.: Child pedestrian safety: the role of behavioral science. Med. J. Aust. **182**(7), 317–318 (2005)
36. Schieber, R.A., Vegega, M.E. (eds.): Reducing childhood pedestrian injuries: summary of a multidisciplinary conference. Inj. Prev. **8**(Suppl. 1), i1–i10 (2002)
37. Clamann, M., Aubert, M., Cummings, M.L.: Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles. Paper Presented at the Transportation Research Board 96th Annual Meeting, Washington DC, United States (2017)

The Role of Context and Perception of Road Rules in the Pedestrian Crossing Decision: A Challenge for the Autonomous Vehicle

Marie-Axelle Granié[1(✉)], Florent Varet[2,3], Béatrice Degraeve[4], and Achot Khalafian[1]

[1] Univ. Lyon, IFSTTAR, TS2, LESCOT, 69675 Lyon, France
{marie-axelle.granie, achot.khalafian}@ifsttar.fr
[2] IFSTTAR, TS2, LMA, 13300 Salon de Provence, France
florent.varet@ifsttar.fr
[3] Aix Marseille Univ, LPS, 13100 Aix-en-Provence, France
[4] Lille Catholic University, OCeS, 59016 Lille, France
beatrice.degraeve@univ-catholille.fr

Abstract. Pedestrian behavior is based on both cognitive processes and the construction of social knowledge and representations. The results of various studies we have conducted showed that pedestrian compliance with road rules varied according to crossing regulation, built environment, gender and age. They also addressed the role of rule perception and internalization. Finally, they showed how informal learning, through observation, of road rules can explain the construction of conventional level rules, the application of which is contextual. These different elements demonstrated how knowledge of social norms can be an issue for the autonomous vehicle.

Keywords: Pedestrian · Compliance · Rule perception · Social norms

1 Introduction

The decisions we make are not always the most rational and prudent. We all know that we do not comply with all existing health, hygiene and safety rules, even if we are fully aware of them. Knowing the traffic rules is not enough to guarantee their application. To fully understand user behaviors, taking into account cognitive mechanisms is not enough. Pedestrian behavior is also based on the construction of social knowledge and representations of the roles of different users, rules and risks.

In this chapter, we propose to review studies that highlights the role of individual and social factors in pedestrian crossing decisions – and how these factors can challenge the implementation of the autonomous vehicle. Specifically, we begin by presenting the results of various studies we have conducted showing that pedestrian compliance with road rules varies according to crossing regulation, the built environment, gender and age. We then address the role of rule perception and internalization. Finally, we emphasize how informal learning, through observation, of road rules can explain the

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 221–233, 2020.
https://doi.org/10.1007/978-3-030-20040-4_20

construction of conventional level rules, with contextual application. We end by discussing how knowledge of social norms can be an issue for the autonomous vehicle.

2 Situational and Individual Variations in Compliance with the Rules

2.1 Spatial and Temporal Compliance

It is a truism to say that user behavior does not always comply with the rules of the road. However, there is a need to better understand the situations in which this lack of compliance occurs and the reasons for it. For example, in a study examining reported and actual behavior of children aged 5–6 years [1], we observed that children are aware of pedestrian rules and report compliance with them, but do not apply the same rules when they are observed on the way home from school with their parents. Knowing a rule does not necessarily mean complying with it. This compliance, which we studied especially among adults, varies in particular according to the physical environment and the perception of road rules.

In a real-life study [2], we observed 400 pedestrians (200 men and 200 women) aged 18 to 55 years at 4 intersections, including 2 without traffic lights. The results showed that spatial conformity (use of the pedestrian crossing) varies according to the configuration of the intersection: pedestrians are more compliant with the use of the pedestrian crossing in the presence of a traffic light than in its absence. Crossing regulation is most often implemented when the infrastructure is complex and/or traffic or pedestrian density is high. The presence of regulation not only acts as a reminder of the injunctive norm, but can be interpreted as a sign of a significant dangerousness of the crossing, leading to increased spatial compliance among pedestrians, whose compliance thus varies according to the road layout.

A subsequent study [3] observed pedestrian behavior at crossings regulated by pedestrian traffic lights by combining in situ observation of the behavior of 422 French adult pedestrians with questionnaires collected after the crossing. The results showed that the size of pedestrian groups and traffic density are associated with safer behaviors, while the presence of vehicles parked near the crossing site are associated with higher risk behaviors. In addition, the driving experience appears to increase safety behaviors and alertness. Pedestrians who report difficulty crossing a road look more towards the traffic light and tend to cross the road in a straight line rather than diagonally. In addition, pedestrians who have already fallen into the street - reported more by older pedestrians - look more closely at other pedestrians before and during the crossing. In terms of temporal compliance (pedestrian green light crossing), pedestrians who cross against the red pedestrian light look less towards the ground and the light, but more towards the traffic before and during the crossing than those who complied with the traffic light. They also run and jaywalk more often. It appears that pedestrians anticipate pedestrian red light violations before they arrive at the intersection to be crossed, depending on traffic conditions and not the color of the pedestrian light, as they look less towards the traffic light before crossing. These behaviors observed before and during the crossing could reveal a deliberate time-reduction objective, where the intentional violation of pedestrian red lights goes hand in hand with the vigilance of incoming traffic before and during the crossing, suggesting an awareness of the risk pedestrians take. The role of two gaze targets (traffic light and incoming traffic) should be further examined as possible indicators of an intention to violate the crossing time rule.

2.2 Variation According to Gender and Age

In addition to these variations in spatial and temporal compliance, it appears that pedestrian behavior also varies by sex. The results of Tom and Granié [2] showed differences between men and women in temporal compliance and in visual research before and during the crossing. Men's visual focuses are mainly on moving vehicles, while women's visual attention is mainly focused on other pedestrians. In addition, these results show that the visual attention of males does not vary with the crossing configuration (with or without pedestrian light) or the crossing phase and remains focused on the dynamic elements of the environment, such as moving vehicles. On the contrary, women, when crossing with traffic lights, mainly look at the pedestrian light before the crossing and at other pedestrians during the crossing. They seem to be more attentive than men do to the static elements of the scene and - in the absence of external regulation to facilitate decision-making - to the social environment rather than the physical environment. The visual centering of female participants on other pedestrians, particularly in the absence of a time rule for crossing, can be interpreted as the result of a stronger social influence and normativity on decision-making among women than among men.

In another in-situ study [4], we observed, using the same observation grid as in Tom and Granié [2], 682 pedestrians (375 women) of two age levels (middle-aged and older adults) at five sites in a major French urban center. The sites all included pedestrian crossings, in mid-block or intersection, regulated or not by traffic and pedestrian lights. The results showed that, in the situation of regulated pedestrian crossings, older people comply with the legal and safety rules more often than middle-aged adults do. They focus on a more diversified number of different visual indicators in preparation for their crossing. Their gazes are more focused on the ground, both before and during the crossing. They also stop more on the sidewalk before crossing, cross when the light is red for traffic, without running. However, the more cautious behavior of older pedestrians is accompanied by less traffic control than among younger adults, who are more likely to violate the pedestrian light rule. Behavioral control and compliance with the rules of caution seem to be accompanied by a delegation of responsibility to the regulatory system in crossing decision-making and a concomitant decrease in traffic condition monitoring.

2.3 Variation According to the Built Environment Crossed

This contextual compliance seems to be based, at least among pedestrians, on their perceptions of the crossing environment and, more particularly, on the inferences that pedestrians make from certain clues found in the environment about the pedestrian-driver balance of power.

In an exploratory study [5], we analyzed the role of the environment, understood in a broader definition including the built environment, on pedestrians' perception of the situation. We used a focus group methodology with two groups of five and six pedestrians. We had them discuss the walking pleasantness and facility of crossing 20 environments, photographed in order to give a 180° view of the road scenes. The results showed that two types of crossing environments are positively rated by the pedestrians interviewed. Some pedestrians prefer environments that are complex for the driver to deal with (wide sidewalks, lack of parking, markings and pedestrian crossings, context animated by the presence of coffee houses and shops), i.e., environments containing ambiguity and uncertainty [6], requiring increased attention of the drivers towards pedestrians. On the contrary, other pedestrians prefer environments that organize and simplify pedestrian use (presence of pedestrian crossings, well-delimited sidewalks, clear visibility, in a rather calm context, and with moderate traffic). In these environments, legibility and predictability are important, both for the driver and for the pedestrian - thus facilitating decision-making, thanks to the presence of a highly regulated space.

In a second part of the same research project, we set up an experimental study [7] whose objective was to identify the variations produced in decisions to cross, depending on different types of urban environments. This experiment also aimed to identify the environmental characteristics that pedestrians take into account and the inferences they develop and use to explain their decision to cross a road. A series of photographs presenting five different types of environments (i.e., city center, inner suburbs, public housing in the outskirts, commercial zone in the outskirts and countryside) were presented to 77 participants divided into three age groups (pre-adolescents, young and middle-aged adults). Their decision to cross or not to cross, their perception of pleasantness and safety, and the elements they considered when making a decision were collected for each environment presented. The quantitative results showed that pedestrians' perception of the pleasantness and safety of public spaces, in terms of walking, varies widely across urban environments. In addition, the decision to cross varies significantly depending on the environment. Pedestrians were much more likely to make the decision to cross the street in the city center than in the other sites presented. The qualitative analysis of the interviews showed that the presence and function of buildings, the quality of sidewalks and the presence of marked parking spaces are key factors in explaining their decision to cross, allowing them to deduce pedestrian and traffic density and vehicle speed.

In these studies too, it is clear that the rules are used flexibly by pedestrians, depending on the context, and especially on the situation in which the pedestrian and the driver find themselves. This reveals how the pedestrian, a member of a dominated and vulnerable group in the road space, finds himself in need of inferring information about the drivers, a dominant group in terms of density, potential nuisance to the pedestrian and urban development. Thus, pedestrians place great emphasis on environmental analysis, looking for clues to understand their place in the environment in the balance of power with the driver, and driver's current and predictable behavior.

We found further evidence of this estimation of power balance in the study of perceptions of crossing environments among pre-adolescents [8]. Three hundred and forty-two secondary school students aged 11 to 16 were asked about their perceptions of the level of pleasantness, safety and convenience of five crossing environments near secondary schools (i.e., the three schools from which the participants came, one from another French region and one from Quebec, Canada). Each environment was represented through five photographs, reproducing a 180° vision from the crossing point. Analysis of the comments showed that the most highly positively rated aspects of the presented environments concern the aesthetic appeal of residential housing, the impression of space emerging from the width of the lanes, as well as the large number of pedestrian crossings and recent road markings.

On the contrary, the relative anarchy of parking also gives pre-adolescents the feeling that pedestrians are not taken into account and respected by drivers. Vehicle parking and the type of built environment therefore seem to be important indicators, allowing both high school students and adult pedestrians to infer the current and future behavior of drivers [7]. Similarly, the presence of traffic lights can be negatively perceived by some pre-adolescents who perceive them as an indicator of traffic density (lights are only used when there is a need to regulate vehicle flows) and therefore of more complex and dangerous crossings. On the contrary, according to other pre-adolescents, the presence of traffic lights makes crossing easier, because it allows a delegation of decision-making to the environment, lowering the feeling of danger. This protective effect of traffic lights had already been pointed out in pre-adolescents who are not very confident about independent travel [9] and observations of senior pedestrians seem to show that the regulation of pedestrian-driver interactions by the layout is also sought when perceptual-cognitive abilities decline [3, 4].

Thus, one of the five environments presented is denigrated by all pre-adolescents, including those who frequent it on a daily basis. Participants explained this negative perception by the absence of sidewalks and the degradation of ground markings, including pedestrian crossings, which is perceived as the symbol that pedestrians are not taken into account in the environment. Inversely, the environments perceived positively by pre-adolescents have in common the strong presence of recent road markings, which, like the width of the pavements, not only mark and differentiate everyone's spaces [5] but also reinforce the real but also symbolic visibility of pedestrians.

3 The Role of Rule Perception

Compliance with traffic rules is not only subject to the local context at the time of decision-making. It also varies according to the perception of road rules themselves. Thus, perception of danger and risk associated with pedestrian behavior, normative beliefs and reported behaviors differ according to whether or not the danger to oneself is intentional.

3.1 Differences Between Errors, Transgressions and Lapses

As part of a research on the influence of educational practices on risky behavior, we interviewed 258 adult parents (104 males and 154 females), aged 25 to 54 years old, about their perception of danger, self-risk, normative beliefs and reported behaviors in response to injury-risk pedestrian behavior, using the Road User Behavior Perception Scales (EPCUR) [10]. EPCUR is composed of four subscales, each randomly presenting 16 pedestrian behaviors. This list of behaviors, based on Elliott and Baughan's tool [11] and previous research on pedestrians [10], differentiate between non-hazardous transgressions (for example, "crossing at a red pedestrian light in a street without a car") and dangerous behaviors without transgression (for example, "crossing while forgetting to look"). The first subscale (Declared Behavior Scale) measures the frequency with which the individual reports manifesting each behavior. In the second subscale (Hazard Perception Scale), the individual must assess the level of danger they perceive about each situation. In the third subscale (Risk Perception Scale), the individual must intuitively assess his or her own probability of having an accident if he or she engages in the above-mentioned behavior. The fourth subscale (Normative Belief Scale) measures individual cognitive norms about the acceptability of behavior [12, 13]. The individual must estimate the level of severity of the transgression he or she attributes to each of the behaviors presented.

Analyses of the responses at these different scales showed that individuals differentiate two groups of behaviors for each scale: (1) self-endangerment behaviors, including lack of visual information taking, (2) offence and risk taking behaviors, including absence of temporal or spatial conformity (lights or pedestrian crossings) and errors during pedestrian-vehicle interaction. These analyses of the EPCUR scales thus revealed that pedestrian behavior, whether in terms of perception of danger and risk, normative beliefs or declared behavior, is differentiated by adults on a single dimension: intentionality or not of the endangerment. Moreover, it appears that intentional behaviors include both offences against legal rules and diagnostic errors, concerning the estimation of intervehicular time and the influence of other pedestrians. This lack of differentiation among pedestrians between infringements and errors as shown in this study raises questions that we addressed to other studies. Using the theoretical framework used by Reason in the construction of the Driver Behavior Questionnaire [14] and the scales of aggressive [15] and positive [16] driving behaviors towards other users, we developed the Pedestrian Behavior Scale (PBS) [17, 18].

The 47 items of this scale are based on existing validated versions of the PBQ [19–21], as well as other reported pedestrian behavior scales [10, 11], and differentiate between 5 types of pedestrian behavior: 1/Violations, defined by a deviation from the legal rules of the highway code concerning pedestrian behavior (10 items); 2/Errors, defined as decision-making that places the pedestrian in danger, but without violations of the legal rules (11 items); 3/Inattentions, defined by inappropriate behaviors related to a lack of concentration on the task (8 items); 4/Civic behaviors, defined by behaviors to calm social interactions (5 items); 5/Aggressive behaviors, defined by conflicting behaviors with other users (6 items). Following Torquato and Bianchi [20], "filter" items were added (7 items).

Study 1 [17] was conducted with 276 students (175 female and 101 male) aged 26 to 40 years. Factor analyses revealed five axes: "violations or errors during the crossing", "inattention", "violations or errors during travel", "aggressive behavior", and "positive behavior". Study 2 [18] was conducted with 343 participants (217 female and 126 male) aged 15 to 78 years. Here, factor analyses revealed four axes. Factor 1 "transgression" includes legal violations and error items. Factor 2 includes "inattention" items. Factor 3 is composed of items of "aggressive behaviors" and factor 4 includes items of "positive behaviors".

The main difference between the factor structure of the DBQ and the Pedestrian Behavior Scale (PBS) is that, contrary to the DBQ for drivers, the transgression axis of the PBS includes both offence and error items. However, this result is consistent with Elliott and Baughan's study [11] on the adolescent pedestrian behaviors, whose "unsafe street crossing" dimension is composed of both error and violation behaviors. Similarly, these results can be compared to the previously cited validation study [10], which also differentiated between self-endangerment behaviors (inattention) and intentional risk-taking behaviors (error-transgression). In their perception of pedestrian rules and compliance behavior, it seems that individuals differentiate behavior according to whether or not there is prior intention (present in errors and violations, but not in inattentions) rather than on the basis of objective (consequences of the act for themselves and others) or legal (infringement or not of a traffic rule) criteria. These risk-taking behaviors (i.e., the deliberate engagement in behaviors in which accidental risk is perceived [22]), are either deviations from legal rules or from informal rules of caution. In other words, French pedestrians do not seem to clearly differentiate legal rules and cautionary rules. In contrast to pedestrians, the studies using the DBQ conducted on drivers (and particularly on French drivers [23]), all showed that drivers differentiate, within intentional risky acts, between behavior that violates legal rules and error behavior. This suggests that legal and social rules are less differentiated among pedestrians than among drivers.

3.2 Differences in Social Value Between Self-endangerment and Risk-Taking

This lack of differentiation between transgression and error in pedestrians in favor of a broader differentiation between intentional risk behavior and self-endangerment related to inattention is perceptible from adolescence onwards.

A recent study we conducted analyzed the evolution of relationships with parental and peer behaviors on a sample of 2,473 secondary school students aged 10 to 16 years, equally distributed in terms of gender and age [24]. Secondary school students were asked about their perceptions of the frequency - for themselves, their parents and their peers - of intentional or unintentional risky behaviors as pedestrians. The results showed that secondary school students systematically assess their behavior as being more often risky than they reported for their parents, and less often risky than they reported for their peers, regardless of the type of behavior. However, the results showed differences according to age, gender, but also the type of behavior mentioned. Indeed, regarding intentional risky behaviors (and unlike unintentional risky behaviors), the gap between Self and parents is larger for boys and increases with age, while the gap between Self and peers decreases with age and is greater for girls. This suggests that different logics are involved, for each type of behavior (intentional versus unintentional), depending on the way in which relationships with parental and peer norms are constructed. They also show that there is a gap between actual social behavioral norms and adolescents' perceptions of them. By gradually conforming to peer norms on intentional risk behaviors but not on self-endangerment, adolescents, particularly boys, seek to confirm themselves as individuals separated from their family group and to confirm themselves as members of their sex group. Our results among adults [10] seemed to confirm that for men: intentional risk-taking was more acceptable than unintentional endangerment. This shows that intentional risky behavior also seems to be a way of social valuing, at least for adolescent boys.

As we will see, this valuation of intentional risky behavior among men may be the consequence of the socialization process experienced by young pedestrians, who are more confronted during their acculturation on the road with the normative behaviors of their social environment than with the legal rules concerning pedestrian behavior.

4 Socialization to Pedestrian Rules and Behaviors

4.1 Informal Learning of Traffic Rules

To better understand the state of skills and forms of behavioral learning in childhood, we used the auto- and allo-confrontation method [25] with nine pre-adolescents aged 11–12 (5 girls and 4 boys) in the first year of secondary school, observed and then questioned about their behavior during their walk from home to school [9]. Our results showed first of all that journeys seem to be affected by a form of routinization based on behavioral and travel habits and a strong knowledge of the specific traffic contexts on the route taken, leading to a strong sense of control of the situation in everyday travel environments. However, children are aware that they cannot use this knowledge in unknown environments, which in turn leads to greater compliance with legal rules when the context changes. In addition, while some important pedestrian skills to handle crossing situations [26] appear to be fairly well mastered by the children in our sample (i.e., crossing site selection, environmental analysis and time gap estimation), compliance with road rules is far from systematic.

Children comply with the rules according to their perception of the constraints they involve, their knowledge of the road environment and their assessment of traffic conditions. Moreover, their observation of offences committed by pedestrians and motorists leads them to put into perspective both the validity of the rules and the need to comply with them. Their expectations of driver behavior are all the more limited because, in accordance with previous studies [27], they perceive drivers as not respecting the rules, particularly those governing their interactions with pedestrians. Thus, the road environment is often seen as hostile, insecure, unpredictable, and vehicle-dominated and the relationship with the driver is not perceived as an interaction but as a relationship of driver domination over pedestrians [28]. In this respect, these results are in line with the results previously presented [7, 8] in which pedestrians' judgments about environments are based mainly on the estimation of the pedestrian/driver balance of power. The 11-year-old pedestrians interviewed thus use their own rules, which are defensive in nature, to the detriment of legal rules. Moreover, this and other studies [29] show that, in this situation as in others, the child acquires knowledge both through direct confrontation with the situation and through observation of the behavior of others.

During this informal learning, the child can reproduce the behaviors that can be observed in experienced adult pedestrians, but not the skills that lead to their decisions. Traffic knowledge among the young pedestrians interviewed is only based on informal learning of the road space, as part of their usual daily journeys. They identify rules of action, based solely on the behavioral patterns they can observe in other pedestrians, without formal learning of the rules and cognitive skills involved in pedestrian behavior.

4.2 Representation of Traffic Rules and Compliance

The studies we have conducted show the contextual variability that children observe in the behavior of others leads them to categorize road rules in the conventional domain, defined in social domain theory as gathering rules of an arbitrary, relative, modifiable and contextual nature [9, 30]. This categorization of traffic rules at the conventional level affects the compliance of individuals, because conventional rules remain external and are not internalized (i.e., integrated into the individual's value system) [31].

In one of the studies mentioned above, conducted with 162 children (83 boys and 79 girls) in the last year of kindergarten (5–6 years) [1], we observed both the knowledge of road rules, their compliance and internalization and their behavior during the accompanied home-school journey. The rule governing the behavior is considered internalized if the individual avoids to commit this behavior even in the absence of any external disapproval [32]. The results revealed that children with higher internalization of traffic rules report more cautious and more controlled behaviors on the street.

However, internalization appears to vary with age. Our results on young adolescents aged 11 to 15 years [33] showed that, like the perception of danger, internalization is an important inhibitor of risky pedestrian behavior. However, the reported risky pedestrian behavior increases during adolescence, while the perception of danger and the level of internalization decrease [34].

5 The Challenges for the Autonomous Vehicle

All the studies presented show that pedestrian behavior at the time of crossing is not based on the traffic rules of the Highway Code. They are based on social norms in which the behaviors to be adopted vary according to the situations, gender and age of individuals, with some intentionally transgressive behaviors being valued by certain social groups. These elements of knowledge about social norms are important to take into account in current thinking about autonomous vehicles for several reasons.

First, the individual has access to these social norms regarding travel in road space before learning the legal rules of the Highway Code. The individual can observe these social norms in action early on, during his or her first travel experiences, as a pedestrian but also as a vehicle passenger. It is therefore important to understand their learning, which will guide the perception of risk and the categorization of legal rules and thus the individual's behavior and compliance with legal rules, throughout life.

Secondly, it must be taken into account that these social norms govern the actual functioning of road space, more than legal rules. These social norms (what the subject may define as “normal” behavior to adopt when travelling in an urban environment) are not necessarily in line with legal norms. The more the social norms to which users refer and comply differ from legal norms, the more difficult their behavior may be to anticipate, if these social norms are not known by the users with whom they interact. Individuals must therefore acquire knowledge both of the social norms to which they must comply, but also of the social norms to which other types of users comply.

Making decisions in road space – whether as pedestrian or driver – requires anticipating the behavior of others according to the role they play in that space (for example, the use of another mode of transport) and the social rules that are linked to that role [35]. Thus, knowledge of legal norms is not sufficient to anticipate correctly the behavior of others, it is also essential to know and understand the different social norms by 1/identifying them through the behavior patterns of different types of users [29]; 2/within the same type of users, differentiating them according to certain criteria of belonging to social groups: women, men, young people, the elderly, motorcyclists, truck drivers, drivers of powerful cars, etc. [9]. These social categorizations will allow a classification of observed behaviors that will facilitate the anticipation of other users’ behaviors, depending on the situation, their roles in the urban space, and refined by taking into account their membership of a certain social group [36]. As a result, expectations of others are not based on legal norms, but on the social norms corresponding to their social group and mode of travel. Knowledge of social norms is thus essential to understand and anticipate the behavior of others during interactions between users in the road space. These elements should be considered in the ongoing discussion on the autonomous vehicle in all industrialized countries.

This knowledge can also be taken into account in the necessary discussions on the pedestrian behavior models that will have to be implemented in the autonomous vehicle, so that it correctly anticipates the behavior of the pedestrians with whom it is likely to interact. The research presented in this chapter showed that pedestrian behavior is strongly related to context (presence or not of regulated crossings, built environments of crossing locations) and individual variables (gender, age, but also time pressure). They also show that pedestrians modify their behavior according to the balance of power with the driver, which they consider more or less favorable depending on the situation. The autonomous vehicle must therefore take into account the fact that the pedestrian’s behavior towards it may vary according to the spaces crossed but also according to individual characteristics.

They can be taken into account in the discussions on the behavior models that must be implemented in the autonomous vehicle so that its behavior is understandable and predictable by users of the different transport modes. Using social norms in the model of driver behavior would make the behavior of the autonomous vehicle more understandable and easier to anticipate by other drivers but also by pedestrians. For interactions with other vehicles, this consideration of social standards should at least be necessary on a transitional basis, pending a complete homogenization of the fleet of vehicles on the road. It will also be necessary to secure interactions with users of other modes of travel that will remain un-automated. This requires a correct understanding of the content of the social norms used by drivers (which may vary according to culture), interacting with other drivers but also with other types of users. This could be studied through observations in real or virtual situations, but also through surveys of reported behavior, including gaps between actual behavior and legal rules (i.e., contextualized rules for the use of different modes of transport).

Moreover, since the arrival of the autonomous vehicle is highly publicized, pedestrians may already have built a model of the behavior of the autonomous vehicle, of which they can anticipate a more rational behavior, more compliant with the rules and more attentive to their presence than a vehicle driven by a human. Pedestrians can already attribute a more cautious behavior to the autonomous vehicle than that of other vehicles and adapt their crossing behavior to these beliefs, thinking, for example, that they will always be perceived by the autonomous vehicle, which will not suffer, like humans, from attentional blindness in certain situations. These representations of the autonomous vehicle among pedestrians will have to be studied, as they may ultimately put the pedestrian at risk.

References

1. Granié, M.A.: Gender differences in preschool children's declared and behavioral compliance with pedestrian rules. Transp. Res. Part F: Traffic Psychol. Behav. **10**, 371–382 (2007)
2. Tom, A., Granié, M.-A.: Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads. Accid. Anal. Prev. **43**, 1794–1801 (2011)
3. Dommes, A., Granié, M.-A., Cloutier, M.S., Coquelet, C., Huguenin-Richard, F.: Red light violations by adult pedestrians and other safety-related behaviors at signalized crosswalks. Accid. Anal. Prev. **80**, 67–75 (2015)
4. Granié, M.-A., Dommes, A., Cloutier, M.-S., Coquelet, C., Huguenin-Richard, F.: Etude des effets de l'âge et du contexte de traversée de rue sur les comportements observés sur passages piétons régulés. In: Cloutier, M.S. (ed.) La ville sous nos pieds: connaissances et pratiques favorables aux mobilités piétonnes, pp. 275–284. Institut National de la Recherche Scientifique - Centre Urbanisation et Société, Montréal (Canada) (2014)
5. Granié, M.-A., Montel, M.-C., Brenac, T., Coquelet, C., Millot, M., Monti, F., Pannetier, M.: Qualitative analysis of pedestrians' perception of the urban environment when crossing streets. Advances in Transportation Studies, vol. XXXI, pp. 17–34 (2013)
6. Engwicht, D.: Intrigue & Uncertainty. Towards New Traffic-Taming Tools. Version 2.1. Creative Communities International, Brisbane, Autralia (2003)
7. Granié, M.-A., Brenac, T., Montel, M.C., Millot, M., Coquelet, C.: Influence of built environment on pedestrian's crossing decision. Accid. Anal. Prev. **67**, 75–85 (2014)
8. Granié, M.-A.: Perceptions des environnements de marche connus et inconnus chez des collégiens piétons en France. In: Huguenin-Richard, F. (ed.) Place aux piéton.nes, pp. à paraître. L'Harmattan, Paris (2019)
9. Granié, M.-A., Espiau, G.: Etude qualitative du comportement piéton de collégiens par la méthode de l'autoconfrontation. Territoires en Mouvement. Revue de Géographie et d'Aménagement 2008, pp. 39–57 (2010)
10. Granié, M.-A.: Influence de l'adhésion aux stéréotypes de sexe sur la perception des comportements piétons chez l'adulte. Recherche - Transports - Sécurité **101**, 253–264 (2008)
11. Elliott, M.A., Baughan, C.J.: Developing a self-report method for investigating adolescent road user behavior. Transp. Res. Part F: Traffic Psychol. Behav. **7**, 373–393 (2004)

12. Guerra, N.G., Huesmann, L.R., Hanish, L.: The role of normative beliefs in children's social behavior. In: Eisenberg, N. (ed.) Review of Personality and Social Psychology. Social Development, vol. 15, pp. 140–158. Sage Thousand Oaks, CA (1995)
13. Huesmann, L.R., Guerra, N.G.: Children's normative beliefs about aggression and aggressive behavior. J. Pers. Soc. Psychol. **72**, 408–419 (1997)
14. Reason, J.T., Manstead, A.S.R., Stradling, S., Baxter, J.S., Campbell, K.: Errors and violations on the roads: a real distinction? Ergonomics **33**, 1315–1332 (1990)
15. Lawton, R., Parker, D., Stradling, S.G., Manstead, A.S.R.: Predicting road traffic accidents: the role of social deviance and violations. Br. J. Psychol. **88**, 249–262 (1997)
16. Özkan, T., Lajunen, T.: A new addition to DBQ: Positive Driver Behaviours Scale. Transp. Res. Part F: Traffic Psychol. Behav. **8**, 355–368 (2005)
17. Granié, M.-A., Pannetier, M., Guého, L.: Validation française d'une Echelle de Comportements Piétons. In: Granié, M.A., Auberlet, J.M., Dommes, A., Serre, T. (eds.) Qualité et sécurité du déplacement piéton: facteurs, enjeux et nouvelles actions, pp. 289–298. Les collections de l'IFSTTAR, Paris (2012)
18. Granié, M.-A., Pannetier, M., Guého, L.: Developing a self-reporting method to measure pedestrian behaviors at all ages. Accid. Anal. Prev. **50**, 830–839 (2013)
19. Moyano Diaz, E.: Teoria del Comportamiento Planificado e intencion de infringir normas de transito en peatones. Estudos des Psicologia **2**, 335–348 (1997)
20. Torquato, R.J., Bianchi, A.S.A.: Comportamento de Risco do Pedestre ao Atraversaar a Rua: Um Estudo com Universitarios. Transporte: Teoria e Aplicaçao, vol. 2, pp. 19–41 (2010)
21. Yildirim, Z.: Religiousness, conservatism and their relationship with traffic behaviours. Department of Psychology, vol. Master of Sciences. Middle East Technical University, Ankara (2007)
22. Saad, F.: Prise de risque ou non perception du danger. Recherche - Transports - Sécurité septembre, pp. 55–62 (1988)
23. Guého, L., Granié, M.-A., Abric, J.C.: French validation of a new version of the driver behaviour questionnaire. Accid. Anal. Prev. **63**, 41–48 (2014)
24. Granié, M.-A., Apostolidis, T.: Effet de l'âge et du sexe sur la relation aux normes des parents et des pairs: étude sur des adolescents piétons. 7èmes entretiens francophones de psychologie. Symposium "L'espace routier: un lieu d'étude de la socialisation pour la psychologie sociale et la psychologie du développement", Université de Lille 3 (2017)
25. Mollo, V., Falzon, P.: Auto- and allo-confrontation as tools for reflective activities. Appl. Ergon. **35**, 531–540 (2004)
26. Thomson, J.A.: Promoting pedestrian skill development in young children. In: Durkin, K., Schaffer, R. (eds.) The Wiley Handbook of Developmental Psychology in Practice, pp. 311–340. Wiley, Londres (2016)
27. Platt, C.V., Clayton, A.B., Pringle, S.M., Butler, G., Colgan, M.A.: Road safety education for children transferring from primary to secondary school. Road Safety Research Report no. 35. Department for transport (2003)
28. Granié, M.-A., Varet, F., Torres, J.: Les trajets à pied comme temps et objets de socialisation chez les collégiens français. Le Sujet dans la Cité **1**, 73–86 (2018)
29. Granié, M.-A.: La construction des règles comportementales sur le port de la ceinture chez l'enfant: analyse du contenu d'entretiens auprès d'enfants de 5 et 8 ans. Recherche Transports - Sécurité **83**, 99–114 (2004)
30. Granié, M.-A.: Socialisation au risque et construction sociale des comportements de l'enfant piéton: éléments de réflexion pour l'éducation routière. Enfances, Familles, Générations **12**, 88–110 (2010)

31. Grolnick, W.S., Deci, E.L., Ryan, R.M.: Internalization within the family: the self-determination theory perspective. In: Grusec, J.E., Kuczynski, L. (eds.) Parenting and children's internalization of values, pp. 135–161. Wiley, New York (1997)
32. Turiel, E.: The development of morality. In: Eisenberg, N. (ed.) Handbook of Child Psychology. Social, Emotional and Personality Development, vol. 3, pp. 863–932. Wiley, New York (1998)
33. Granié, M.A.: Sex differences, effects of sex-stereotype conformity, age and internalization on risk-taking among pedestrian adolescents. Saf. Sci. **47**, 1277–1283 (2009)
34. Nucci, L., Guerra, N., Lee, J.: Adolescent judgments of the personal, prudential, and normative aspects of drug usage. Dev. Psychol. **27**, 841–848 (1991)
35. Foot, H.C., Thomson, J.A., Tolmie, A.K., Whelan, K., Morrison, S., Sarvary, P.: Children's understanding of drivers' intentions. Br. J. Dev. Psychol. **24**, 681–700 (2006)
36. Mundutéguy, C., Darses, F.: Perception et anticipation du comportement d'autrui en situation de conduite automobile. Le Travail Humain **70**, 1–32 (2007)

Towards the Development of a Universal Testing Environment for Attention Guiding Techniques

Max Bernhagen(✉), Philipp Hein, André Dettmann, and Angelika C. Bullinger

Chair for Ergonomics and Innovation, Chemnitz University of Technology, Erfenschlagerstr. 73, 09125 Chemnitz, Germany
{max.bernhagen,philipp.hein,andre.dettmann, angelika.bullinger-hoffmann}@mb.tu-chemnitz.de

Abstract. Contexts as diverse as control rooms, driving and picking in logistic systems provide the backdrop to the potential of attention guiding to improve human performance. Due to recent technological progress in the field of head-mounted displays for augmented reality, attention guiding in mobile and context-sensitive situations has become widely applicable. Accordingly, the gap in existent literature concerning design, development and in particular evaluation and comparison of attention guiding techniques is surprising. Our literature reviewed showed that most studies were conducted within specific use cases and accordingly were using a specific technique. In order to start building knowledge on the comparison of techniques, we review the literature and run a focus group. From both, we derive pointers for the design and development of a universal testing environment to evaluate attention guiding techniques. We build a prototype with selected evaluation methods and test both the environment and two combined attention guiding techniques.

Keywords: Attention guiding · Augmented reality · Head-mounted display

1 Introduction

Augmented Reality (AR) devices were destined to be the next technological step for enhancing human performance in industrial applications, but today's real-world usage is still negligible. Devices providing AR content need to be mobile, connected and context-sensitive, so that they can help users by guiding them to target objects, supplying task-specific information, controlling and automatically registering each action taken. For industrial applications like manufacturing, maintenance or order-picking, head-worn devices offer the hands-free operation, workers need. Until recently these head-worn devices, classified as head-mounted displays (HMDs), were not commercially available, and were often only built for research purposes or not sufficiently mobile [1, 2]. This changed with the introduction of the Google Glass and especially the Microsoft HoloLens, a consumer-ready, self-sustained AR-HMD. The HoloLens with its big set of integrated, out of the box features like spatial perception and gesture recognition provides better access to the technology and its advantages. As mentioned

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 234–243, 2020.
https://doi.org/10.1007/978-3-030-20040-4_21

above, many industrial tasks profit from hands-free devices. Many tasks contain search or navigation components, which can be improved by implementing attention guiding techniques supporting the worker. In this context, augmented reality head-mounted displays are a promising technology because people can operate them remotely. State-of-the-art HMDs provide the important benefit of stereoscopic depth cues. As a result, many use cases address picking, assembly or maintenance tasks. Here, guiding the user's attention is crucial to quickly find specific objects. Especially for industrial applications, time reduction is an important part of companies' technological endeavors. From an ergonomic viewpoint, the users' physical stress and mental demand should always be considered when optimizing work tasks. As a result, researchers already investigated several designs for attention guiding systems, mainly focusing on single aid techniques.

In our ongoing research towards developing multi-aid attention guiding techniques with a separation between near and far navigation, we discovered that techniques are usually developed and tested for specific use cases. A universal testing environment that allows evaluating and comparing techniques, does not exist. This is a problem, as this hinders the adaptability for comparing other techniques in the same environment. Comparing techniques is crucial for improving existing and developing new techniques. Solely Renner and Pfeiffer [3] create a testing environment suitable for comparing techniques. Their approach is limited to the use case of assembly with only rotational user movement. In this work, we analyze existing attention guiding techniques with regard to their use cases and experimental setups. We further identify requirements and influences of testing environments for attention guiding techniques and propose a concept for a standardized testing environment to compare single- and multi-aid techniques.

2 Related Work

While related literature is rare, we present seminal papers which provide the foundation for the first environment design. Grogorick, Stengel, Eisemann and Magnor [4] present a method for gaze guiding in a 360° virtual environment. They differentiate between a static and a dynamic scenario. The static scenario describes stationary use cases where only little head movement is necessary. Use cases that require dynamic user movement are represented by the dynamic scenario. Figure 1 shows the virtual testing setup. The setup consists of 200 objects with four different geometric shapes, which are placed on a 270° spherical segment around the user. The objects have randomized colors, orientations and positions, but have the same size. The distance between objects and user varies between two and three meters. The procedure Grogorick et al. used to evaluate their gaze guiding technique was a search task, where users had to find a specific sphere within the array of distractor objects. Participants were placed in a swivel chair, which only allowed rotatory motion. Testing environment and gaze guiding method were visualized by a modified HTC Vive head-mounted display.

Fig. 1. Virtual testing setup for evaluating a gaze guiding method [4]

Another approach to attention guiding is the omnidirectional attention funnel [5]. By the help of a funnel, which is decreasing in size towards the target, users will see the shortest path to a singular target. To test the technique a 360° the authors created a workspace as the experimental setup. The setup consists of four tables and shelves and twelve objects (six abstract objects and six everyday objects) on each table/shelve combination. Participants had to find and touch 24 objects within the test environment. Objective measurements for search time and error, as well as the measurement for mental workload using the NASA TLX was taken (Fig. 2).

Fig. 2. Outer perspective of the omnidirectional attention funnel (left [5]) and SWAVE technique (right [3]) implemented in the testing environment

Renner and Pfeiffer [3] compared different attention guiding designs for manual assembly tasks. They simulate an augmented reality device with the help of a virtual reality headset. Participants of a user study had to assemble a virtual birdhouse, using components in a given order. The components were placed on a virtual desk and shelf around the participants. Renner and Pfeiffer compared their own SWAVE technique to images indicating the next component, the omnidirectional attention funnel, a 2D arrow and a simple display. The user study investigated task completion time, head movement as well as the subjective assessment of the speed, accuracy, learnability and usefulness of the different techniques (Fig. 3).

Henderson and Feiner [6] analyzed attention guiding for complex maintenance tasks. By the help of a head mounted display, mechanics were given assistance for repair sequences. Graphical illustrations and text information were provided as well as arrows shown to guide the attention to the components. Mechanics were placed inside a military vehicle to complete 18 maintenance steps. Tasks were placed in different places inside the vehicle, requiring high degrees of head movement. Task completion time, head movement, error rate and subjective ratings for ease of use, satisfaction level, and intuitiveness were recorded.

Fig. 3. Mechanic inside a military vehicle for maintenance (left [6]); Experimental warehouse for order picking (right [7])

Using attention guiding in a logistics environment was analyzed by Schwerdtfeger et al. [7]. Their work differentiates between coarse and fine navigation. Finding the right shelf in the warehouse is considered coarse navigation. Locating a specific box in the shelf is considered fine navigation. For coarse navigation, the shelf number is shown on a monocular AR-HMD. Fine navigation is supported by a two-dimensional circular adaption of the omnidirectional attention funnel in combination with highlighting the box label. The testing environment for a user study was an experimental warehouse consisting of two aisles with shelves on either side. The procedure was an order picking process, which consists of finding the right shelf and selecting the target object. The picking process requires translational and rotational user movement. Measurements were taken for picking time, error rate as well as subjective strain and constraints by the visualization.

Current related work already created several designs for attention guiding techniques which improve user performance during search and navigation tasks. The differentiation between different navigation modes was only implemented by two navigation concepts. Henderson and Feiner [6] intuitively combined 2D and 3D arrows, but no comparison between single aid and combined methods has been made. Combining techniques that were developed for specific navigation modes provides the possibility to further increase performance. Considering aids for far navigation, unobtrusiveness is an important feature. As per our definition, far navigation encompasses longer distances and in turn longer navigation times. Especially in industrial work environments, navigation paths can contain safety concerns, if the user's view is partially obstructed by augmented reality content. Aids suited for far navigation are arrow, wedge and halo techniques. They offer sufficient navigation clues for the less precise far navigation, without occupying the user's field of view. Near navigation requires less body motion and aids can take up more visual space to allow precise guiding. Attention funnel and SWAVE are suitable techniques to support near navigation. The combination of techniques requires a switching mechanism that chooses the appropriate aid depending on the navigation mode. The mechanism therefore needs information about the user's motion and checkpoints in the navigation path.

3 Towards the Development of a Universal Testing Environment

For the development of a universal testing environment, important characteristics on the design of this testing environment need to be gathered from literature and experts. Following, use cases can be distinguished by these characteristics and the testing environment should be able to evaluate most common use cases. Using the results of a focus group with six usability experts, we identify four main characteristics for the design of the universal testing environment in order to imitate real-world use cases. Experts are asked to name characteristics as well as present and discuss them. Afterwards, participants are instructed to find high-ranked links and cluster them (for summary of the results, cf. Fig. 4).

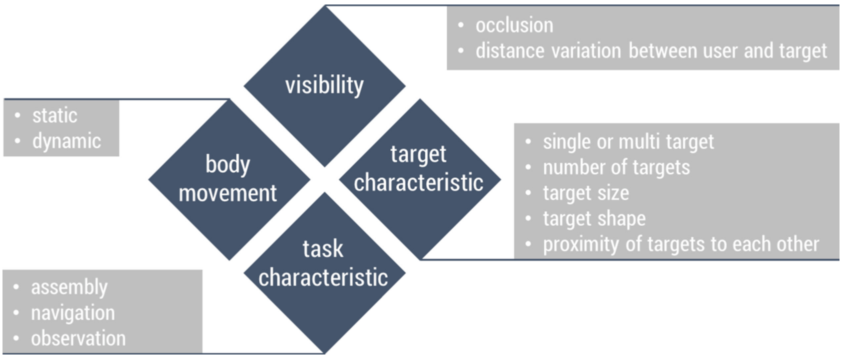


Fig. 4. Main characteristics for the design of the universal testing environment

Visibility copes with different viewing conditions. Therefore, the factors occlusions and large distances will increase the time needed to find targets. Examples in application areas are picking tasks where targets may be occluded or on the other side control room guiding without any occlusion on a screen. Target characteristic had the most mentions, as the user may be guided towards single or multiple targets, targets may vary in size, shape, frequency of appearance or the total number. Another characteristic is body movement. On the one hand, the testing environment can either imitate static scenarios then the body stands still and only rotational (head) movements occur. Corresponding use cases are control room observations, seated assembly tasks or in-vehicle attention guiding. On the other hand, dynamic body movement refers to scenarios like attention guiding in order picking tasks or assembly tasks at the assembly line. Tasks characteristics refer to possible use cases and the need to utilize different qualitative or quantitative evaluation methods.

Following the workshop participants' argumentation, body movement, visibility and target characteristic are factors concerning the design of the virtual environment. Task characteristics is important for choosing suitable evaluation methods. For example, studies investigating performance in control rooms or vehicles may focus on time to fixation [9]. Evaluation of order picking tasks in busy factories or navigation in open environments on the other hand require information about mistakes [10], movement directions [11], picking time [12] or total assembly time [8]. Most common data used for all task characteristics and therefore very basic are head rotation, time to target and the NASA TLX questionnaire.

In order to create a complete summary of recommended evaluation methods as well as suggested environment designs for the task characteristics, one has to validate distinct variants. As mentioned above, the analysis of the performance of attention guiding techniques, which differs between near and far navigation, will be the first use case. Here, we chose a static scenario with good visibility and relatively complex target characteristic. The resulting environment is shown in Fig. 5 and described in the following paragraph.

Fig. 5. Subject in the augmented testing environment

4 Prototype and Evaluation

Based on the presented testing environment, we created a complex scenario in order to evaluate effects of combining different attention guiding techniques for near and far navigation. We configured our testing environment (see Fig. 5) to enable testing of use cases that require static rotational user motion. The target objects are white colored cubes in three different sizes, placed on a cylindrical plane around the user. The cubes served as targets and distractors simultaneously. One cube at a time was marked as the target object by a red frame (see Fig. 6). We choose the attention funnel (AF) and SWAVE (SWAVE) techniques for modification in our study. Both methods offer precise near navigation at the cost of obstructing the field of view. For the far navigation component, we implemented a simple arrow navigation (AF+/SWAVE+). The concept of arrows is widely used in navigation tasks and thus easily understandable for users. Our differentiation mechanism switched from the far to the near navigation aid, when the user's line of sight was within 60° of the target vector. We implemented the environment in the Microsoft HoloLens via the Unity development platform and the HoloLens clicker was used as the input device.

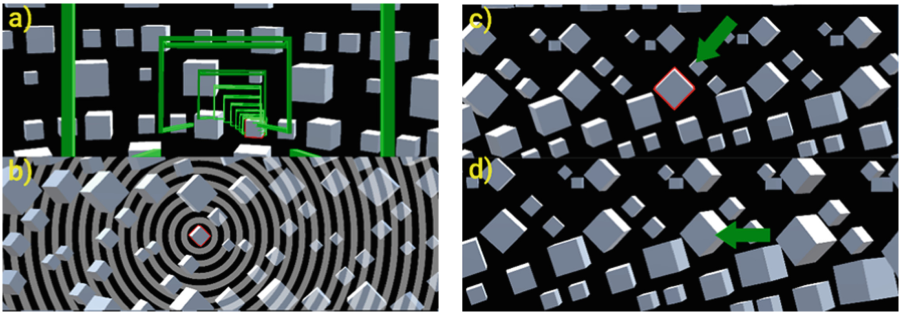


Fig. 6. Attention guiding techniques used in the preliminary evaluation: (a) Attention Funnel (b) SWAVE (c) Arrow (d) 2D-arrow used as far navigation aid in AF+, SW+ and AR techniques

The user study we conducted was set up in the style of Renner and Pfeiffer [3]. Participants had to successively find and select 25 cubes within the testing environment. Besides the combined methods attention funnel plus (AF+) and SWAVE plus (SW+), the baseline techniques attention funnel (AF), SWAVE (SW) and arrow (AR) were the treatments in the within-subjects study design (Fig. 6). To ascertain the efficiency of user guidance, the total rotation the user performs, and the time needed to find the target have been recorded, alongside the most efficient rotation the user could have performed. In total 25 able-bodied participants aged from 19 to 35 years, with a mean age of 27.24 (SD = 4.33; 10 female; 15 male) were recruited. All participants passed the TNO stereopsis test [13], ensuring their ability to correctly view the 3D images presented by the HoloLens. All participants came from an academic background and scored a mean rating of 4.5 (SD = 0.99) out of 6 on the ATI-scale, demonstrating a high affinity for interaction with technology [14]. Participants received no compensation.

Findings for search time (Fig. 7) and rotation overhead (Fig. 8) show, that mean values were reduced by adding the switching mechanism to AF and SW techniques. A repeated measures ANOVA [$F(2.72, 65.38) = 4.55$, $p = .007$)] shows a statistically significant reduction for search time for the AF and AF+ conditions. The repeated measure ANOVA for the rotation overhead [$F(2.71, 65.06) = 6.50$, $p = .001$] shows no significant difference of mean values for the AF/AF+ and SW/SW+ comparisons.

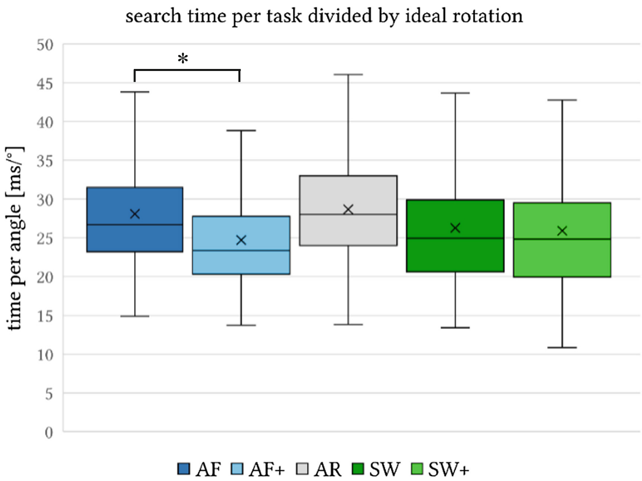


Fig. 7. Boxplots of the normalized search times for the five attention guiding techniques. The "x" mark represents the mean value (* = significant differences with $\alpha = 0.05$).

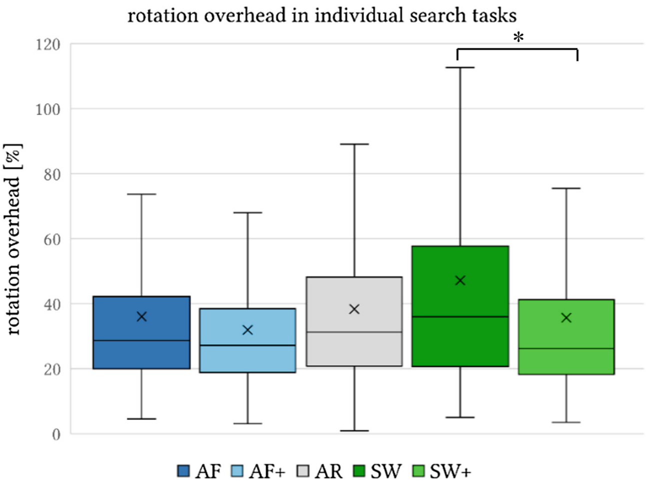


Fig. 8. Boxplots of the normalized rotation overhead for the five attention guiding techniques. The "x" mark represents the mean value (* = significant differences with $\alpha = 0.05$).

5 Discussion and Outlook

The target of this paper was to build knowledge on a universal testing environment to compare different attention guiding techniques with each other. We conducted a literature review and an expert workshop in order to find characteristics for the environment as well as evaluation methods. We presented both, design and development of a first prototype of the platform and performed a study on this platform. Study showed that platform is suitable to research attention guiding techniques and provide useful data for the comparison with regard to the characteristics. A missing aspect for answering the research question is the uncaptured rotation speed per technique. Here, it would be helpful to differ between near and far navigation speed and compare these values within the attention guiding techniques. In terms of the differentiation between near and far navigation, preliminary results indicate that the differentiation reduces search time and rotational motion of users in comparison to singe aid techniques. In a following step the testing environment we presented will be extended to allow the research of more complex navigation patterns, with translational and rotational motions. Another important research question that has been omitted in our study and should be addressed in future work is the influence of target characteristics (e.g. size, number and proximity) on the performance of the techniques. Following studies should not only differ the target but also examine a standardized amount of targets, shapes and colors to vary in order to produce different complexity of the targets. As a result, the platform should be considered as modular in terms of investigating several use cases within one tool. Thus researcher can vary the above mentioned influences to imitate the real-world scenarios and derive the best attention guiding technique for their purpose. The corresponding software will contain basic attention guiding techniques and influence characteristics.

Besides the suggested evaluation methods for attention guiding, the existence of attentional blindness is of growing interest. Distractions from hazards are a huge threat, especially for the usage at the crowded job floor. Therefore, the platform should integrate the evaluation method for investigating, whether effects of attentional blindness or other effects of visual distraction occur. In terms of developing attention guiding techniques with the concept of differentiation of near and far navigation first results seem to be promising. However, results need to be considered in other standardized virtual environments in order to get a comprehensive assessment of their potential.

References

1. Kishishita, N., et al.: Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. In: 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE (2014)
2. Sukan, M., Elvezio, C., Oda, O., Feiner, S., Tversky, B.: ParaFrustum: visualization techniques for guiding a user to a constrained set of viewing positions and orientations. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST 2014), pp. 331–340. ACM, New York (2014). https://doi.org/10.1145/2642918.2647417
3. Renner, P., Pfeiffer, T.: Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In: 2017 IEEE Symposium on 3D User Interfaces (3DUI). IEEE (2017)
4. Grogorick, S., et al.: Subtle gaze guidance for immersive environments. In: Proceedings of the ACM Symposium on Applied Perception. ACM (2017)

5. Biocca, F., et al.: Attention funnel: omnidirectional 3D cursor for mobile augmented reality platforms. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM (2006)
6. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE Trans. Vis. Comput. Graph. **17**(10), 1355–1368 (2011)
7. Schwerdtfeger, B., et al.: Pick-by-vision: a first stress test. In: 2009 8th IEEE International Symposium on Mixed and Augmented Reality. IEEE (2009)
8. Tang, A., et al.: Comparative effectiveness of augmented reality in object assembly. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM (2003)
9. Schwarz, F., Fastenmeier, W.: Augmented reality warnings in vehicles: effects of modality and specificity on effectiveness. Accid. Anal. Prev. **101**, 55–66 (2017)
10. Tonnis, M., Klinker, G.: Effective control of a car driver's attention for visual and acoustic guidance towards the direction of imminent dangers. In: 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality. IEEE (2006)
11. Lubos, P., Bruder, G., Steinicke, F.: Analysis of direct selection in head-mounted display environments. In: 2014 IEEE Symposium on 3D User Interfaces (3DUI). IEEE (2014)
12. Hanson, R., Falkenström, W., Miettinen, M.: Augmented reality as a means of conveying picking information in kit preparation for mixed-model assembly. Comput. Ind. Eng. **113**, 570–575 (2017)
13. Okuda, F.C., Apt, L., Wanter, B.S.: Evaluation of the TNO random-dot stereogram test. Am. Orthoptic J. **27**(1), 124–130 (1977)
14. Franke, T., Attig, C., Wessel, D.: A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. Int. J. Hum.-Comput. Interact. **35**(6), 456–467 (2019)

Specificities of Organisational Communication in European Universities' Digitalised Environments

Olga Burukina[1,2](✉)

[1] The University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland
obur@mail.ru
[2] The Institute of Legislation and Comparative Law Under the Government of the Russian Federation, B. Cheremushkinskaya Street 34, 117218 Moscow, Russia

Abstract. In the contemporary digital era, with universities being at the core of the global changes, boosting and leading them, issues of university digitalisation and organisational communication are topics of particular interest, as well as the presence of artificial intelligence in their policies and strategies and communicating it both to their employees and students. The presented research focuses on issues of university digitalisation, in particular when it comes to organisational communication practices. The relevance of the topic is stipulated by the need to study the existing practices in organisational communication and approaches to the establishment and development of digitalisation policies and strategies and analyse the current practices in the chosen Russian and Finnish universities as part of their digitalisation processes aimed at ensuring their competitiveness in the globalised world.

Keywords: University digitalised environment · Digitalisation processes · Organisational communication · Digitalisation strategy · Staff involvement

1 Introduction

In the increasingly globalised and hyperconnected world, every organisation striving for successful development needs a communication strategy, which at its core relies on ongoing organisational conversations. Therefore, it is important for organisation leaders to empower their staff and support them with the right tools and systems to ensure employees can not only communicate effectively, but constantly evolve and get better by improving both internal and external organisational communications [1].

The preliminary results of the pilot research reported in this paper have revealed a variety of attitudes of administrators and academics to the increasingly digitalized university environment – from wholehearted welcoming to 'cool toleration' at Finnish universities and even opposition to some aspects of the newly introduced digital environment at some departments of Russian universities. The research has pinpointed a number of push and pull factors encouraging academics to and discouraging them from welcoming certain aspects of their universities' digitalised environment, with some of

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 244–256, 2020.
https://doi.org/10.1007/978-3-030-20040-4_22

them revealing far deeper reasons of their cool or opposed attitudes to unsystematic introduction of digital technologies at both Finnish and Russian universities.

2 Digitalised University Environment

A modern university cannot be imagined without a well-developed digital environment comprising digitalised work places and digital learning environments, information and knowledge management and sufficient regular digital training.

University digitalisation when based on systemic approach inevitably becomes an integral component of the university sustainability plan. Though the idea of digitalisation-sustainability convergence in organisations and society challenges every stakeholder, it gives them unseen chances within and across organisational boundaries [2].

2.1 Deeper Engagement and Higher Productivity

A well-elaborated digital environment can improve university staff's engagement, agility and productivity [3]. Digitalisation can turn the university into a free and flexible organisation, drastically decreasing time spent on implementation of work tasks, raising its employees' productivity, contributing to greater integration of university divisions and tighter collaboration of its staff members, making the operations transparent, seamless and faster, and boosting the employees' professional development.

2.2 Profound Sustainability

Sustainability has gradually turned into an all-too-familiar theme, almost a byword mainly used correctly, yet now and then abused. More and more European organisations including universities speak of sustainability as an integral part of their vision.

In the opinion of Mervyn King, ex-governor of the Bank of England, Sustainability is the primary moral and economic imperative of the 21st century. It is one of the most important sources of both opportunities and risks for businesses. Nature, society, business are interconnected in ways that should be decoded by decision-makers [2].

Besides alignment of decentralised campuses and setting clear goals and priorities in the areas of campus operations, emissions and energy, ecosystems and nature, university sustainability plans should focus on the issues of health and well-being, culture and learning, – all concerning their employees, students and local communities.

Besides, the idea of sustainability is inseparable from the idea of social responsibility, which in the university's case is threefold – before the employees, the local community and the national society as every university is responsible for its staff and its graduates along with their contribution to the local and national development.

3 Universities' Organisational Communication

Both digitalisation processes and organizational communication cannot be fully successful without a systemic approach that, when applied properly, guarantees identification and inclusion of all the most important components and taking into account every stakeholder with their real needs, aims and challenges. Therefore, systemic approach provides solid ground for elaboration of an effective system of university digitalisation and consecutive systemic changes in organizational communication.

3.1 Digitalisation Strategy

A digitalisation strategy can be defined as both a form and means of strategic management and an adequate response to the challenges of the continuing digital disruption, which has primary importance for key industries and fundamental fields of human activities including education, with higher education being in its vanguard. However, without specifying university vision (according to its specifics), goals, strengths and weaknesses, opportunities and threats, along with a scope of tactics, i.e. related activities aimed at maximizing the benefits provided to the university by the external digital environment and internal digital initiatives.

Universities need a well-elaborated digital strategy to advance their digital transformation. In the digital era, the digital strategy is a foundation of the organisation's sustainable development and its effective communication, both internal and external. Basing on a well-elaborated strategy, the university can develop its own digital roadmap and a portfolio, set priorities and start changing the digital environment [4].

3.2 University Specificity

Quite a number of universities have elaborated digital strategies, some others, including the University of Vaasa are working on it. Nevertheless, the very understanding of the need for a digital strategy proves the university's strive for sustainable development. Both the process of a digital strategy elaboration and the process of its implementation involve communication with senior managers to coordinate digital projects with the university's development as an integral organisation [5]. However, taking into account the specific character of the university (the diverse character of the main staff's occupation – teaching and researching and the utter dependence of the whole university on them), it is highly advisable to use their active position and professional imitativeness and involve academics and researchers in the process of a strategy elaboration and implementation for the benefit of the whole university.

4 Research Outline

The presented pilot research was designed to both revealing the most important issues of university digitalisation in Europe basing on analysis of previous research results and discourse and analysing the ongoing digitalisation processes at Finnish and Russian universities, which seem to have been facing serious challenges.

The scope of methods used in the pilot research includes a questionnaire-based survey aimed at identifying the current problems faced by academics and administrators in the new university digital environment, case studies of four universities in question – a Finnish and a Russian universities, cross-case analysis of two universities in the international perspective, content analysis of the participants' responses, and documental analysis of the universities' digitalisation policies and websites.

4.1 Research Hypotheses

The undertaken pilot research based on two hypotheses:

Hypothesis 1 – the Finnish respondents are more deeply involved in the digitalisation processes going at the University of Vaasa and are more willing to contribute to its development.

Hypothesis 2 – the Finnish respondents' attitudes to the impact of digitalisation on their work life are more positive than those of the Russian respondents, who are more sensitive to decisions habitually taken behind their backs.

The analysis of the global, Finnish and Russian research, discourse and university practices helped to pinpoint four challenges generated by the previous and current digitalisation processes in university digitalised environments. The challenges created by the increasing reliance on digital communication tools in the absence of a comprehensive university digitalisation strategy are the following:

- Digital tools and platforms keep people tethered to work all the time, which is not healthy, to say the least. Therefore, every organisation's digital environment needs clear boundaries to help the employees to unplug periodically.
- Taking into consideration the swiftly growing number of apps and other digital tools, choosing the right tools for the organisation and its workflow can be a frustrating process [1].
- At present, some organizations face a situation with some of their employees spending all of their time communicating and hardly any time actually working [6]. (This situation reveals common students' behaviour in most Russian universities.)
- Today's multigenerational workforce has vastly different rituals, routines, and expectations with regard to communication, which can create situations where people feel alienated, disrespected or just plain left out of the loop [ibid.].

All these prompt a need for well-developed and highly adaptive organizational communication strategies and policies, benefitting organizations and their employees.

4.2 Research Methods and Processes

The research methodology bases on systemic approach to the idea of university digitalisation and its research and on comparative approach to analyzing its main achievements and core challenges as comparative approach is able to highlight the sore points more vividly, as well as a pilot survey intended to identify the perceptions and attitudes of academics and university administrators and managers to the ongoing digitalisation processes and reveal their (mis)understanding of the current situations.

The designed questionnaire contained 7 background questions (Age, Gender, Occupation Type, Work experience in HE, Work experience for the current employer, Type of contract, Organisational change) and the main part comprising 4 sections, consisting of 14 groups and 70 statements, most of which were scaled according to Likert bipolar scaling, measuring their positive or negative responses from ‘Strongly agree’ to ‘Strongly disagree’ and multiple-choice questions. The sections included were titled “Digital strategies”, “Digital tools in work”, “Organisational communication”, and “Artificial Intelligence”. As the length of this article is strictly limited, only a part of the survey can be presented in the article – two sections of the four, namely “Digital strategies” and “Organisational communication” (however, partly), yet the presented results highlight several very important trends, perceptions, and attitudes.

4.3 Respondents

The survey was planned as pilot and voluntary, so the sample was only limited by the time frame and employees’ willingness to participate in it. As the pilot research was designed as comparative, with two universities engaged – the University of Vaasa, Finland and Lipetsk State Pedagogical University, Russia. The two universities were chosen due to their similarity in a number of characteristics, including their location in Finnish and Russian provinces, size and other characteristics. The University of Vaasa was established in 1968 as a public university. The amount of its academic staff totals to 327, and administrative staff – to 184 employees. The number of students totals to 5,000. The University includes four schools with a range of departments and multiple programmes: School of Management, School of Accounting and Finance, School of Marketing and Communication, and School of Technology and Innovations.

The Lipetsk State Pedagogical University was established in 1949 as a public university. The amount of its academic staff totals to 356 persons (with 288 full-timers and 68 part-timers), and administrative staff – to 150 employees. The number of students totals to 5,000, and doctoral students – to 70. Since 1.2.2016, the university has comprised six specialized institutes created on the basis of the 13 previously functioning faculties: Institute of Philology, Institute of Psychology and Education, Institute of Natural, Mathematical and Technical Sciences, Institute of History, Law and Social Sciences, Institute of Culture and Art, and Institute of Physical Culture and Sports.

The period open for the Finnish employees in their online survey was 5 weeks, with the first 2 weeks providing a majority of the participants – 33 and the other 3 weeks adding 4 more participants. The 33 Russian respondents were provided with questionnaires printed out on paper and asked to return them on the same day at a suitable time. All the 33 questionnaires returned valid for analysis.

5 Research Results

The Finnish and Russian university employees’ participation was based on voluntary and free expression of their opinions, and the fact that the 37 Finnish respondents have taken part in the survey and all the 33 Russian respondents have returned the received questionnaires proves their interest in the topic of digitalisation and digital changes in organisational communication and their willingness to be heard.

5.1 Digitalisation Strategies

The pilot survey aimed at reaching several goals, including primary identification of the thematic relevance and the level of the thematic importance for the academics and university administrators and managers, second – revealing university employees' perceptions and attitudes to the ongoing digitalisation processes and finally – highlighting the (dis)advantages of the taken decisions and interim achievements.

The background questions intended to identify the respondents' age, gender, occupation, and type of contract as this information has highlighted their relevance to the research of the universities' digital environments could explain their attitudes to the internal digitalisation processes and provided food for further research of the topic.

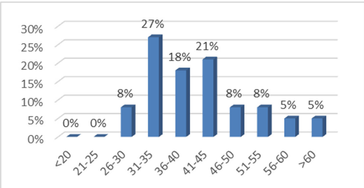


Fig. 1. Finnish respondents' age groups

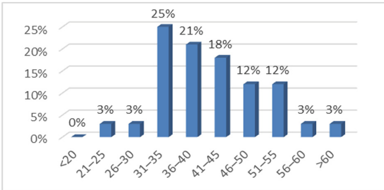


Fig. 2. Russian respondents' age groups

The histograms above show that the pool of the Russian respondents seems to be older than the Finnish respondetns, but the difference is irrelevant for the research purposes and the two clusters can be compared safely (Figs. 1 and 2).

Gender is not an important variable in the given research, though it is important for the validity of the survey (the samples were supposed to be as close to each other as possible). The gender characteristics of the two clusters are very close, too, with 62% of the Finnish respondent being female and 38% – male, and the Russian cluster consisting of 67% of female respondents and 33% – males (Fig. 3 and 4 below).

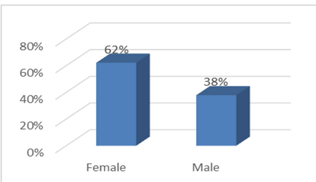


Fig. 3. Finnish respondents' gender

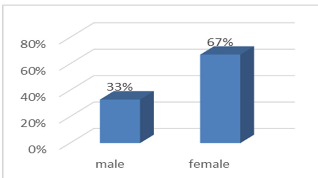


Fig. 4. Russian respondents' gender

The responses of the Finnish participants have highlighted that the majority (41%) consists of academics and researchers, 16% are researchers proper, 13% identified themselves as academics, 2% – managers, 5% – general administrators, 8% – student administrators, 10% – research services, and 5% – other categories. Similarly, the majority of the Russian respondents (66%) are academics, 17% are managers, 8% are general administrators, 6% are academics and researchers in one, and 3% belong to other staff categories. The histograms show that the occupations of the Finnish respondents are more diverse and a considerable number of the participants are focused on pure research work (16%) and teaching (13%). When clustered, the respondents can be united into two major groups – Finnish academics and researchers (70%) and administrators and managers (25%), and corresponding Russian groups (72% and 25%), which makes the two groups valid for further analysis (Figs. 5 and 6).

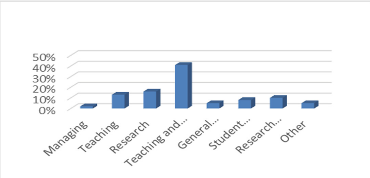


Fig. 5. Finnish respondents' main occupation

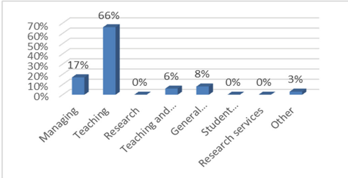


Fig. 6. Russian respondents' main occupation

The next background question concerned he respondents' types of contract because it reveled both their loyalty and involvement in the university's current issues and second, their intentions to further contribute to their employer's development. Answering the question about their type of contract, 63% of the Finnish respondents have permanent contracts, 32% have fixed term <5 years, and 5% have some other types of contract (not identified), while the responses of the Russian participants are more diverse, with 61% indicating the they have permanent contracts, 9% – fixed-term contracts >5 years, 24% – fixed term <5 years, and 6% have some other types of contract. The responses of the Russian participants do not seem to be quite valid as the majority of the respondents (66%) are academics who do not have permanent contracts nowadays at Russian universities (all academics have been gradually transferred to the fixed-term contracts of 5 years maximum). It means that most academics choosing the option 'permanent' do not fully recognize their status either due to their previously habitual status of permanent contract or because the contract prolongation in some departments of LSPU is done almost automatically unlike in many other Russian universities (Figs. 7 and 8 below).

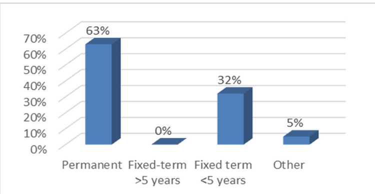


Fig. 7. Finnish respondents' type of contract

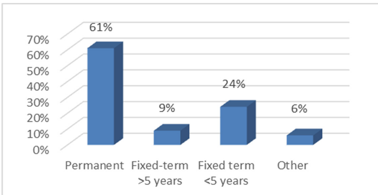


Fig. 8. Russian respondents' type of contract

The first statement *"Digitalisation is strategically managed in my university in general"* in the section *"Digitalisation strategy"* and the first group with the same title (comprising six statements) showed a higher level of confidence (76%) of the Russian respondents in the strategic management of digitalisation processes at their university: 49% agreed, 27% strongly agreed, with 9% only disagreeing with the statement and 15% being uncertain. The Finnish respondents were more cautious: though the same quantity 49% agreed, 5% only strongly agreed, with 18% disagreeing, 5% strongly disagreeing and 23% unsure about their response. These results seem to have revealed both the high level of management of both universities' administrators on the one hand and the positively disposed respondents' belief in their universities' being guided by some comprehensive strategies, which presence contributes largely to their employers' sustainability, while their absence would mean lower level of stability in the employees' lives.

The responses of both clusters of participants to the statement *"My University has a university-level digitalisation strategy"* revealed their loyalty to the employer and the high level of the university administrations' skills as almost a half (47%) of the Finnish respondents (39% agreed and 8% strongly agreed) and 2/3 of the Russian respondents (76%–55% agreed and 21% strongly agreed) expressed their confidence that both Vaasa and LSPU have digitalisation strategies. However, 13% of the Finnish respondents disagreed with this statement, 8% strongly disagreed and 32% were uncertain about is, while only 3% of the Russians disagreed and 21% were unsure.

The reaction of the Finnish respondents to the statement *"My university follows the national guidelines in digitalisation"* was unpredictable: 1/3 of the respondents only agreed with it (24% agreed and 8% strongly agreed), with an overwhelming majority being unsure (64%) about their university attitude to the national guidelines in digitalisation or about the existence of national guidelines in higher education, and 4% stating that their university didn't follow the national guidelines (2% disagreed and 2% strongly disagreed). On the contrary, an overwhelming majority (70%) of the Russian respondents stated that LSPU follows the national guidelines in digitalisation (55% agreed and 15% strongly agreed), with 27% being unsure and 3% disagreeing with the statement. The availability of the Russian national guidelines in digitalisation, particularly for HE digitalisation seems to be doubtful, so these responses most likely reflect the Russian employees' loyalty to their university and national guidelines at large.

The responses to the statement *"My university is at the forefront in digital developments"* highlighted the higher level of adequacy of the Finnish respondents' estimation of their university position in the national and global digital developments, yielding 18% of positive results only (16% agreed, 2% strongly agreed), with 47% of negative results (34% disagreed and 13% strongly disagreed) and 42% being in two minds or uncertain about their estimation. 30% of the Russian respondents expressed brighter estimations of LSPU's leading position (27% agreed and 3% strongly agreed), with 36% perceiving the reality more adequately (30% disagreed and 6% strongly disagree) and 34% were uncertain in their response.

The reaction of the Finnish respondents to the statement *"I would like to participate in drafting a new digitalisation strategy for my university"* was quite predictable – their willingness (56%) to be involved in the core projects influencing their lives is famous and understandable: 51% agreed, 5% strongly agreed, 13% did not express their willingness or unwillingness, 18% disagreed and 13% strongly disagreed. However, the reluctance of a majority of the Russian respondents (51% – 21% disagreed, 6% strongly disagreed and 24% did not express their attitude) was not totally unexpected. It can be partly explained by the current work overload of academics at Russian universities, which has negatively affected and cardinally diminished their imitativeness. Nevertheless, the high percentage of the Russian respondents willing to take part in drafting a new digitalisation strategy highlighted their active position and loyal attitude to their employer despite the current hardships, as well as their understanding of the importance of the issues and readiness to contribute to the university's advance.

Analysing the responses to the statement *"I can influence digitalisation decisions concerning my own work"*, we can see that app. the same percentage of the respondents was in two minds (Finns – 21%, Russians – 27%), but the number of positive and highly positive responses is higher among the Finnish participants (49% and 18% vs. 37% and 12%) and the number of negative and utterly negative responses is twice higher among the Russian participants (18% and 6% vs. 10% and 2%). These results, particularly the Russian respondents/negative estimation of their ability to influence digitalisation decisions concerning their work has revealed the current status of academics in Russian universities, where they can hardly influence any decisions concerning their work, most digitalisation decisions included.

The second group of statements was titled *"Knowledge of Strategy"* and comprised six statements. The responses of the Finnish participants to the first statement in this group *"I know where to find my university's digitalisation strategy"* were quite truthful: the digitalisation strategy of Vaasa is being under elaboration. That was why 34% of the Finnish respondents did not agree they knew where to find the Vaasa digitalisation strategy, 27% strongly disagreed and 13% hesitated. The positive responses of 26% of the Finns were unexpected and welcome by the digitalisation strategy elaborators who would be happy to know the strategy's exact location. The confidence of 49% of the Russians who confirmed their knowledge of the location of LSPU's digitalisation strategy (46% agreed and 3% strongly agreed) was quite unexpected, as it has not been written yet. 51% of the Russians seem close to the true state of affairs.

The responses to the next statement *"I have read my university's digitalisation strategy"* puzzled both the researcher and the administrators in charge for elaboration of digitalisation strategies at both universities. The digitalisation strategy has not been written yet in either of the universities and yet, 21% of the Finnish respondents and twice as much – 49% of the Russian participants have read their university's digitalisation strategies (16% of the Finns and 43% of the Russians agreed, and 5% and 6% straongly agreed accordingly), with 13% of the Finns and 27% of the Russians being uncertain, while 29% of the Finns and 24% of the Russian disagreed and 37% of the Finns strongly disagreed. Here again, the estimation of the Finnish respondents was more adequate and closer to the reality than that of the Russian respondents, probably due to their belief in the quality of their strategic management.

The responses to the statement *"I have participated in drafting my university's digitalisation strategy"* proved to be mostly adequate, with an overwhelming majority disagreeing (29% of the Finns and 40% of the Russians) and strongly disagreeing with it (45% of the Finns and 6% of the Russians). However, the amount of the hesitating respondents (8% of the Finns and particularly 36% of the Russians) and even more so, the equal number of the respondents supporting the allegation (13% (strongly agreed) + 5% (agreed) of the Finns and 9% + 9% of the Russians) give food for thought and prompt further research of the issue in some semi-structured interviews.

The next statement *"My University's digitalisation strategy is clear"* was supported by 16% of the Finnish respondents and twice as many Russians (33% agreed and 3% strongly agreed), with a majority in both clusters uncertain of the clarity of their strategies (53% of the Finns and 43% of the Russians) or denying the idea (13% of the Finns disagreed, 18% strongly disagreed, 21% of the Russians disagreed, too). The high estimation of the strategy's clarity by the Russians may reveal their strong belief in the quality of their university's strategy regardless of whether it exists or does not.

The statement *"I support my university's digitalisation strategy"* again revealed a very high level of loyalty of the Russian respondents: 49% of them agreed and 18% strongly agreed with this statement, while 33% hesitated and no one disagreed with it. Such unanimity in the Russian respondents' positive attitude to the digitalisation strategy, which is purely virtual and speculative, can be referred to as *'blind loyalty'*. However, 1/3 of the Finnish respondents also expressed their adherence to the non-existent strategy (21% agreed and 8% strongly agreed), with an overwhelming majority (66%) being doubtful and 5% strongly disagreeing with the idea.

The responses to the last statement in the group *"My University's digitalisation strategy needs to be updated"* was needed to reveal the true situation with the state of affairs as regards to the digitalisation strategy at both universities – the strategies as comprehensive well-elaborated documents have never existed, that was why a majority of the respondents either agreed or were uncertain in their opinions: 32% of the Finns and 36% of the Russians agreed, 13% of the Finns and 22% of the Russians strongly agreed, 48% of the Finns and 36% of the Russians were not sure of their estimation of their digitalisation strategy's need for a update, while only a small amount of responses contradicting this trend (5% of the Finns and 6% of the Russians disagreed and 2% of the Finns, 0% of the Russians strongly disagreed).

Group 3 in the "Digitalisation Strategy" section was actually a multiple-choice question *"What aims do your university decision makers pursue when developing the digital environment?"* The respondents' choice of the options was as follows (Table 1):

Table 1. Aims pursued by university decision makers developing the digital environment.

Possible aims	Finns	Russians
Support the staff with the right tools	31%	13%
Ensure effective communication	24%	21%
Ensure that employees constantly evolve	16%	12%
Increase employees' productivity	22%	25%
Increase sustainability	7%	29%

The results indicate that twice as fewer Russian respondents believe that their university decision makers strive for supporting them, however four times as many

Russian respondents believe in the decision makers' aiming at increased sustainability. The rest of the options gained approximately similar number of supporters.

The next was the question *"What is the main development focus of digitalisation at your university?"* which offered five options and revealed the following preferences (Table 2):

Table 2. The main development focus of digitalisation at the UniVaasa and LSPU.

Possible development focus	Finns	Russians
Investing in digital learning spaces	21%	8%
Developing virtual university	13%	19%
Introducing new digital tools regularly	15%	26%
Offering digital support to staff and students	32%	28%
Ensuring that staff has sufficient digital skills	19%	19%

The responses have highlighted two core differences in the situations at the Finnish and Russian universities – the digital learning space has not been developed well enough at LSPU, while the Finnish respondents expressed their dissatisfaction with the regularity of introduction of new digital tools. Both clusters of respondents agreed that the main development focus of digitalisation at their universities is offering digital support to the staff and students, so that they have sufficient digital skills.

The next multiple-choice question "What should be the main development focus of digitalisation at your university?" intended to reveal the participants' understanding of the digitalisation processes at their universities, urge their active engagement into the processes and reveal their tactical preferences. The preferred choices were as follows (Table 3):

Table 3. Expected core development focus of university digitalisation.

Required development focus	Finns	Russians
Involving the staff in taking decisions concerning university digitalisation	16%	24%
Focusing on the administrators' and academics' real needs	23%	24%
Strengthening collaboration between learners, curriculum teams, departmental heads, and support services	18%	25%
Investing in the physical spaces and places	16%	16%
Investing in the staff's digital skills	27%	11%

The Russian respondents proved to show a more active attitude to their participation in decision-making processes – 1/4, adherence to strengthening collaboration between learners, curriculum teams, departmental heads, and support services – 1/4 and more attention to the real needs of the academics and university administrators – 1/4, while the Finnish respondents paid more attention to the university investments in the staff's digital skills (almost 1/3) and the focus on the academics' and university administrators' real needs (almost 1/4).

The last question "What are the main challenges facing academics at your university?" highlighted the core problems facing the employees, with a majority of the Finnish respondents being dissatisfied with the university's digitalisation policy, 1/4 Finns and over 1/5 Russians distressed because of undetermined benefits of digital tools, 1/4 Finns and Russians dissatisfied with their insufficient ICT skills, and over 1/5 Russians displeased at the absence of technical support (Table 4):

Table 4. The main challenges facing the university academics.

Core challenges	Finns	Russians
University's unclear digitalisation policy	32%	18%
Undetermined benefits of digital tools	25%	22%
Distrust to ICT technologies	5%	12%
Insufficient ICT skills	24%	26%
Absence of technical support	14%	22%

5.2 Digitalisation Impact on Organisational Communication

The Section "Organisational Communication" comprised 10 statements divided into 2 groups, but the limits of the article allow presenting the results of the first two statements only, which characterized the general attitude of the respondents.

The statement *"Digitalisation has led to more efficient communication"* was supported by 48% of the Finnish respondents and 43% of the Russian ones (with strong support of 10% and 6% accordingly). However, 32% of the Finns and 36% of the Russians were quite dubious about their response and 10% of the Finns and 15% of the Russians did not agree to support this statement (8% of the Finns disagreed and 2% strongly disagreed, and 15% of the Russians disagreed).

The statement *"Digitalisation has led to more open communication"* was supported by a majority of the Russian respondents (52% agreed and 3% strongly agreed) and the Finnish respondents (45% agreed and 5% strongly agreed), with and almost equal share of hesitating respondents (24% of the Finns and 27% of the Russians), while 24% of the Finns disagreed with the statement and 2% strongly disagreed, with 18% of disagreeing Russians. The survey highlighted some more positive results of digitalisation at both universities but these prove to be the most important ones.

6 Conclusions and Discussion

The results of the discourse analysis supported by the survey have highlighted a scope of topical problems facing European university senior managers, academics and other employees, with the core problem being the absence of a comprehensive university digitalisation strategy at both universities. However, with a digitalisation strategy being elaborated, kept in mind and guiding the digitalisation processes at the University of Vaasa, its employees seem to enjoy a more sustainable digitalised university environment, characterised with a lower level of stress and higher level of integrity of the e-environment and its processes including organisational communication, which contribute to the Finns' more positive attitudes to the impact of digitalisation on their work life, revealed in the responses of the survey participants proving Hypothesis 2. The Finnish respondents have demonstrated a deeper involvement in their university communication and greater willingness to solve its problems, proving Hypothesis 1.

The four challenges identified above and the urgent need for a digitalisation strategy, allow setting priorities in the universities' digitalisation processes, with primary attention to be paid to elaboration of a comprehensive university digitalisation strategy to guide the current and future digitalisation projects aimed at creating a highly beneficial digitalised environment at both universities in question in order to let their academics, researcher, administrators and other employees work with the highest efficiency possible without being driven away by possible deficiencies of the current digitalised environment for the universities' and the local communities' prosperity.

References

1. Cansialosi, C.: Digital Communication in the Workplace is no Longer Optional (2016). www.forbes.com/sites/chriscancialosi/2016/08/01/digital-communication-in-the-workplace-is-no-longer-optional/amp/
2. Kiron, D., et al.: The Convergence of Digitalization and Sustainability (2018). https://sloanreview.mit.edu/article/the-convergence-of-digitalization-and-sustainability/
3. Tandon, R.: How a Truly Digital Workplace can Improve Engagement, Agility and Productivity of your Employees (2015). https://www.netsolutions.com/insights/building-pathway-towards-a-truly-digital-workplace/
4. Kube, B.: A digitalization strategy – in three steps. Lufthansa Industry Solutions (2018). https://www.lufthansa-industry-solutions.com/de-en/solutions-products/technology-consulting/a-digitalization-strategy-in-three-steps/
5. Abudi, G.: Managing communications effectively and efficiently. Paper Presented at PMI® Global Congress 2013 – North America, New Orleans, LA. Project Management Institute, Newtown Square (2013)
6. Clark, S.: Effective Communication in a Digital World (2017). www.digitalistmag.com/future-of-work/2017/12/08/effective-communication-in-digital-world-05609824/amp

A Dialogue System for Elderly People Considering Impression Formation and User's Character

Huta Takeda, Manabu Kurosawa, and Hiroshi Yajima(✉)

School of Science and Technology for Future Life, Tokyo Denki University,
5 Senju-Asahi-cho Adachi-ku, Tokyo, Japan
kurosawa@cck.dendai.ac.jp, yajima@mail.dendai.ac.jp

Abstract. In recent years, lack of communication among solitary elderly has become a big problem. In order to deal with the above situation, development of a dialogue system for solving the lack of communication of the elderly has been drawing attention in recent years. The concept of the system proposed in this paper is to induce communication in consideration of the situation of elderly people during dialogue. An agent who is friendly to the elderly conducts a conversation taking into account the feelings of the elderly, thereby enabling dialogue close to the elderly.

Keywords: Dialogue system · Elderly · Character · Impression formation · Primacy effect

1 Introduction

In recent years, Japan is entering an aging society. The proportion of elderly people aged 65 or over in Japan's total population is 27.3%. As the aging progresses, the proportion of elderly people is expected to become 33.3% in 2025. In other words, it is said that one out of every three people will become an elderly person. With the progress of the aging population, the proportion of lonely elderly people is increasing [1]. The proportion of elderly living alone in the elderly population was 4.3% for males and 11.2% for females in 1980. Thereafter, in 2001 the number increased to 13.3% for men and 21.1% for women. Senior citizens tend to be lacking in communication [1]. We conducted a survey on the frequency of conversation among elderly people aged 60 and older. We analyzed those who were elderly people living alone and who responded that the frequency of conversation (including telephone and e-mail) was less than once every two or three days. As a result, the corresponding elderly people were 28.8% for males and 22.0% for females [2].

It is said that the risk of dementia is high in elderly people with poor communication. According to Saito et al.'s research, the risk of dementia for elderly people whose frequency of conversation is less than once a week is 1.4 times that of elderly people who often talk [3].

In recent years, many chat dialogue systems for the elderly have been studied to solve the lack of communication of the elderly with aging [4], but the conventional

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 257–268, 2020.
https://doi.org/10.1007/978-3-030-20040-4_23

system has the following problems: (1) system does not take into consideration factors such as the process of communication of the elderly and the character of the elderly, and (2) Because the system is not used to the elderly, it is difficult for the elderly to talk with the system that they first met.

The aim of this research is as follows. The first aim is to propose a system that is easy for the elderly to use, which has no dialogue experience with the dialogue system. The next aim is to propose a system that conducts dialogue in consideration of elderly personality. The effect of the proposed system is as follows: (1) the system eliminates the lack of communication of the elderly, and (2) the elderly has a good impression on the characters in the proposed system through conversation. Due to the effect of (2) above, higher elderly people can obtain higher degree of dialog satisfaction.

Experiment confirmed the effectiveness of “gentleness presentation function within conversation”.

2 Related Research

Dialogue system research can be divided into two types. (1) a task - oriented dialogue system on the premise of achieving a goal, (2) a non - task - oriented dialogue system aimed at talking itself [5]. In recent years, a dialogue system research to solve the lack of communication among elderly people was proposed [4]. This system aims at conversation itself such as chatting and counseling for the elderly. Therefore, this system is classified as a non-task-oriented dialog system.

The chat dialogue system for the elderly generally considers the characteristics of elderly people. Many of these systems are studied based on the reminiscence method [6] which is psychotherapy [4]. The reminiscence is a psychotherapy advocated for American psychiatrist Robert Butler in the 1960s. An important element of the reminiscence method is that elderly persons will share memories of the past and others through speaking; the mental state of the elderly is stable It is well known that the retrospective method is effective for communication between the elderly and the system. For this reason, many chatting systems for the elderly listen to the remarks of the elderly, and adopt a system of striking the remarks of the elderly (in this paper, the interactive system is referred to as a listening interactive system). In the listening interactive system, the response of the system consists of questions about elderly people’s remarks or striking to the elderly’s remarks. In other words, in the listening interactive system, the conversation between the elderly and the system progresses mainly by the elderly.

3 Issues of Listening Interactive System for the Elderly

As mentioned in Sect. 2, many researches on interactive dialogue are in chat dialogue system research for elderly people. In developing a dialogue system with elderly, we found there are the following three problems.

Issues A: The process of communication is not considered in designing interactive dialogue system. In the case of dialogue between humans, there is a clear difference between the dialogue with the partner for the first time and the dialogue with a friend. Also in the dialogue between the dialogue system and the elderly, considering the process of communication is important. Also, it is important for the dialogue system to use the optimal dialogue method in each process to realize natural communication. However, in the conventional research, communication process is hardly considered in construction of dialogue system.

Issues B: It is difficult for the elderly to talk to the dialogue system for the first time. In the case of an interactive dialogue system, the response of the system is composed primarily of incompatibility and questions. Therefore, it is necessary for the elderly to speak to the dialogue system by themselves, but most elderly people do not have experience to dialogue with the dialogue system. For this reason, when the elderly speaks for the first time with the dialogue system, there may be a case of embarrassment on the part of the elderly, or elderly can not recognize the dialogue system as a partner according to the character of the elderly. It is difficult for the elderly to speak to the dialogue system autonomously when the initial dialogue with the elderly and the dialogue system has begun.

Issues C: There are elderly persons with a personality whose dialogue system is difficult to talk to. For elderly people who prefer to talk about themselves, interactive methods of listening interactive systems are effective. However, there are some elderly people who do not like to talk about themselves. The conventional interactive dialogue system without considering the character of such elderly persons, is doubtful in effectiveness for all elderly people if the system uniformly respond to elderly.

4 Proposed Method

4.1 Proposal Concept

In this paper, we propose a dialogue system which makes conversation close to the elderly, in order to solve the problem of current dialogue system for the elderly. Specifically, we propose a system that promotes conversation with emphasis on the feelings of elderly people. This proposal incorporates a consideration of the communication process and a psychological point of view such as the character of the elderly.

4.2 Outline of the Proposed System

In this section, we outline the proposed system to solve the three issues mentioned in Sect. 3.

In order to solve the issues A, the proposed system divides the communication process into the following two stages (Fig. 1).

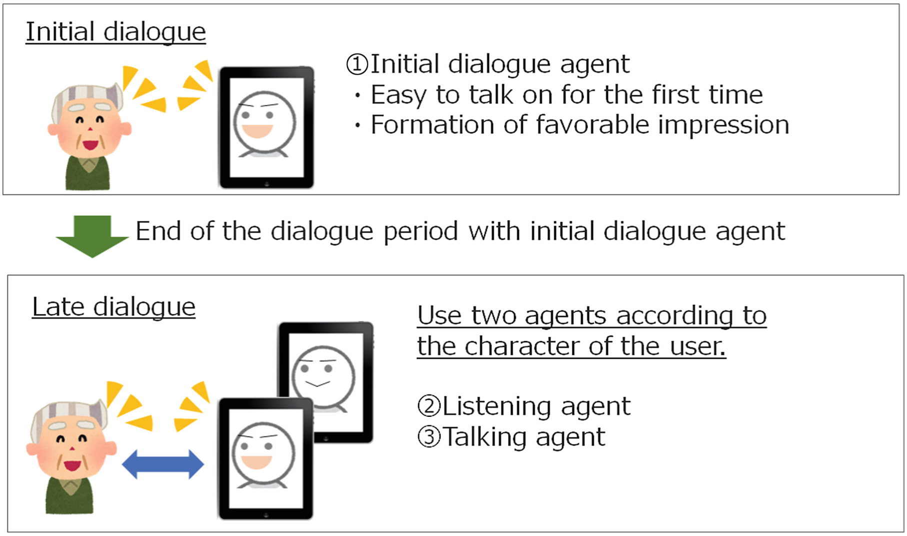


Fig. 1. Image of the proposed system

(1) "Initial Dialogue" phase for the elderly to become accustomed to dialogue with the information system, (2) "middle-late dialogue" phase for the elderly to enjoy conversation with the system.

We developed a dialogue method tailored to the characteristics of these dialogue phases.

In order to solve the issue B, we developed the "initial dialog agent" used in the initial dialogue. "Initial dialog agent" was developed considering the characteristic that elderly people are difficult to get used to the new system. The initial dialog agent adopted a dialogue method that makes it easier for elderly people to use the system for the first time by using the psychological knowledge of the elderly. In particular, the function of forming a favorable impression on the agent of the elderly is an important factor.

In order to solve task C, we developed "agents that take into consideration the character of the elderly" used in the middle and late dialogue.

Specifically developed agents are the following two types.

(1) "listening agent" which listens attentively to the elderly's story. (2) "Spoken agent" who conducts dialogue on its own initiative.

In the dialogue of the middle and latter periods, the proposed system provides dialogues of a form that matches the character of the elderly by separately using these agents according to the character of the elderly.

5 Detail of the Proposed Method

5.1 Basic Policy of Initial Dialog Agent Development

The purpose of the initial dialog agent is to provide an easy-to-talk environment for elderly people who are not accustomed to talk with the dialogue system. Therefore, in the early stage dialogue, basically, agent-based episode talk (hereinafter referred to as scenario response) is the main subject.

In the scenario response, the proposed system basically advances the dialogue in accordance with the scenario prepared in advance, so that the elderly will talk to the topic provided by the dialogue system and answer the questions. In this way, dialogue will proceed. Through this initial dialogue, even elderly people who are not used to dialogue with dialogue system can easily adapt to speak with dialogue system.

As an agent in the dialogue system, we adopted "children attending elementary school" because elderly people are naturally familiar with children. The scenario response has the following structure. First of all, the topic in dialogue was an event in school. For elderly, the timings of dialogue with agent are twice in the morning and night in the day. In the morning dialogue, the agent makes simple greetings. In the evening dialogue the agent speaks to the elderly episodes about events at school.

5.2 Change Method of Agent Scenario Response Using Elderly Input

Basically, the scenario response of the initial dialog agent does not depend on the input contents of the elderly. However, if the input content of the elderly can be assumed in advance, the dialogue system changes the response content of the agent according to the input contents of the elderly.

For example, if the agent asks "Do you eat breakfast yet?", Three patterns of "Eaten", "Do not eat", "Other answers" are considered as input intention of the elderly. In the case where the person's input intention is "eaten", it can be assumed that "eaten" or "ye" etc. are answered by elderly people. The input content in the case where the input intention of the elderly is "not eating" is also assumed by the agent in the same way.

In this case, the agent's response corresponding to the case where the supposed input content matches the input contents of the actual elderly person is set in advance. With this setting, the agent can respond to the input contents of the elderly.

In the case where the input content of the elderly does not exactly correspond to the input content assumed in advance, the dialogue system classifies the input intention of the elderly into "other responses". Then, in that case, a reply with no sense of incompatibility is made regardless of the input contents of the elderly. For example, in the case where breakfast is the topic, agent reply "The doctor said that it is important to eat breakfast" reply.

In addition, the dialogue system not only uses the input contents of the elderly for the immediately following reply as described above, but also applies to the dialogue of the following day.

For example, if the agent makes a question saying "Have you eaten breakfast yet?", and the elderly answer is judged to be "not eating", the dialogue system will record the responses of the elderly. Then, in the next morning, agent ask elderly “You seem to have not eaten breakfast yesterday, but did you eat today?”.

By making such a response, the agent gives the elderly people the impression that the agent understands and remembers the elderly’s talk answer.

5.3 Gentleness Expression Function

In order to realize improvement of the degree of satisfaction of dialogue which is one of the objects of this research, ingenuity to improve the impression that the elderly holds against the agent is indispensable. Therefore, the initial dialog agent has a function to express gentleness, which is based on psychological knowledge.

For the impression formation, process of interpersonal recognition is studied in the field of social psychology [7]. Studies on interpersonal perception have revealed that there is a phenomenon called initial effect. Due to this phenomenon, the information presented at the initial stage of impression formation is found to have greater significance for the formation of impressions than the information given at a later time [7]. Various characteristics exist as personality characteristics of a human being as an element of impression formation. Among them, it has been clarified that characteristics having a central function exist in forming impressions such as “warmth”, “cold”, etc. [7] (Since the expression “warm-hearted” is an abstract expression, in this paper we express “warm” by replacing “warm” with almost synonymous “gentle”).

To summarize such findings, it can be said that a favorable impression is easily formed in the subsequent impression formation by giving the impression “gentle” at the beginning of impression formation.

Based on such knowledge, in order to improve the impression that the elderly has against the agent, we made the initial dialogue agent handling the initial dialogue function to express “gentleness”. “Gentleness” in this paper is defined as “compassion for others”. Example of gentleness expression is to apply care-of words to others who are in trouble or sadness or make actions of gentleness.

As the method of Initial dialog agent, there are two ways of expressing the gentleness expressed by the agent. “(1) expression of gentleness to the elderly” and “(2) expression of gentleness to other than the elderly”. In (1), The agent responds such as worrying words when the senior citizens are bad. In (2), the agent expresses gentleness by telling the elderly about the behavior to others other than the elderly such as “I helped a friend”.

6 Agents Considering the Character of the Elderly

6.1 The Purpose of the Dialogue Agent in the Latter Period of Dialogue

As shown in Sect. 4.2, the proposed system uses different dialogue methods in the early dialogue and the mid-latter term dialogue. In the middle-and-later term dialogue, the dialogue system changes the dialogue method according to the character of the elderly. The aim of this conversion is to give elderly people a higher level of dialog satisfaction.

6.2 Character Classification of the Elderly

There are various methods to classify the personality of elderly people, but in this paper "aggressiveness to dialogue" is used for character classification of the elderly. The grounds for this selection is made by the following two considerations.

1. Other Personality Characteristics Classification affects emotional expression and change of ending in agent response.
2. On the other hand, in the case of "aggressiveness to dialogue", the response of the agent affects the interactive form and the amount of dialogue Therefore, influence on dialogue is strong using this method.

Specifically, using the utterance tendency scale [8] which is a psychological measure, we classify the character of the elderly into "speaking liking" and "discreetness".

When using the proposed system, questionnaires are given to elderly subjects before system use, and senior citizens are classified.

6.3 Agent Classification

We classified the elderly into two in terms of aggressiveness towards dialogue. In other words, we developed two types of agents in consideration of compatibility with individual elderly people: concretely, "listening agent" who makes dialogue with elderly people who are "talkative" and "Talking agent" that makes dialogue with "conservative" elderly people.

6.4 Listening Agent

The listening agent takes an interactive form suitable for the elderly with talkative character. Specifically, the listening agent performs utterance mainly about repetition and striking a match (see Fig. 2).

The flow of listening agent dialogue is shown below. (1) The elderly talks to the agent first (2) The agent responds. (3) After the agent 's response, there are sometimes times when elderly people cannot find the topic of an appropriate conversation.

At that time, when a certain period of time elapses, the agent automatically issues new questions related to the topic of the past to the elderly.

(4) If there is a remark from the elderly and the previous system response is a new question, agent asks deep question such as "Why is that?". (5) Next, when the content of the elderly's speech is a question, the agent responds in an ambiguous manner to the question, and then a new question is subsequently made by agent. The reason why the agent responds ambiguously is that many questions from the elderly tend to be irrelevant to the current topic. As an agent, it is difficult to respond to everything to the remarks of elderly people irrelevant to topics.

Next, when the elderly's input matches a response rule including a specific word such as "I like AA", agent performs a parrots such as "You like AA". In addition, if the input of the elderly does not include a noun, etc., the agent makes a new question related to the topic in order to broaden the talk, and in the scenes where the input of the elderly does not match these rules. The agent strikes a simple counterfeit, for example, “I see, is not it?”.

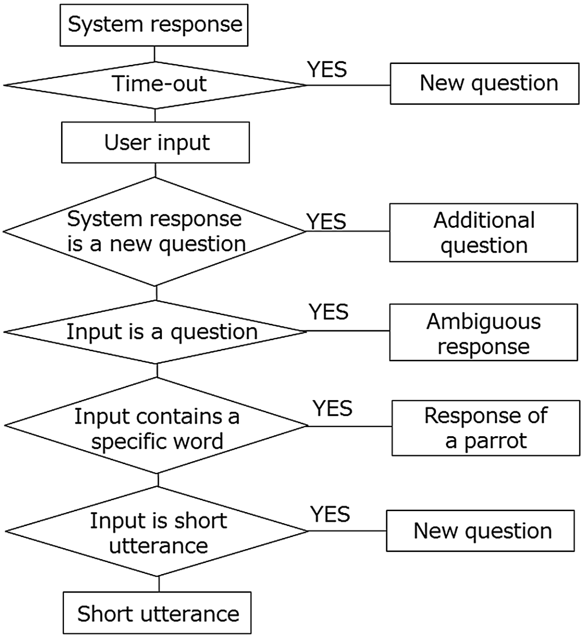


Fig. 2. Dialogue flow of listening agent

6.5 Spoken Agent

The speaking agent takes an interactive form suitable for elderly persons of modest character. Therefore, responses of agents are basically centered on scenario-like responses that do not depend on the input contents of elderly people.

As a dialog flow, firstly, the speaking agent has processing of ambiguous response to the timer processing and the question like the listening agent. As a function other than that, the speaking agent makes an utterance concerning the topic according to the scenario.

The scenario has a branch point. If the elderly ‘s response to the agent’ s question matches the rule prepared in advance at that branch point, the agent responds to express empathy, self - disclosure, etc., and changes the topic.

For example, when talking about food as a topic, agent asks question “What is the summer food?” in scenario. And for this question, system assumes the elderly’s reply such as “I like shaved ice”. And if that word is included in the actual reply, system will self-disclose such empathy as “I also like shaved ice” and “I like the strawberry taste”. Also, Change the topic from “food” to “shaved ice”.

7 Configuration of the Proposed System

This chapter describes the proposed system configuration. The proposed system uses MMD Agent [9] from Nagoya Institute of Technology. MMD Agent is a tool kit for spoke n dialogue system integrating speech recognition, speech synthesis, 3D model display. The feature of MMD Agent is that description of response rules in dialogue and change of model can be easily performed. However, MMD Agent alone cannot describe complicated response rules required to realize the proposed system. For this reason, the proposed system uses MMD Agent as an interface. And we developed our own dialogue control system using complicated response rules in cooperation with MMD Agent.

The dialogue control system consists of dialogue control system, agent system ("initial dialogue agent", "speaking agent", and "listening agent") and dialogue history DB. In addition, the dialogue control system was configured to perform morphological analysis on the input of the elderly using the Japanese morphological analysis system MeCab [10]. Japanese morphological analysis is used because part-of-speech information is used for agent response rules.

8 Experiment

8.1 Purpose of Experiment

The dialogue system proposed in this paper covers two kinds of dialogue between the elderly and dialogue system (agent), that is, the initial dialogue and the middle and latter term dialogue. In this proposal, different dialogue methods are used for two kinds of dialogue. In this paper, we conducted experiments to verify the effectiveness of the initial dialogue agent used for the initial dialogue among the two types of dialogue. In the experiment, in particular, we verified that the "tenderness expression function" of the initial dialog agent influences the impression formation for the agent from the viewpoint of the elderly.

8.2 System Used for Experiment

In the proposed system used for experiments, an initial dialog agent is used. For the 3D model used for the agent, we selected, from among the free materials on the Internet, what matches the appearance of elementary school student who is the setting of this agent. For the motion during the dialogue, we selected the appropriate one from the motion files included in the MMD Agent sample and applied it according to the dialogue contents. The synthesized speech of the agent used the voice file included in the sample of MMD Agent. Furthermore, the height of the voice was adjusted to a young impression that fits the image of the elementary school student's low grade.

In Sect. 5.3, we showed two ways of expressing ways of expressing the gentleness used by the initial dialog agent: (1) expression of gentleness to the elderly, and (2) expression of gentleness to other than the elderly. In this experiment, only the method of (2) was used, as described in detail in Sect. 8.4. Basically, the subjects of this experiment are students, and the utterance of an agent who is concerned with physical condition by the method of (1) is inappropriate.

8.3 Comparison System

The purpose of the experiment is to verify the effectiveness of the gentleness expression function possessed by the initial dialog agent. For this reason, we used an agent replacing the episode expressing the gentleness of the initial dialog agent with a general episode as the comparison system.

8.4 Experimental Procedure

Eight people in their twenties who never used a dialogue system were taken as subjects as the first step to verify the effectiveness of the system. Subjects were divided into groups A and B. Group A used the proposed system and group B used the comparative system. The experimental period is 5 days and we conduct daily dialogue between morning and night. The average number of utterances in one dialog is 5 and the average dialog time is 1 min 55 s.

8.5 Evaluation Scale and Evaluation Method

We conducted a questionnaire on the impression of the dialogue system for the subjects after the end of the dialogue on the first day night and the fifth day night of the experiment. The questionnaire was created independently by asking psychology experts. Each questionnaire item adopted a method of evaluating the impression of the subjects in five stages. The effectiveness of the proposed system is verified by comparing questionnaire results for each group and correlation analysis.

Table 1 Analysis of the questionnaire

Question items	Median of evaluation		Correlation coefficient
	A	B	
1: The agent is gentle	5	3	
2: Conversation is fun	4.5	4	0.548
3: Conversation can be relaxed	4	3.5	0.287
4: The agent is friendly	5	4	0.523
5: The agent understands a dialogue	3	3	0.333
6: Interaction is easy to talk	3.5	3	0.738
7: I'd like to talk to the agent more	4	4.5	0.076
8: The agent can be trusted	3.5	3	0.491
9: I am attached to the agent	4.5	4.5	0.746

8.6 Experimental Results

The results of the questionnaire were tabulated for each group (see Table 1). Questionnaires were also analyzed by the following procedure.

1. Impression analysis on the tenderness of agent response First, by using the gentleness expression function, it is necessary to verify whether the subject evaluated the character of the agent as gentle. We compared the effectiveness of gentleness expressions in groups A and B by "1" item of Table 1 (Subject thinks the agent is a gentle personality). As a result, the evaluation value of group A using the proposed system was median value 5, and the evaluation value of group B using comparison system was median value 3. From this result, by using the gentleness expression function, we found that subject could feel agent more gentle.
2. Correlation analysis of gentleness and other impression formation Rank correlation analysis was carried out to verify the correlation between the impression of the agent and the impression of other factors. The definition and calculation method of the two variables used in the analysis are shown below.

[Variable 1]: Evaluation of the gentleness that subject felt during the dialogue period.

[Calculation method]: With respect to the evaluation of question 1, the average value (value by subject) on the first day and the fifth day.

[Variable 2]: Improvement degree of each impression evaluation during the experiment period [Calculation method]: For questions other than question 1, the difference between day 1 and day 5 (value by subject).

In general, correlation coefficients are evaluated in the following categories. [0.7 or higher: fairly strong correlation], [0.7 to 0.4: strong correlation], [0.4 to 0.2: slightly correlated], [0.2 or less: almost no correlation].

As a result of the questionnaire, it was found that questions 6 and 9 had a strong correlation, and questions 2, 4 and 8 had a strong correlation.

In addition, at the end of the experiment, we conducted a questionnaire asking five degrees of satisfaction with conversation when "listening agent" and "speaking agent" were used. As a result, when the initial agent was added, the evaluation was 4.5, whereas when not using the initial agent, there was a big difference of 3.0.

9 Discussion

As a result of the impression evaluation questionnaire, the proposed system was gained high overall evaluation. From the interview after the questionnaire it became clear that the subject recognized the agent as a gentle character in the initial dialogue. Based on this recognition, it is considered that favorable impressions were formed as a whole dialogue in subjects. Average score for the questionnaires was 4.5 at proposed method, and 2.5 at the conventional method. This result shows the effect of the proposed method.

Even in the experiment, as a result of the correlation analysis, it was found that giving the impression that the agent is a gentle character at the beginning of the dialogue has a strong correlation with ease of talking and attachment formation.

The index of ease of talking is a very important index for users without dialogue experience with the dialogue system. In addition, the index of whether or not you can afford attachment is also very important in considering long-term communication with the dialogue system. So, It is considered that the usefulness of the "gentleness expression function" proposed in this paper is high.

10 Conclusion

In this paper, we proposed a dialogue system for conversation closely to the elderly in order to solve the problem of the current dialogue system for elderly people. Specifically, we proposed a system to promote dialogue with emphasis on the feelings of elderly people. Specifically, the communication was divided into two. The initial phase is the initial dialogue phase impressing the tenderness of the agent.

From the experimental results, it became clear that (1) subjects received an expression of the gentleness of the agent, and a familiar relationship was born between the subject and the agent in the initial conversation, (2) the continuation of the conversation became smooth.

References

1. 2017 - year elderly society white paper. http://www8.cao.go.jp/kourei/whitepaper/w-2017/html/zenbun/index.html
2. 2016 Edition Aged Society White Paper. http://www8.cao.go.jp/kourei/whitepaper/w-2015/html/zenbun/index.html
3. Saito, M., Kondo, K., Ojima, T., Hirai, H.: Criteria for social isolation based on associations with health indicators among older people. A 10-year follow-up of the Aichi Gerontological Evaluation Study. Jpn. Public Health Mag. **62**(3), 95–105 (2015). (in Japanese)
4. Shitaoka, K., Tokuhisa, R., Yoshimura, T., Hoshino, H., Watanabe, N.: Active listening system for a dialogue robot. Nat. Lang. Process. **24**(1), 3–47 (2017). (in Japanese)
5. Inaba, M., Kamizono, S., Takahashi, K: Candidate utterance acquisition method for non-task-oriented dialogue systems from twitter. J. Artif. Intell. Soci. **29**(1), 21–31 (2014). (in Japanese)
6. https://www.ninchisho-forum.com/knowledge/iryou/008.html
7. Daibo, I. (ed.): Social psychology perspective 1, Sei Shin Shobo, May 1989
8. Moriko, Y., Hori, Y. (eds.): Psychometric Measure ScaleI. Science, June 2001
9. MMDAgent. http://www.mmdagent.jp/
10. MeCab. http://taku910.github.io/mecab/

Design and Evaluation Methodology for Cockpit Human Factor of Civil Transport Aircraft

Jin Zhefeng(✉), Dong Dayong, Zhang Yinbo, and Meng Hua

COMAC Shanghai Aircraft Design and Research Institute,
No. 5188 Jinke Road, Pudong New Area, Shanghai, China
jinzhefeng@comac.cc

Abstract. Human factors is ubiquitous and cannot be avoided completely. The ultimate goal of human factors design is to effectively reduce human errors caused by design. To carry out error prevention design based on human-machine interface design is the core of human factors design. The engineering design phase is an important stage to effectively control and solve design induced human error problem. The key elements of human factors design include five aspects: layout, control device, information display, alerting, automation, following specific design principles and enhancing integration design, to increase the human-machine interface design efficiency, and to apparently reduce the likelihood of human errors. At the same time, human factor evaluation is an effective complementary measure to find design defect. This article puts forward the methods of civil transport aircraft cockpit human factors design and evaluation from the perspective of engineering designers. Practice has proved that the method is feasible and it can provide certain guidance for civil transportation cockpit human factors design field.

Keywords: Civil transport aircraft · Cockpit · Human factor

1 Introduction

Data shows that over the past century, as many as 75% of the accidents were caused by human factors. Errors (human error) caused by human factors have become the main threat to aviation safety. For this purpose, the aviation industry, including the FAA, EASA airworthiness authorities have carried out a large number of researches for aviation human factors issues, in order to improve the level of aircraft safety through the analysis of the causes and solutions of human errors. Many “pilot error” is actually “designer error”. Although many of the accident investigation reports often draw “pilot error caused the accident” research conclusion. While in fact, a lot of the “pilot error” are essentially “designer error”. In the man-machine system, human error is not only a simple process of man-made mistakes, more importantly, the unreasonable system design provides the probability of existence or continued development of a potential behavior that may result in accident. This kind of human errors can often be designed to be avoided.

In the cockpit environment, the flight crew realizes information and operation interaction mainly through the display system (display information, alerting), computer (automation) and control system (control device). From the design engineer’s point of view, by clearly understanding the design elements of the human factors and following

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 269–275, 2020.
https://doi.org/10.1007/978-3-030-20040-4_24

certain design principles, and integrating the idea of error prevention design into the human factors design elements can effectively improve the degree of friendly human-machine interface and reduce the probability of human errors emerged from the source.

2 Cockpit Human Factors Design Elements and Principles

The key elements of human factors design include the five aspects of layout, control device, information display, alerting, automation, which by follow certain design principles can effectively increase the degree of friendly human-machine interface design, in order to obviously reduce the probability of human errors.

2.1 Arrangement and Layout

Under the premise of meeting the requirements of the pilot's visibility, accessibility, comfort, and enough operation space, the cockpit layout should carry out comprehensive arrangement of the cockpit system and equipment. The cockpit layout can follow the following principles:

The cockpit layout can decorate the instrument panel, display device, and control and manipulation equipment based on the design eye point of target market standard pilot (the 50th percentile) as the benchmark and in order to meet the pilot external visibility requirements to design the windshield, satisfy the visibility and accessibility and comfort requirements.

The cockpit layout should consider the limitations of human body. Due to the relatively large differences in human body sizes, the cockpit design should be based on the arithmetic difference measurement data statistics of anthropometric data statistics relative to the target market people on 50th percentile pilot that can make sure the 5th percentile and 95th percentile limit the statistics of the size of the pilot to achieve good accessibility and visibility (Fig. 1).

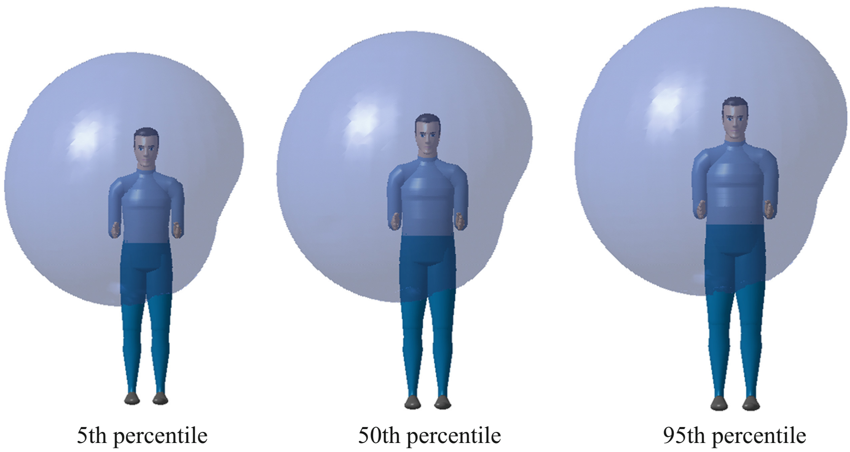


Fig. 1. This shows a figure about pilot's accessibility of the right index finger.

2.2 Controls

Controls are defined as devices the flight crew manipulates in order to operate, configure, and manage the airplane and its flight control surfaces, systems, and other equipment. This may include equipment in the flight deck such as: buttons, switches, knobs, keyboards, keypads, touch screens, Cursor control devices, Graphical user interfaces, such as pop-up windows and pull-down menus that provide control functions, voice activated controls [1]. The cockpit controls design suggestions follow the following principles:

(a) Controls design should maintain the cockpit "static dark";
(b) Controls Should have fast and convenient way to keep the pilot/flight crew to exercise control the plane;
(c) Controls layout should match the pilot/flight crew tasks;
(d) Controls layout should consider about same driving safety from left and right side seat;
(e) Controls layout should make flight crews more comfortable manipulation;
(f) Controls layout should consider operating procedures, to avoid the pilot upper substantial movement;
(g) Controls layout and display layout should be considered the relative position of relationship, as far as possible use nearby principle;
(h) Controls should be good visibility in various environments;
(i) Controls should fully consider the accessibility, shall ensure that the pilot work in the position to full trip control necessary controls;
(j) Controls should have a clear identity;
(k) Controls movement direction and system response should be consistent with the pilots' forecasts;
(l) The effect of the controls action after flight crew should be made clear and quick perception;
(m) Controls action and feedback of the system should be to keep the flight crew situational awareness;
(n) The current state of controls must be able to make the flight crew clear unambiguous perception;
(o) Different forms of operation prevention measures of controls should be taken to reduce the flight crew operating errors;
(p) Controls design should minimize the flight crew potential manipulation mistakes;
(q) Controls design should have corrective actions after the flight crew found any mistakes; Such as typing errors, etc.;
(r) Controls design should avoid excessive energy dispersion due to manipulation of controls, to reduce situational awareness, to lose their stance, etc.;
(s) Controls should consider state changes caused by environmental factors such as vibration;
(t) Different forms of controls should have different forms of feedback such as mechanical, sound, light, touch to make flight crew to perceive controls action;
(u) Controls design should consider the operating force which could avoid to bring extra physical load to the flight crew;
(v) Controls should not have sharp bumps which could avoid physical injury to flight crew.

2.3 Displays

Displays provide the pilots with visual information which is needed to complete the mission, including status of altitude, attitude, airspeed, horizontal situation, the engine thrust and rpm, hydraulic, fuel system, etc. Displays design should follow the following principles:

(a) Displays Should provide sufficient and necessary information when the flight crew make a final decision to exercise control authority so that could make sure flight crews are able to control the aircraft safely;
(b) Flight crews should be able to access all useful information for decision making;
(c) Information about the system current state that matches with flight crew task should be provided;
(d) Information display mode should consider flight crews cognitive characteristics;
(e) Information display design should consider the collaboration between the flight crew, display information should not make the flight crew to understand the state of the aircraft;
(f) The layout of the information display should consider the pilot's visual scanning features, important information should display in the easiest way to observe the location of the pilot;
(g) Primary flight displays should adopt the "T" type of typical layout;
(h) Both on the ground and air, information display for the pilot/flight crew should be easy identification;
(i) Information display should use the strong contrast of foreground and background colors that can make the characters, symbols, identification, calibration and so on are easier to identify;
(j) Information display should use different color to identify meaning and importance;
(k) The response time of display information system equipment which can be in a timely manner to reflect the state of the system should be considered. The pilot's physiological characteristics should be considered at the same time that can help to enhance the flight crew situational awareness;
(l) Any input for flight crew by multi-function keyboard, knobs, buttons should be defined in a clear and quick response to flight crew input effect;
(m) Information display content should be intuitive, minimize the additional workload caused by the flight crew reworking information;
(n) Information display timeliness should be considered to avoid too much useless information appear on the display at the same time;
(o) Flight crew should have a simple way to show the required information;
(p) All kinds of symbol position and direction of movement should be consistent with the plane movement or consistent with the flight crew cognitive habits;
(q) Units of display information should be paid attention to, especially English unit and the unit should be paid attention to;
(r) Displays Should avoid or reduce glare caused by lighting and external light;
(s) Displays should be considered flight crew unconscious damage such as water splashing, foreign bodies, touch screen, etc.;

(t) When one or more than one display device failure, information display should still be able to provide necessary information for flight crew to safely flight;
(u) Information display should consider to reduce the flight crew's visual fatigue.

2.4 Alerting

The role of the alerting system is to attract the attention of the flight crew, make them know the failure or malfunction of aircraft and aircraft systems, or other abnormal conditions. The design of modern aircraft crew alerting also need to prompt flight crew possible measures that can be taken. It is suggested that the design of alerting system follow the following principles:

(a) The alarm information should let the pilot/flight crew clear about aircraft fault program;
(b) Warning (especially voice and display alarm) must be meaningful to avoid ambiguity;
(c) The alarm should consider different levels and different forms. The same level alarm should have consistency;
(d) When necessary, the aircraft can provide measures to eliminate the alarm;
(e) Inhibition of alarm should be considered a potential security risk;
(f) The alarm status indicator should make the flight crew rapid perception about current normal or abnormal state;
(g) The alarm sound and light should not bring to the body harm of the flight crew;
(h) The alarm sound shall not affect the flight crew's normal voice communication;
(i) The alarm shall not affect the flight crew to make the right judgments;
(j) The aircraft should provide disposal way of the alarm for the flight crew;
(k) The aircraft should provide controls which can quickly remove sound and light alarm for the flight crew;
(l) The aircraft should have the alarm logging to look back.

2.5 Automation

Automation refers to the process of machine, equipment or system achieving intended target following the designer's requirement, without direct flight crew intervention, through automatic test, information processing, analysis and decision, manipulation and control. The introduction of automation technology changed aircraft cockpit human-machine interface and way of flight operation, which increased the possibility of flight crew making errors. Therefore, design principles for automation should be established at the early stage of cockpit design, as guidance for the application of automation technology. It is suggested that the design of automation follow the following principles:

(a) Automation should be designed as a complement and/or as an aid to the flight crew who decide when and how to use it;
(b) Automation concepts should answer the three essential questions of why, what and how to automate;

(c) The flight crew should be given sufficient information to allow monitoring and management of the automation. These indications may be visual, tactile, or auditory;
(d) The goal of automation is not to bring crew workload to zero but rather to achieve an optimal level of workload. Too little workload can be as hazardous as too much workload.

3 Cockpit Human Factor Evaluation

Cockpit human factors evaluation work run throughout the entire cockpit human factors design development and validation process. In the evaluation of cockpit initial system design and optimization, system design, evaluation and the evaluation of the optimization and integration of the cockpit design optimization phase, there are different types of evaluation activities planned. The evaluation of cockpit evaluation work scope includes two main parts: physical efficiency and human-computer interaction.

Physical efficiency mainly includes spatial placement, equipment layout, external visibility, cockpit environment, etc., focusing on the cockpit central pedestal, instrument panel, display units, overhead panel, side console, seat, pedal, sidestick, and the arrangement of corresponding controls. Physical efficiency evaluation considers human characteristics, including the physical parameters, operating habits, visibility, auditory, fatigue, etc., and consider all kinds of usage scenarios for evaluation.

Human-machine interaction includes control device, display information, the alerting, etc., focusing on control device and display information used by the flight crew for flying, navigation, communication, and management of system. Human-machine interaction evaluation considers people's cognitive characteristics, including perception, understanding, decision-making, implementation, and considers a variety of tasks scenario for evaluation.

Cockpit human factors assessment mainly adopts graphic and text evaluation, static evaluation, subtask evaluation, and full task evaluation. The evaluation work of planning include assessment with stakeholders, specific as follows:

Graphic and Text Evaluation. Graphic and text evaluation is mainly conducted in the form of meeting, by inviting pilots and human factor experts to review the system's control, display, alerting, etc. Graphic and text evaluation is mainly based on description document, design drawings, etc. Graphic and text evaluation requires 2 to 3 experienced test pilots with engineering background.

Static Evaluation. Static evaluation focuses the evaluation of visibility and accessibility, by inviting a reasonable sample of pilots to the static prototype for evaluation. Static evaluation is performed on platforms installed with corresponding static controls, such as the engineering mock-up, engineering simulator, etc. Static evaluation requires 2 to 3 experienced test pilots with engineering background.

Subtasks Evaluation. Subtask evaluation is based on flight crew task, focusing mainly on a single system, evaluating all its usage scenarios and tasks to be carried out. Subtask evaluation need to be conducted on platforms equipped with the appropriate functions, such as the function simulation platform, functional integration test platform, full-function engineering simulator, etc. Subtask evaluation needs to define scenarios and tasks. Subtask evaluation requires 2 t o 3 experienced test pilots with engineering background, as well as the participation of cockpit human factor experts.

Full Task Evaluation. Full task evaluation is based on the flight crew task assessment, focusing on multiple systems, evaluating norma, abnormal, and emergency usage scenarios and tasks to be carried out. Full task evaluation is conducted on platforms with full cockpit functions, such as a full-function integrated test platform, full-function engineering simulator, or on the aircraft. Full task evaluation needs to define scenarios and task Full task evaluation requires 2 to 3 experienced test pilots with engineering background, as well as the participation of cockpit human factor experts.

4 Conclusion

From the perspective of engineering designers, this article puts forward a method for human factors design and evaluation of civil transport aircraft cockpit, which can provide certain guidance in the civil transportation aircraft cockpit human factors design field. At present, although human factors design cannot guarantee that errors will not appear in the process of all operations or maintenance, integrating the error prevention design considerations into the five aspects of human factor design elements such as cockpit layout, control, display, alerting, automation, which no doubt is one of the best ways to reduce human errors from the design source. Especially for aircraft designer, to improve the level of aviation safety, they should work harder to design the aircraft to be more “friendly” and “error preventive” and “error tolerant”. In view of this, this article suggested that we should attach great importance to introduce integrated design and error prevention considerations in aircraft design, learn from all kinds of aviation accidents or incidents, and look for human factors design solutions. At the same time, we should carry out research on cockpit human factors evaluation method to detect design flaws that will lead to human errors as early as possible in aircraft design phase, effectively prevent incorrect design, reduce the probability of human errors in the subsequent use, and save the cost of aircraft design change.

Reference

1. Federal Aviation Administration: Installed Systems and Equipment for Use by the Flightcrew. ANM-111.5/3/2013 (2013)

System Accessibility Design

Accessibility Assessment in Mobile Applications for Android

Patricia Acosta-Vargas[1](✉), Luis Salvador-Ullauri[2], Janio Jadán-Guerrero[3], César Guevara[3], Sandra Sanchez-Gordon[4], Tania Calle-Jimenez[4], Patricio Lara-Alvarez[3], Ana Medina[1], and Isabel L. Nunes[5]

[1] Intelligent and Interactive Systems Lab, Universidad de Las Américas, Quito, Ecuador
{patricia.acosta,anagabriela.medina}@udla.edu.ec
[2] Centro de Educación Continua, Escuela Politécnica Nacional, Quito, Ecuador
lsalvador@cec-epn.edu.ec
[3] MIST, Universidad Tecnológica Indoamérica, Ambato, Ecuador
{janiojadan,cesarguevara,patolara}@uti.edu.ec
[4] Escuela Politécnica Nacional, Quito, Ecuador
{sandra.sanchez,tania.calle}@epn.edu.ec
[5] CTS/UNINOVA, DEE, Nova University of Lisbon, Monte de Caparica, Portugal
imn@fct.unl.pt

Abstract. At present, the lack of adequate methods to test whether a mobile application is accessible has become a major challenge for accessibility experts. This study was applied to ten mobile applications, the most popular according to PCMAG. We propose to use the Web Content Accessibility Guidelines 2.1 through manual review and automatic review with the Google Play Store Accessibility Scanner validator for the Android. The evaluation results of the mobile applications indicate that the applications are not accessible because they do not comply with the minimum required level proposed by WCAG 2.1. The research proposes suggestions to improve and raise awareness among the designers of mobile applications, in such a way that more inclusive mobile applications accessible to all types of users are built.

Keywords: Accessibility · Assessment · Accessibility Scanner · Mobile applications · WCAG 2.1

1 Introduction

At present, the use of intelligent devices as well as digital technology allow access to essential services and information resources of the web [1], so that it continuously improves people's lives, according to the World Starts Pocketbook [1] in 2018, approximately 45.7% of the world's population uses the Internet.

According to the report of network technology provider Cisco [2], the number of mobile phone users was 5 billion in 2017, but that number will increase to 5.5 billion users in the next five years.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 279–288, 2020.
https://doi.org/10.1007/978-3-030-20040-4_25

The GlobalWebIndex report [3] reveals that 50,000 Internet users between the ages of 16 and 64, or approximately 78% of the population [3] interact on social networks from a mobile device.

The accelerated advance of technology and the creation of new computer applications make software engineering a field with much future. On the other hand, the development of mobile applications faces a significant challenge because there are many software programs and users on the web, but these programs lack accessibility. In the last five years, global digital growth shows no sign of slowing; the usage of smartphones, and mobile applications [4] has bigger significantly. The programmers of these applications have forgotten to design the software thinking about the principle of universal access [5].

The few studies related to accessibility methods for smartphone software are attributed to the lack of adequate tools, guidelines, and policies for valuing mobile applications. This study proposes to apply the Web Content Accessibility Guidelines (WCAG) 2.1 [6, 7] for the advance of accessible mobile applications.

On the other hand, the Google Play Store provides the tool to validate specific guidelines for smartphones based on Android with Accessibility Scanner [8]. In this research, we adopted the definition of "Mobile Accessibility" [7, 9] by considering that applications should be accessible to all types of users, regardless of physical, cognitive or motor disabilities, when using smartphones.

In our study, accessibility in mobile applications is related to the state in which any user can interact with a smartphone or software.

In this study, we applied the manual revision with the WCAG 2.1 [6, 7] and the Google Accessibility Scanner validator for Android, updated to March 29, 2018, [7] with version 1.2.

The proposed method to find the value of the level of accessibility of mobile applications contains the following phases:

- select mobile applications to evaluate
- install the Google Play Store applications
- configure and activate the accessibility scanner
- open the mobile application to evaluate
- save the results in a spreadsheet
- analyses the results of the evaluation
- Suggest accessibility improvements

This research is organized as follows: Sect. 1 presents the introduction to the proposed research, Sect. 2 contains the background and related studies, Sect. 3 presents the method applied, and the case study, Sect. 4 details the analysis and discussion of the results, and Sect. 5 provides conclusions and an approximation of future work for this research.

2 Background and Related Work

A mobile application is a software program [7] for use on a portable device or smartphone. The application can be hybrid or native, depending on the technologies for which it was designed [10].

On the other hand, accessibility [11] refers to the barriers that exist between user communication and mobile application. It refers to a design of mobile applications that allow users [12] to perceive, comprehend, navigate and interrelate with the application while providing content.

Web Content Accessibility Guidelines 2.1 [6], contain four principles:

Perceivable: The first principle indicates that users should be able to distinguish content visually, tactilely and sonorously [6].

Operable: The second principle conditions that users should be bright to navigate appropriately using the interface components [6].

Understandable: The third principle is related to content and how to interface controls should be understandable for user management [6].

Robust: The fourth principle refers to content, which must be robust adequate to be consistently understood by users of applications that allow them to work handily with current and future technologies [6].

Web Content Accessibility Guidelines 2.1 [6], consist of four principles, 13 guidelines and 76 compliance criteria [13], an indeterminate number of advisory techniques and sufficient techniques. These are associated with one of the following levels of compliance:

Level "A" is the minimum level of access to the web. This level is reached when all the criteria related to the level are met. Otherwise, a group of operators will not be bright to access the contents of the website.

Level "AA" is the middle level of accessibility. If it is not reached, it will be a challenge for a group of users to access the website and its contents.

Level "AAA" is the top level of access to the web. Involves a greater degree of ease for users to access the web as content. A website or mobile application that reaches level "AAA" is a website or application that can be accessed by all users [11].

Previous studies by several authors related to accessibility in mobile applications were reviewed, described below:

According to de Oliveira et al. [14], there are visually impaired people who use screen readers. Regardless of the growing number of accessibilities guiding principle and recommendations, there are now accessibility problems in mobile applications. Also, accessibility guidelines, tools are needed [15] to automate the procedure of integrating accessibility into applications. The authors propose a systematic literature review to identify [15] state of the art in native mobile applications.

According to Damaceno et al. [16], people present a series of visibility barriers when using mobile devices. The authors propose, to solve this problem, a set of recommendations.

Acosta-Vargas et al. [7] apply Accessibility Scanner and WCAG 2.1 to assess accessibility in mobile applications, the results obtained reveal that not all these mobile applications are accessible, suggest correcting the barriers identified to improve the level of accessibility.

3 Method and Case Study

In this case study, we evaluated ten mobile applications installed on smartphones for the Android operating system. We randomly selected applications from popular applications around the world according to PCMAG[1]. The Accessibility Scanner[2], version 1.2 of the Google Play Store, was applied in the accessibility evaluation; this tool applies some WCAG 2.1 [7, 11]. Data collection [7] took place in November 2018. Table 1 contains the tools valued in this case study. It contains the application identifier, tool name, logo, tool version, number of downloads and application type.

Figure 1 contains an outline of the method applied in this study to evaluate accessibility in mobile applications, randomly selected from among the most popular according to PCMAG magazine, contains seven phases structures as follows:

Select the Applications [7]. In the first phase, we randomly selected ten applications from the group of most popular applications used around the world according to PCMAG[3]. The more detailed description of the tools evaluated, please refer to Table 1, which contains the data related to the mobile applications assessed in this case study.

Install the Applications from the [7] Google Play Store. In the second phase, the selected applications are in Google and installed with the requirements of each tool according to Table 1. In the case study, we used version 7.0 of the Android operating system.

Install and Activate [7] Accessibility Scanner. In the third phase, to assess the accessibility of each mobile application, it is essential to install the Accessibility Scanner validator, developed by Google, used the updated version on September 21, 2018, has more than 100,000 downloads, as a requirement for installation is necessary that smartphones have as a minimum requirement the Android 6.0 operating system. Once we installed the application, we activate the accessibility test; then the application is finally initialized, we proceed with the accessibility evaluation on the applications detailed in Table 1.

Open the Application to Evaluate. In the fourth phase, we interact with each mobile application; when testing the functionality of the application, we run the accessibility test. The application automatically captured the data for analysis, then storing the information in the Accessibility Scanner application log. The data for analysis has the options to share via email, social networks or to store directly in the cloud.

[1] https://www.pcmag.com/article/362295/the-100-best-android-apps.

[2] https://support.google.com/accessibility/android/answer/6376570?hl=en-GB.

[3] https://www.pcmag.com/article/362295/the-100-best-android-apps.

Table 1. Mobile Application most used according to PCMAG

Id	Tool	Logo	Updated	Download	Application type
A	WhatsApp		October 24, 2018	1.000.000.000+	Messenger service
B	Skype		October 25, 2018	1.000.000.000+	Messenger service
C	Facebook		October 31, 2018	1.000.000.000+	Social net-works
D	Twitter		October 29, 2018	500.000.000+	Social net-works
E	Snapseed		June 29, 2018	50.000.000+	Photos
F	Instagram		October 31, 2018	1.000.000.000+	Photos
G	YouTube		November 2, 2018	1.000.000.000+	Video and movies
H	Google Apps		November 1, 2018	1.000.000.000+	Productivity
I	Microsoft Office		January 18, 2019	500.000.000+	Productivity
J	Angry Birds		December 5, 2018	100.000.000+	Entertainment

Record the Results. In the fifth phase, we record the results of the evaluation of the mobile applications in a Microsoft Excel spreadsheet. Table 3 contains the summary information, related to the evaluation of mobile apps with the [7] Accessibility Scanner.

Analyze the Results. In the sixth phase, we analyzed the results of the accessibility evaluation applied in the previous phase. This phase is essential to study and relate the problems identified with the Accessibility Scanner tool. We identify the accessibility barriers, then relate them to the four accessibility principles, the success criteria and the respective accessibility levels for the mobile applications according to WCAG 2.1 [6].

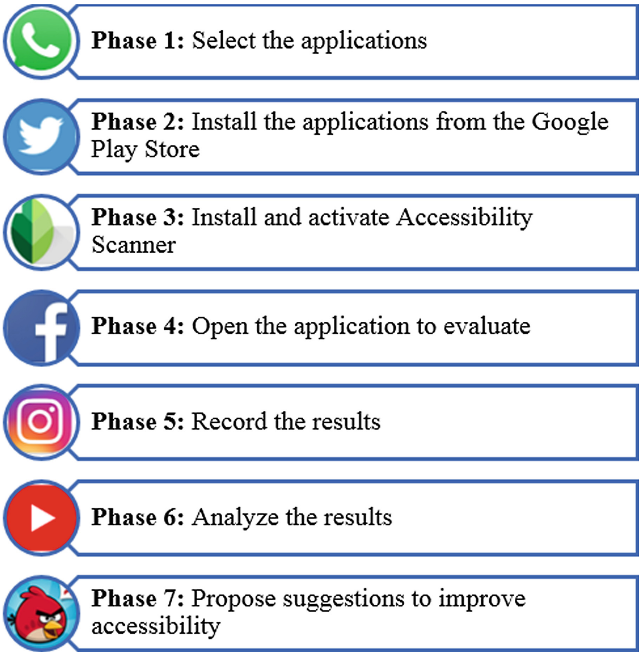


Fig. 1. Method for accessibility assessment

Propose Suggestions to Improve Accessibility. In the seventh and final phase, we reviewed the barriers identified in the assessment of mobile applications. In the section on conclusions and future work, we have detailed the proposals for improving accessibility in mobile applications.

4 Results and Discussion

Table 2 contains the details of the principles and criteria of success identified in the evaluation. The analysis can help understand the applicability of WCAG 2.1 [6, 17] and [18] User Agent Accessibility Guidelines (UAAG) 2.0 to the mobile environment. The information found can be useful for designers, developers, and app content evaluators for mobile devices. The study identifies additional accessibility needs in the mobile environment related to the principles of WCAG 2.1. The dataset of this study is available in the Mendeley[4] repository, in Microsoft Excel format.

Table 3 contains the Accessibility Scanner evaluation of the initial screens of this case study detailed in Table 1. Table 3 contains the mobile application identifier, the number of elements evaluated, the touch lens, the contrast, the label, the description, the image contrast and the type of element not supported.

[4] https://data.mendeley.com/datasets/m39bfzbzgc/3.

Table 2. Barriers related to WCAG 2.1

#	Principle	Mobile accessibility considerations	Success criterion	Level	Error
1	Perceivable	2.1 Small Screen			Item label
		2.2 Zoom	1.4.4 Resize text [6]	AA	Text contrast
		2.3 Contrast	1.4.3 Contrast [6]	AA	Text contrast
			1.4.6 Contrast (Enhanced) [6]	AAA	Text contrast
2	Operable	3.1 Keyboard Control [6]	2.1.1 Keyboard [6]	A	
			2.1.2 No Keyboard [6]	A	
			2.4.3 Focus Order [6]	A	
			2.4.7 Focus Visible [6]	AA	
		3.2 Touch Target Size			Touch target
		3.3 Touchscreen Gestures			Touch target
		3.4 Device Manipulation Gestures	2.1.1 Keyboard [6]	A	
		3.5 Placing buttons			
3	Understandable	4.1 Changing Screen Orientation			Item descriptions
		4.2 Consistent Layout	3.2.3 Consistent Navigation [6]	AA [7]	Item descriptions
			3.2.4 Consistent Identification [6]	AA [7]	
		4.3 Positioning essential page elements			
		4.4 Grouping operable elements	2.4.4 Link Purpose	A [7]	
			2.4.9 Link Purpose	AA [7]	
		4.5 Provide a clear indication	3.2.3 Consistent Navigation [6]	AA [7]	Item label
			3.2.4 Consistent Identification [6]	AA [7]	
		4.6 Provide instructions for custom touchscreen	3.3.2 Labels or Instructions [6]	A [7]	Item label
			3.3.5 Help [6]	AAA [7]	
4	Robust	5.1 Set the virtual keyboard			
		5.2 Provide natural methods			

Table 3. Data obtained when evaluating accessibility

Id	# Elements	Touch target	Text contrast	Item label	Item descriptions	Image contrast	Type of element not supported
A	10	5	4	0	0	1	0
B	8	1	6	0	1	0	0
C	7	3	1	3	0	0	0
D	6	2	0	0	1	3	0
E	3	1	2	0	0	0	0
F	12	9	0	1	1	0	1
G	6	5	0	0	1	0	0
H	11	2	6	0	1	2	0
I	3	1	1	1	0	0	0
J	1	0	0	1	0	0	0

When applying descriptive statistics to the most frequently identified errors, the average number of errors corresponds to 11.17. The typical error corresponds to 4.44; the median is 6, the mode is 6, the standard deviation is 10.87, the variance is 118.17. The kurtosis is −0.19, the asymmetry coefficient is 1.12, the range is 28, the minimum value is 1, the maximum corresponds to 29 with a confidence level of 95.0%.

Figure 2 shows the most frequently repeated accessibility failures in mobile applications. It is observed that the errors that are repeated with greater frequency correspond to "Touch target" with 29 failures that correspond to 43.3% and "Text contrast" with 20 failures that correspond to 29.9%. Followed by "Item label" and "Image contrast" with six failures that represent to 9%, "Item descriptions" with five failures that represent 7.5% and finally, "Type of element not supported" with one failure that corresponds to 1.5%.

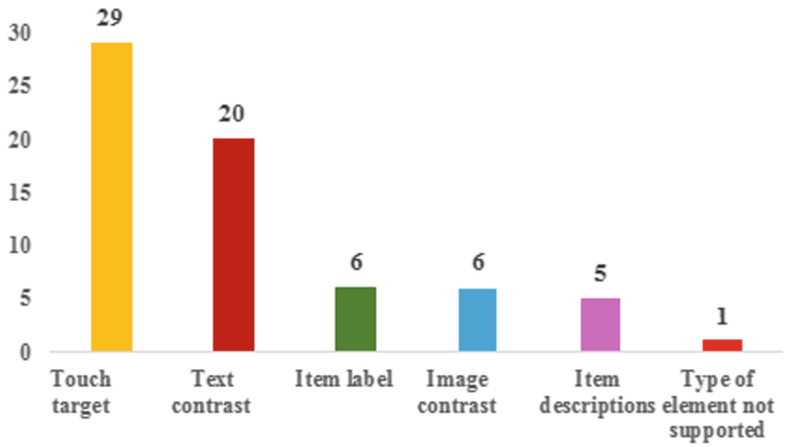


Fig. 2. Accessibility barriers detected

5 Conclusions and Future Works

Digital inclusion of users is easy to achieve with a list of identified barriers in mobile applications, which must be corrected before the application goes into production. The results of this research show that rated mobile applications violate the accessibility levels recommended by the World Wide Web Consortium. It is, therefore, essential to raise awareness among mobile application developers to apply WCAG 2.1 throughout the application development cycle in such a way as to guarantee at least a minimum level of accessibility.

To solve the problem of the "item label," it is proposed to minimize the amount of information in the dedicated mobile version or to propose a sensitive design. To solve the problem of "text contrast" concerning the browser it is essential to set the default text size, render and enlarge the browser window. In general, it is suggested that text be resizable without assistive technology by up to 200%. Besides, it is essential to apply a contrast of at least 4.5:1.

The "touch target" problem can be solved by ensuring that trace marks are at least 9 mm high by 9 mm wide [6].

To solve the problem of "item descriptions," developers must be able to ensure that all users can quickly change the orientation of the application. Also, we suggest that the components that are repeated in several web pages offer a more consistent and stable design.

In order to fix a problem related to the "item label," it is necessary that the elements to which the changes are applied are sufficiently distinct from distinguishing them from unprocessable elements. Also, device manipulation gestures and custom touchscreens can help developers create efficient interfaces.

Finally, for the "Robust" principle, it is essential that the standard keyboard allows device configurations to be customized.

For future work, we suggest combining the best methods in the assessment of mobile applications. During the evaluation phase of the ten mobile applications, we identified that the barrier that repeats most frequently relates to the "text contrast" second we have the "tactile target."

In future research, we recommend that for the evaluation of accessibility for mobile applications we consider (i) users with different types of disabilities, (ii) the barriers for each type of disability and (iii) different scenarios to be evaluated in the application. This study can serve as a starting point for evaluating accessibility in mobile applications for Android and iOS. The perception is that the Android system is more complicated than iOS, but both platforms include a variety of accessibility features. Programmers and designers will make a difference in the development of mobile applications when applying WCAG 2.1.

Acknowledgments. The authors thank UDLA for funding this study through the project "FGE. PAV.18.10".

References

1. United Nations: World Statistics Pocketbook 2018 Edition (2018)
2. Cisco: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 White Paper – Cisco. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html
3. GlobalWebIbdex's: GWI Social (2017)
4. Choudhary, S.R., Gorla, A., Orso, A.: Automated test input generation for Android: are we there yet? In: Proceedings - 2015 30th IEEE/ACM International Conference on Automated Software Engineering, ASE 2015, pp. 429–440 (2016)
5. Zein, S., Salleh, N., Grundy, J.: A systematic mapping study of mobile application testing techniques. J. Syst. Softw. **117**, 334–356 (2016). https://doi.org/10.1016/j.jss.2016.03.065
6. World Wide Web Consortium (W3C): Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/
7. Acosta-Vargas, P., Zalakeviciute, R., Luján-Mora, S., Hernandez, W.: Accessibility evaluation of mobile applications for monitoring air quality, pp. 1–11. Springer, Cham (2019)
8. Google: Make apps more accessible. https://developer.android.com/guide/topics/ui/accessibility/apps
9. Mobile Accessibility at W3C—Web Accessibility Initiative (WAI)—W3C. https://www.w3.org/WAI/standards-guidelines/mobile/
10. Jabangwe, R., Edison, H., Duc, A.N.: Software engineering process models for mobile app development: a systematic literature review. J. Syst. Softw. **145**, 98–111 (2018). https://doi.org/10.1016/j.jss.2018.08.028
11. World Wide Web Consortium (W3C): Introducción a la Accesibilidad Web. https://www.w3c.es/Traducciones/es/WAI/intro/accessibility
12. Acosta-Vargas, P., Rybarczyk, Y., Pérez, J., González, M., Jimenes, K., Leconte, L.: Towards web accessibility in telerehabilitation platforms. In: ETCM (2018)
13. Acosta-Vargas, P., Jadán-Guerrero, J., Guevara, C., Sanchez-Gordon, S., Calle-Jimenez, T.: Technical contributions to the quality of telerehabilitation platforms: case study—ePHoRt project. In: Assistive & Rehabilitation Engineering, pp. 1–22. IntechOpen (2019)
14. de Oliveira, A.F.B.A., Filgueiras, L.V.L.: Developer assistance tools for creating native mobile applications accessible to visually impaired people. Presented at the (2018)
15. Ferreira, C., Ferreira, S.B.L.: Mobile application accessibility in the context of visually impaired users
16. Damaceno, R.J.P., Braga, J.C., Mena-Chalco, J.P.: Mobile device accessibility for the visually impaired: problems mapping and recommendations. Univers. Access Inf. Soc. **17**, 421–435 (2018). https://doi.org/10.1007/s10209-017-0540-1
17. World Wide Web Consortium: Mobile Accessibility: How WCAG 2.0 and Other W3C/WAI Guidelines Apply to Mobile. https://www.w3.org/TR/mobile-accessibility-mapping/
18. World Wide Web Consortium (W3C): Leading the web to its full potential. https://www.w3.org/

Development of an Accessible Video Game to Improve the Understanding of the Test of Honey-Alonso

Luis Salvador-Ullauri[1], Patricia Acosta-Vargas[2](✉), Janio Jadán-Guerrero[3], Cesar Guevara[3], Sandra Sanchez-Gordon[4], Tania Calle-Jimenez[4], and Patricio Lara-Alvarez[3]

[1] Centro de Educación Continua-Escuela Politécnica Nacional, Quito, Ecuador
lsalvador@cec-epn.edu.ec
[2] Universidad de Las Américas, Quito, Ecuador
patricia.acosta@udla.edu.ec
[3] MIST, Universidad Tecnológica Indoamérica, Quito, Ecuador
{janiojadan, cesarguevara, patolara}@uti.edu.ec
[4] Escuela Politécnica Nacional, Quito, Ecuador
{sandra.sanchez, tania.calle}@epn.edu.ec

Abstract. When evaluating the learning styles of several individuals using the Honey-Alonso test, some users did not understand the meaning of several of the questions. This may be due to problems of context, tiredness in front of the extension of the test, lack of understanding or disinterest. The Honey-Alonso test consists of four groups of twenty questions each. Each group of questions allows identifying the level that an individual possesses on each one of the four learning styles. These styles are: active, reflective, theoretical and pragmatic. Answering a questionnaire of eighty questions is not an easy task from an andragogical point of view. This article proposes the creation of an educational video game designed with a script based on the questions of the Honey-Alonso test. The answers selected by the player are taken as a condition to determine the order of the next questions presented to the player.

Keywords: Human factors · Human-systems integration · Systems engineering

1 Introduction

The way a group of students learns is often not the same for everyone. For this reason, the activities in class are usually distributed considering the learning styles of each one. One strategy for distributing activities correctly among the students in a class is to subject them to a test that determines what learning style each student has. The Honey-Alonso questionnaire is used for that purpose [1]. It is a questionnaire of 80 questions, the validity of which has been statistically proven. A questionnaire is a tool that the teacher can use to identify the psychoeducational characteristics of the students. However, the number of questions it contains means that students do not answer it conscientiously. The Honey-Alonso questionnaire has been used in other works to select the activities that best relate to each student within the object of study [2]. This

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 289–298, 2020.
https://doi.org/10.1007/978-3-030-20040-4_26

type of selection process can be promising if the result of the questionnaire reflects the learning style of the students. Although there is no time limit for the questionnaire to be answered, many students prefer to answer their questions randomly. It is a problem if we consider that the activities assigned to the students depend on the answers to the questionnaire. The lack of motivation in the students to answer the questionnaire makes it necessary to search for strategies to facilitate this task. To help students respond sincerely and calmly to the questionnaire, a video game that allows students to internalize their ideas before answering each question has been generated [3]. The video game contains illustrations that make it easier to understand the questions. Besides, the order of presentation of the questions has been configured according to the answers given by the students. The use of technology attracts the attention of students to perform ordinary tasks in an extraordinary way [4]. This article describes the process of analysis, design, development, and testing of a video game aimed at motivating students to solve the Honey-Alonso questionnaire.

The second part describes the method used in this work. The third part describes the analysis phase of the application. The fourth part describes the design phase of the application. The fifth part shows some results obtained when using the video game. Finally, some conclusions and suggestions for future work are presented.

2 Method

This research applies User-centered design (UCD). User-centered design is an iterative design process in which designers focus on the users and their needs in each phase of the design process [5]. Each iteration of the UCD approach involves four distinct phases. These phases are: understand the context of use, specify user requirements, design solutions, and evaluate against requirements [6].

In the phase of understanding the use of context, an interview is conducted with stakeholders to analyze the scope, feasibility, timing, and costs of product development. This phase is essential to know if the project should be implemented.

In the user requirements specification phase, future users of the product explain the requirements that the product design should consider. Product requirements are continuously reviewed throughout product development to determine the correct fulfillment of the request. New requirements can be added. Likewise, existing requirements can be changed in order to improve the initial design.

In the solution design phase, the list of requirements is used to design the different parts of the product. Each of these parts is verified in the preliminary architecture of the product. The architecture of the product allows us to understand what the components of the product are and how they relate to each other.

Each part of the product has a particular design that meets the requirements and communication interfaces specified in the architecture made in the preliminary design.

In the evaluation phase, the design and implementation of each part of the product are reviewed against the list of requirements to determine compliance.

All phases are carried out cyclically in order to refine the product. In this way, it is possible to adapt the temporary needs that appear to the design and development of the final product.

Figure 1 shows the methodology involved in the proposed application.

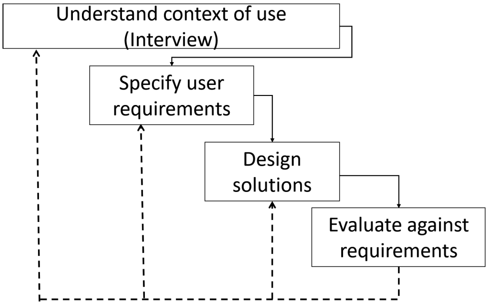


Fig. 1. Phases of User-centered design.

The following sections describe in detail each of the phases carried out in this work.

3 Requirements Analysis

Modern superior education institutions require flexible teaching and learning practices [7], focused on new technologies and providing greater efficiency in the acquisition of knowledge. Addressing real-world problems teaching must focus on a variety of approaches, strategies, materials, and resources to ensure effective learning by most students [8]. Therefore, teachers should consider the different learning styles of students [9]. The Honey-Alonso questionnaire is an easy-to-use tool for determining the learning styles of each student. However, to motivate students to answer each of the questions in the questionnaire consistently, it is necessary to use attractive mechanisms. For this reason, an application was developed that presents the questions in a more enjoyable way.In order to determine what the requirements should be for the design of the application, a series of factors were considered including the degree of understanding of the questions, strategies to improve the understanding of the questionnaire, suggestions to allow an individual answer to each question. Among the main requirements detected at all stages of development are the following:

1. Some of the questions in the Honey-Alonso questionnaire showed a severe level of interpretation among several of the users. It is therefore suggested that illustrations be used to improve the level of comprehension of the written texts.
2. It is advisable to read text aloud to improve its interpretation.
3. The questionnaire should be answered slowly and reflecting on each question.
4. The results of the questionnaire should be immediately shown after it is completed.
5. The own diagnosis should be indicated according to the analyzed results.
6. An indicator of the progress of the questionnaire should be provided.
7. Each question should be identified with the group of questions to which it belongs.

1. Some users may not want every question in the questionnaire to be read.
2. The text of each question must be large enough to be easily read.
3. Each question should be easy to answer.
4. The result of the questionnaire should be transmitted to a virtual education platform through SCORM standard.
5. The application should be able to transmit statistical information about its functioning.
6. The application should work on a web hosting, e-learning platforms or on any distribution medium such as CD or flash memory.
7. The application must use the keyboard or other assistive devices for user interaction.
8. The application must use colors to identify the kind of learning style when information related to them is showed.
9. The questionnaire should be oriented to the context of higher education users in Latin America.

4 Design of the Interface and Functionality of the Application

4.1 Designing a Learning Object

The application with which the questionnaire is represented constitutes a learning activity. This learning activity has content. Moreover, technology is used as a context target for Latin-American students. So, its design must follow the recommendations for the development of a learning object. A learning object is "A digital self-contained and reusable entity, with a clear educational purpose, with at least three internal and editable components: content, learning activities and elements of context. The learning objects must have an external structure of information to facilitate their identification, storage and retrieval: the metadata" [10].

The questions represent the content of the learning object. These questions are classified into four sections. Each section represents a learning style to be evaluated. The four learning styles evaluated are: active, reflective, theoretical and pragmatic.

Each question admits only one answer. The answer can be affirmative or negative. The activity to be developed by the user is to understand each question of the test and answer it with the truth.

It is necessary to design a learning activity in such a way that the technological context attracts the user of the application. The learning activity is designed considering the Kolb model. Kolb's experiential learning style theory is typically represented by a four-stage learning cycle in which the learner 'touches all the bases' [11]. Figure 2 shows the four-stage learning cycle of Kolb.

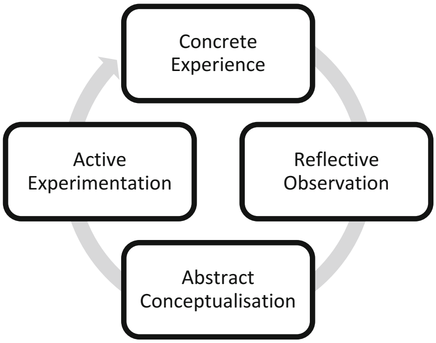


Fig. 2. Four-stage learning cycle of Kolb.

A question is presented to the user every time. Each question has a sound, a color and an image associated with it. The user can read and listen to the question and observe the image associated with it. That is the concrete experience of the user. The presence of an avatar reading the question should allow the user to reflect on all the technological means used to motivate the precise interpretation of the question. When the user answers affirmatively, it is considered that another question related to the previous one can be asked. Each question has a definite point of view. However, the user could find it inappropriate for their context.

A negative answer indicates that the user has interpreted the question in another context and therefore the application will select a new question belonging to another group of questions, thus avoiding that their state of mind influences the rest of the answers. A negative response means that the student is placed in a context where a learning element is exaggerated or underestimated. For this reason, it is necessary not to insist with questions from the same group. The user has reflected and has an abstract concept different from the one we want to transmit.

The Honey-Alonso questionnaire is an essential tool to determine if the user is within the appropriate context to develop active experimentation [12]. Therefore, active experimentation will be subject to the learning styles that the questionnaire evaluates as most emphasized. The result of the test will be used to determine which activities the user should experience.

For a student to learn meaningfully, it is necessary to consider their prior knowledge and ideas, their needs, expectations, learning styles and strategies. Using differentiated instructional strategies, teachers can meet the varying needs of all students and help them to meet and exceed the established standards [13]. Learning styles, explain in part the student failure but also explain, the success achieved by others.

4.2 Designing the User Interface of the Learning Object

The interface design is one of the essential elements for building a coherent and consistent learning object [14]. The design of the learning object interface requires not only considering aesthetic and functional media objects [15]. It must also contemplate metaphors, the representation of knowledge and the meaning of content through appropriate visualization [16]. The user interface of a learning object consists of mediatic objects or areas. These mediatic areas are defined and located according to the information provided directly by the user.

The user interface of the application has been divided into several areas, each of which has a specific purpose. Table 1 shows the areas of the user interface:

Table 1. Areas of the user interface of the application.

Area	Area content
Questions	The text of each question
Avatar	An avatar who read the text of each question
Image	The image represents each question
Tracking	Map with the selected questions and their answers
Configuration	Controls to set the state of the sound, avatar, and text
Response	Options to respond to the questions

The user interface is kept simple to allow users to focus their attention primarily on the questionnaire and the meaning of the questions. Figure 3 shows the position of the different areas of the user interface on the application screen.

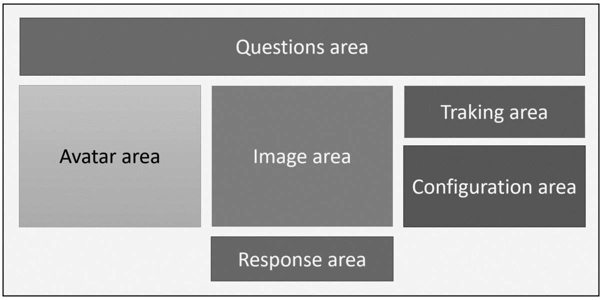


Fig. 3. Areas from the user interface of the application.

4.3 Accessibility Considerations

The design considers the possibility that the application can be used by various types of users, including those who have a disability. The answer to questions can be done using the touch screen, mouse, keyboard, keypad or wireless remote-control presentation. When either option is selected: affirmative or negative answer, the request displays the following question. Each question is assisted by a pre-recorded voice related to the question. Each voice can be activated or deactivated according to the needs of each user. Besides, the application can be used on any device that has a browser compatible with HTML5. Other educational programs have been developed with the problem of accessibility in mind. Many suggestions have been published on how to develop and evaluate this type of programs [17–20]. However, users with deficiencies need simpler interfaces to manage computer applications [21]. Therefore, the need for customization according to the individual need of the user can only be carried out internally, through the modification of the application code.

4.4 Making the SCORM Envelope

The application was developed in HTML5 and Javascript language so that it can be used on the Internet without the need for plugins. For the application to be used on any e-learning platform, the application was required to be wrapped in a SCORM package. To create the SCORM package of the application the tool Reload version 2.0.3 was used. To transfer the results of the questionnaire to the e-learning platform the library SCORM_API_wrapper.js based on the Javascript language was used. The application uses the SCORM 2004 version. However, the game can also be used as a client-server application to transfer the questionnaire results to a PHP-based server program. The transference of information in client-server mode is carried out using the AJAX protocol. Figure 4 shows the architecture of the application.

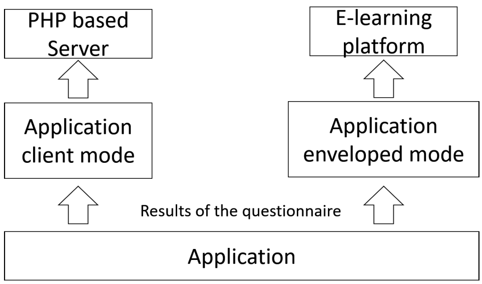


Fig. 4. The architecture of the application.

The users of the application are Spanish speakers; therefore, the application was developed in this language; however, it can be easily translated into any language, due to its simplicity. The final product is shown in Fig. 5 below:

Fig. 5. Final product.

5 Results and Discussion

A group of 120 users was evaluated using the standard physical paper form. These users were then invited to use the application available on the Internet. For the evaluation of the results, the General Preference Scale in Learning Styles was used [22].

The results were grouped by the number of participants belonging to each style category. Two data series were gotten. The first Series corresponds to the information obtained with the physical questionnaire and the second corresponds to the information obtained through digital application. The series of results are shown in Table 2.

Table 2. Series of results with and without application use.

Learning styles: the number of students	Using a paper questionnaire	Using software application
Active	53	69
Reflective	3	1
Theoretical	40	28
Pragmatic	24	22
Total	120	120

To compare both series, we used Pearson's correlation coefficient. Its formula is shown below:

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (1)$$

The value of calculated Pearson's correlation coefficient was $r_{xy} = 0.84$

There is a positive correlation. Therefore, both series are directly related. A direct relation indicates the existence of coherence in the answers of both questionnaires.

6 Conclusions

Users were motivated to use the application and answer all the questions asked the first time they used the program.

The use of the computer allows a more detailed analysis of the results of the questionnaire.

The use of the application modifies the interpretation of the questions, makes them be interpreted in another context, a technological context.

Technology is an essential factor in motivating users to collaborate with their learning process.

For future work, more user requirements would be received from the users in order to improve and customize the current application to the particular needs of each educational center. The results of the questionnaire can be used to automate the assignment of work to each student according to the personal learning style evaluated in the e-learning platforms.

Acknowledgments. The authors thank UDLA for funding this study through the project “FGE. PAV.18.10”.

References

1. Hoffmann, A.F., Liporace, F.: Cuestionario Honey-Alonso de estilos de aprendizaje: Análisis de sus propiedades Psicométricas en Estudiantes Universitarios. Summa Psicológica **10**(1), 103–117 (2013)
2. Marin, P.A.R., Herran, A.O., Dario Duque Mendez, N.: Rules based system to educative personalized strategy recommendation according to the CHAEA test. In: XI Latin American Conference on Learning Objects and Technology. IEEE, Costa Rica (2016)
3. Olguín, J.A.M., García, F.J.C., de la Luz Carrillo González, M., Medina, A.M., Cortés, J.Z. G.: Expert system to engage CHAEA learning styles, ACRA learning strategies and learning objects into an e-learning platform for higher education students. In: Xhafa, F., Barolli, L., Amato, F. (eds.) Advances on P2P, Parallel, Grid, Cloud and Internet Computing, 3PGCIC 2016. Lecture Notes on Data Engineering and Communications Technologies, vol. 1. pp. 913–922. Springer, Cham (2016)
4. Welzer, T., Družovec, M., Kamišalići, A.: Cultural aspects in technology-enhanced education. In: Lhotska, L., Sukupova, L., Lacković, I., Ibbott, G. (eds.) 2018 IFMBE Proceedings on World Congress on Medical Physics and Biomedical Engineering, vol. 68/1, pp. 885–888. Springer, Singapore (2019)
5. Fox, D., Sillito, J., Maurer, F.: Agile methods and user-centered design: how these two methodologies are being successfully integrated in industry. In: Conference on Agile, AGILE 2008, pp. 63–72. IEEE (2008)
6. Uma, V., Suseela, V.J.: UCD approach for the management of user services in University Libraries. Int. J. Adv. Libr. Inf. Sci. **3**(1), 274–290 (2015)
7. Smith, D.E., Mitry, D.J.: Investigation of higher education: the real costs and quality of online programs. J. Educ. Bus. **83**(3), 147–152 (2008)
8. Davis, E.A., Krajcik, J.S.: Designing educative curriculum materials to promote teacher learning. Educ. Res. **34**(3), 3–14 (2005)
9. Boström, L.: Students’ learning styles compared with their teachers’ learning styles in upper secondary school – a mismatched combination. Educ. Inquiry **2**(3), 475–495 (2011)
10. Rehak, D., Mason, R.: Engaging with the learning object economy. In: Littlejohn, A. (ed.) Reusing Online Resources: A Sustainable Approach to E-Learning, pp. 22–30. Kogan Page, London (2003)
11. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development, p. 21. FT Press, Upper Saddle River (2014)
12. Gargallo López, B., Almerich Cerveró, G., Suárez Rodríguez, J.M., et al.: Eur. J. Psychol. Educ. **28**, 1361 (2013)
13. Levy, H.M.: Meeting the needs of all students through differentiated instruction: helping every child reach and exceed standards. Clearing House: J. Educ. Strat. Issues Ideas **81**(4), 161–164 (2008)
14. Rodríguez, V., Ayala, G.: Adaptivity and adaptability of learning objects interface. Int. J. Comput. Appl. **37**(1), 6–13 (2012)

15. Rodríguez, V., Ayala, G.: Design methodology for adaptivity and adaptability of learning object's interface. Int. J. Online Pedagogy Course Des. **3**(2), 77–95 (2013)
16. Rodríguez V.; Ayala G. Diseño de la Interfaz del Objeto de Aprendizaje: Asegurando la Coherencia entre los Objetos Mediáticos. In: LACLO 2008 Tercera Conferencia Latinoamericana de Tecnología de Objetos de Aprendizaje, Aguascalientes, Aguascalientes, México, pp. 269–277 (2008)
17. Salvador-Ullauri L., Jaramillo-Alcázar A., Luján-Mora S.: A serious game accessible to people with visual impairments. In: 9th ICETC International Conference on Education Technology and Computers (2017)
18. Jaramillo-Alcazar A., Lujan-Mora, S., Salvador-Ullauri, L.: Accessibility assessment of mobile serious games for people with cognitive impairments. In: INCISCOS 2017 International Conference on Information Systems and Computer Science (2017)
19. Jaramillo-Alcázar A., Salvador-Ullauri L., Luján-Mora S.: A mobile serious games assessment tool for people with motor impairments. In: 9th ICETC International Conference on Education Technology and Computers (2017)
20. Jaramillo-Alcázar A., Luján-Mora S., An approach to mobile serious games accessibility assessment for people with hearing impairments. In: Advances in Intelligent Systems and Computing, pp. 552–562 (2018)
21. Wolpaw, J.R., et al.: Brain–computer interfaces for communication and control. Clin. Neurophysiol. **113**(6), 767–791 (2002)
22. García, J.L., Santizo, J.A.: Análisis de datos obtenidos a través del cuestionario CHAEA en línea de la página web. Revista Estilos de Aprendizaje **2**(2), 97–98 (2008). www.estilosdeaprendizaje.es

Fuzzy Model for Back Posture Correction During the Walk

Cesar Guevara[1](✉), Janio Jadán-Guerrero[1], Diego Bonilla-Jurado[2], Luis Salvador-Ullauri[3], Patricia Acosta-Vargas[4], Tania Calle-jimenez[5], and Patricio Lara-Alvarez[1]

[1] Universidad Tecnológica Indoamérica, Quito, Ecuador
{cesarguevara, janiojadan, patolara}@uti.edu.ec
[2] Universidad Estatal de Bolívar, Guaranda, Ecuador
dbonilla@ueb.edu.ec
[3] Centro de Educación Continua-Escuela Politécnica Nacional, Quito, Ecuador
lsalvador@cec-epn.edu.ec
[4] Universidad de Las Américas, Quito, Ecuador
patricia.acosta@udla.edu.ec
[5] Escuela Politécnica Nacional, Quito, Ecuador
tania.calle@epn.edu.ec

Abstract. The body posture is not a purely aesthetic problem, due to it can produce a multitude of adverse effects on health, even extremities disorders (injuries) and malfunction of organs. This study focuses on the correct and incorrect detection of the position of the spine and extremities while walking. A database of three patients with lumbago and sciatica, with diagnosis of muscle tension due to poor posture while walking, has been used in this article. These patients underwent physiotherapy treatment and were later filmed taking a short walk of 2 min to see the results. This process was developed for a period of 4 weeks, divided into 2 h of physiotherapy per week and 1 h of compilation of videos with the results obtained. To detect the correct movement of each of the patients, the Kinect Xbox One device was used. It identifies all body points, alignment, speed and angles during the walk. 25 points of human body in three dimensions are detected in real time by the Kinect, which allows to generate a data collection in real time and more efficiently. With the database of patients, a pre-processing of the information is done to identify the most relevant points for our study. A fuzzy model is generated which determines maximum and minimum thresholds for the posture of the back (angle of inclination), shoulders posture (shoulders inclination with respect to the spine), head posture (inclination with respect to horizontal vision) and movement of arms. The model dynamically identifies which position is correct for the movement during the walk, and in addition, the progress that is generated during a time series. This prototype detector is used for rehabilitation of high-performance athletes and is an approximation for the correct posture during long and medium distance races, jumps, among other sports that use the walk as a basis in their workouts. This study was based on the solution of back problems in clinical patients. These preliminary tests have given excellent results in the testing phase, which validates it as an option to prevent injuries in patients with these conditions.

Keywords: Back · Walk · Fuzzy logic · Kinect

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 299–305, 2020.
https://doi.org/10.1007/978-3-030-20040-4_27

1 Introduction

An appropriate body posture is of the utmost importance to maintain good health, because it facilitates breathing, mobility of legs and arms as well as other parts of the body. A proper walking posture prevents a great percentage of back, neck and shoulder pains, degenerating into several other problems such as lumbago and muscle contractures. Multiple problems in the intervertebral discs are caused by an incorrect posture, because it loosens and pushes the muscles in an imbalanced way.

Currently there are several researches in this regard to improve posture during exercise, a walk or while running. In the article carried out by [1], an evaluation of some methods of head posture measurement was presented by correlating the three head position variables (two horizontal separation variables and one variable of orientation). This study analyzed seven postural conditions with 20 patients with symptoms of poor posture. These measurements were made using a three-dimensional motion capture system in several situations such as sitting normally, sitting with the back straight, standing normally, standing with the back straight, walking at a pace of 4 and 6 km/h and running at a speed of 8 km/h, both on a treadmill. The results obtained from this study show the importance of a standardized evaluation protocol for a more reliable assessment of the forward head posture.

Another study on the posture of the upper body is presented by [2], which determines the effectiveness of the photographic measurement of the postures of the cervical spine and the shoulder and its interrelation with the head posture. The study was carried out with young patients between 15 and 17 years of age of both genders. The main objective is to analyze the different angles used in postures and their differences. The study showed a 58% wrong head posture, related to the shoulders, as well as significantly higher cervical and sagittal head angles.

In the work developed by [3], the walking posture was investigated, and the effects of an upright posture against an erroneous walking posture, as well as psychological and physiological states during the execution of the walk, were analyzed. Seventy-three healthy adults who walked normally and erect were used during the study, with reports of physical effects such as blood pressure, temperature, etc.

Physically, the vertical gait posture group exhibited significantly lower systolic blood pressure, a galvanic skin response and a skin temperature slightly lower than the erroneous gait posture group.

In this paper a model of back posture correction during the walk is proposed using the Xbox Kinect and the artificial intelligence technique of Fuzzy Logic with the body points.

This work is structured as follows: in Sect. 2 the materials used to develop the proposed model, that is, the database applied, are described. On the other hand, the techniques used in the creation of the model are detailed. The modelling process of the back-posture detector is explained in Sect. 3. Section 4 presents the results of the model and compares them with papers already published. Finally, the conclusions obtained in this work and the future lines of research that were born from this study are presented.

2 Materials and Methods

This section details the materials used for the study (equipment and information collected). In addition, the techniques of artificial intelligence used for the development of the model and data processing algorithms are described.

2.1 Materials

The materials used in this article have been the Kinect of the Xbox One and all the information collected with that device.

Kinect Xbox One. The Kinect sensor is a horizontal bar of about 23 cm (9 inches) connected to a small circular base with a motorized pivot and is designed to be placed lengthwise above or below the video screen. The device features an RGB camera, a depth sensor, a multi-array microphone and a custom-made processor that runs the patented software, which provides full-body 3D motion capture, facial recognition and voice recognition capabilities. The sensor contains a motorized tilt mechanism and in case of using an Xbox 360, it has to be connected to a power outlet, since the power that the USB ports can supply is insufficient; with the redesign of the Xbox 360 S this is not necessary because this game console has a specially designed port for accommodating the connector of the Kinect and this allows to provide the power required by the device to function properly [4].

Database. The database was compiled over a period of 2 months using the Kinect Xbox One device, where 2 patients (patient A and B) have made short recordings walking 20 m with incorrect back posture and correct posture. The recordings made were 5 per day of each user, obtaining a total of 200 recordings. The distribution of information is 100 normal records and 100 incorrect records. On the other hand, each of the patients has 100 records, that is, 50 normal records and 50 incorrect records per patient.

The information of each patient is based on the skeleton detected by the Kinect, specifically of the points shown in Fig. 1, which are the back, shoulders, neck and head.

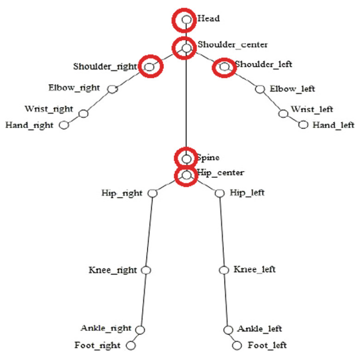


Fig. 1. Skeleton of points detected by the Kinect Xbox One.

The position of each of the points is detected laterally while the camera records the patient during the walk. Figure 2 shows the detection of points during a short walk of patient A and B.

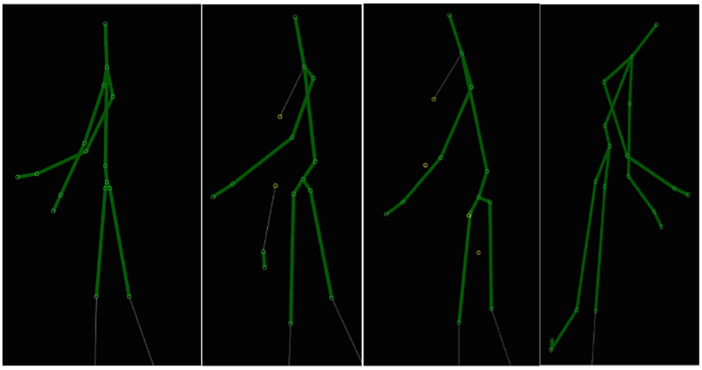


Fig. 2. Skeletons of patients A and B during the walk correctly and incorrectly.

The techniques of artificial intelligence that were applied in this study are presented in the following section.

2.2 Methods

This section describes the artificial intelligence techniques that have been used to generate the back-posture detection model. To determine in a more efficient way the correct position of the skeleton points of a patient, it is necessary to apply the Fuzzy Logic technique.

Fuzzy Logic. The technique of artificial intelligence called fuzzy logic is based mainly on observation from a subjective point of view, that is, it can determine thresholds depending on the decision of the observer [5]. These thresholds are contextualized mathematically to determine periods of confidence for each category (fuzzy sets) selected by the observer. Once the thresholds of these categories have been determined, heuristic rules are generated using the logical functions IF and THEN [6].

Categories are often defined as triangular functions or trapezoidal curves, since each value will have a gradient where the value increases, a peak where the value equals 1 (which can have a length of 0 or more) and a gradient where the value is decreasing. They can also be defined using a sigmoid function. A common case is the standard logistic function defined as:

$$S_{(x)} = \frac{1}{1 + e^{-x}} \tag{1}$$

This technique is widely used in an endless number of real-world applications such as drone engine management, climate prediction, etc.

The preprocessing of the information and development of the proposed model using fuzzy logic to identify the correct and incorrect posture of the patient's back are presented in the next section.

Model of Back Posture Detection During the Walk. In this section we present the fuzzy modelling based on patient walk data for a correct back posture. To identify the correct points of the human skeleton, crucial points and thresholds for body movement during the walk have been identified. According to the study carried out by [3], the head, shoulders and back points should not exceed 10° with respect to the horizontal, as shown in Table 1.

Table 1. Information about correct and incorrect postures of head, shoulders and back.

Head	Shoulders	Back	Posture
$\propto < 7°$	$2° < \beta < 9°$	$\delta < 7°$	Correct
$\propto > 7°$	$\beta > 9°$	$\delta > 7°$	Incorrect

Based on this information, a fuzzy logic model is generated, which determines joint thresholds between head, shoulders and back to determine an optimal detection result. The proposed model is presented in Fig. 3.

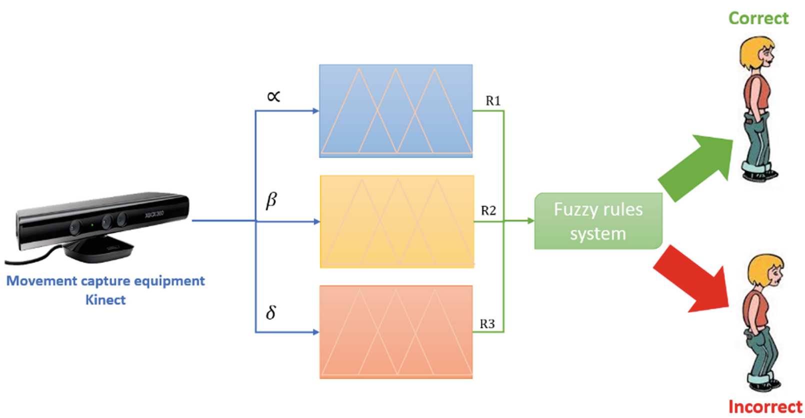


Fig. 3. Implementation of Simulink of the detector of exercises of fuzzy model with Kinect.

The training phase obtained an accuracy of 89% and the testing phase a 90%, considering that it is a controlled environment, as shown in Table 2.

Table 2. Table of results of training and tests of the detection model in shoulder, head and back.

	Shoulder				Head				Back			
	Training		Tests		Training		Tests		Training		Tests	
Patient	CC	IC	CC	IC	CC	IC	CC	IC	CC	IC	CC	IC
Patient A	89,23	90,01	90,42	89,36	86,36	85,76	89,15	87,7	90,22	87,64	87,52	89,98
Patient B	88,98	88,12	90,00	91,05	88,74	88,32	89,00	89,16	88,55	88,49	88,99	90,56
Average	**89,105**	**89,065**	**90,21**	**90,205**	**87,55**	**87,04**	**89,075**	**88,43**	**89,385**	**88,065**	**88,255**	**90,27**

*CC = Correct classified, IC = Incorrectly Classified.

This proposed model was tested in a real environment with 2 people unconnected with the initial database. In the next section the results of the proposal will be presented.

3 Results

The results obtained in the testing phase have been optimal with an average of 92% in the training phase and 91% in the testing phase, as shown in Table 3.

Table 3. Results of the back-posture detection model.

	Training		Tests	
Patient	CC	IC	CC	IC
Patient A	92,13	91,87	91,86	90,04
Patient B	92,56	93,23	90,98	91,44
Average	92,345	92,55	91,42	90,74

These results indicate that the performance of the model is optimal with a reduced database. On the other hand, it has been demonstrated that the selection of attributes for the study has been satisfactory and with an efficient precision.

The conclusions obtained in the study are presented in the following section, as well as possible lines of future research.

4 Conclusions and Future Works

In the present work a model was proposed for the detection of back posture during the walk, where some conclusions have been obtained during the development of the research. First, the use of a Kinect device allows a more efficient capture of the body points of the patient in real time during the execution of the walk. This greatly benefits the collection of real information on the location of each point. Second, the use of specific points in the study allows the model to be more effective and with a more optimal analysis, so it is a fundamental basis for a good result. Third, the application of fuzzy logic for body movements has been an excellent solution, because it can be adaptive to the characteristics required during the tests of the model, as well as being more precise in the detection.

As future research lines, it is intended to use electronic sensors applied to the patient's body, which allows more precise information such as temperature, blood pressure, sweating, etc. Evolutionary neuronal networks will be used to identify more accurately when the patient performs a correct and incorrect movement of a body point.

References

1. Lee, C., Lee, S., Shin, G.: Reliability of forward head posture evaluation while sitting, standing, walking and running. Hum. Mov. Sci. **55**, 81–86 (2017)
2. Ruivo, R.M., Pezarat-Correia, P., Carita, A.I.: Intrarater and interrater reliability of photographic measurement of upper-body standing posture of adolescents. J. Manipulative Physiol. Ther. **38**(1), 74–80 (2015)
3. Hackford, J., Mackey, A., Broadbent, E.: The effects of walking posture on affective and physiological states during stress. J. Behav. Ther. Exp. Psychiatry **62**, 80–87 (2019)
4. Stone, E.E., Skubic, M.: Fall detection in homes of older adults using the microsoft kinect. IEEE J. Biomed. Heal. Inform. **19**(1), 290–301 (2015)
5. Rastelli, J.P., Peñas, M.S.: Fuzzy logic steering control of autonomous vehicles inside roundabouts. Appl. Soft Comput. **35**, 662–669 (2015)
6. Belal, S.Y., Taktak, A.F.G., Nevill, A.J., Spencer, S.A.: A fuzzy system for detecting distorted plethysmogram pulses in neonates and paediatric patients. Physiol. Meas. **22**(2), 397–412 (2001)

Building Hybrid Interfaces to Increase Interaction with Young Children and Children with Special Needs

Janio Jadán-Guerrero[1(✉)], Cesar Guevara[1], Patricio Lara-Alvarez[1], Sandra Sanchez-Gordon[2], Tania Calle-Jimenez[2], Luis Salvador-Ullauri[3], Patricia Acosta-Vargas[4], and Diego Bonilla-Jurado[5]

[1] Centro de Investigación MIST, Universidad Tecnológica Indoamérica, Quito, Ecuador
{janiojadan, cesarguevara, patolara}@uti.edu.ec
[2] Escuela Politécnica Nacional, Quito, Ecuador
{sandra.sanchez, tania.calle}@epn.edu.ec
[3] Centro de Educación Continua, Escuela Politécnica Nacional, Quito, Ecuador
lsalvador@cec-epn.edu.ec
[4] Intelligent and Interactive Systems Laboratory, Universidad de Las Américas, Quito, Ecuador
patricia.acosta@udla.edu.ec
[5] BH Consultores, Ambato, Ecuador
administracion@bhconsultores.com

Abstract. Young children as well as children with special educational needs learn from their environment with social, emotional and physical stimuli. In this context, educational resources and teaching strategies play a main role for them in order to understand the new information. This paper describes the experience of building hybrid interfaces that combine technology with traditional educational resources. A total of 60 teachers divided in two groups completed some tasks which consisted of generating new educative resources with tecnology. Through Design Thinking methodology, teachers designed three hybrid interfaces: 1. Interactive books, combining traditional fairy tales books with mobile devices, where QR codes and NFC tags give life to the stories; 2. Educational Board Games, where augmented reality markers give an extra information to the players; 3. Tangible educational resources, which integrate Makey-Makey device and Scratch with fruit, clay, aluminum foil or water to build laboratory.

Keywords: Hybrid interfaces · Interaction · Young children · Children with special needs

1 Introduction

The sensory stimulation is fundamental during formative brain development of children. They learn from their environment, from way people treat them and from what they see, hear and touch [1]. These experiences prepare children to better integrate themself to the physical and social environment through the development of executive

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 306–314, 2020.
https://doi.org/10.1007/978-3-030-20040-4_28

functions. Executive function is an umbrella term that encompasses the mental abilities that the human being development in behavioral, emotional and metacognitive functioning regulation. It is in the preschool stage where children present a critical and cerebral plastic opportunity [2]. Any cognitive process that is going to develop within a child, is based on the input of sensory information, so the original information must penetrate the sensory channels and be coded quickly in a way that goes into short-term memory [3]. Considering that the multisensory stimulation contemplates the first element of access of a child with his environment, it is indispensable that this process be developed in an integral way and that being this the initial step on which will depend of perceptions that will provide the foundation for learning [4].

Multisensory stimulation uses the senses of sight, hearing, touch, smell and taste to maximize a child's executive function and understanding of the environment. Multisensory environments improve the development of thought, intelligence and social skills, whilst also offering children with cognitive impairments and other challenging conditions, the opportunity to enjoy and control a variety of sensory experiences [5].

In this context, educational resources and teaching strategies play a main role for young children and children with special needs. Among the educational resources used by teachers are fairy tails, cardboard, colored pencils, glue, scissors and clay. Through them, teachers generate interaction and collaboration through stories, games and recreational activities.

Nowadays as technology advances, new educational resources are being integrated into the classroom, such as the computer, projector, interactive whiteboards, tablets and smart phones, which are giving rise to new pedagogical trend where we can find e-learning, inverted classroom, gamification, serious games, laboratories or remote laboratories, makers spaces or ubiquitous education.

These revolutionary technologies open up space for new ways of learning and are changing the teaching-learning process in classroom. However, in developing countries such as Ecuador, using these new technological and pedagogical tendencies is not common, not even the existing ones are fully exploited due to the fact that teachers ignore how to use them or the lack of adequate teaching strategies for a generation of students with technological skills. This problem releases an opportunity to develop technologies that will close the generational gap and also develop didactic resources for classroom.

This article describes the experience with 60 primary and secondary education teachers, who designed hybrid interfaces that combine traditional educational resources with technological ones, with the purpose of strengthening the interaction in learning of young children and children with special educational needs.

The rest of this article is structured as follows: Sect. 2 details the background and related work, Sect. 3 presents the method used, Sect. 4 presents the results and discussion, and finally, Sect. 5 presents the conclusions and future work of this research.

2 Background

In recent years, a new generation of hybrid interfaces for learning are gaining in popularity within classrooms. These spaces have been digitally augmented by replacing blackboards and tables with computer-augmented surfaces provided with multitouch interaction [6]. The use of conventional resources with technology has been conducted on new trends in pedagogy and technology. Trends in pedagogy are: Active Learning, Problem-based Learning, Project-based Learning, Collaborative learning, Learning based on challenges, flipped classroom, Space makers and Gamification. Trends in technology are: Adaptive Learning, Learning in Social Networks and Collaborative Environments, Mobile Learning, Ubiquitous Learning, Virtual Assistant, MOOCs, Big Data and Learning Analytics, Remote and Virtual Labs, Internet of Things, Augmented reality and Virtual Reality [7].

In our study we integrated some of these trends in a Moodle Learning Management System, in which we applied the Methodology of Learning Based on Narrative Metaphors and Gamification [8]. Other pedagogies applied were Collaborative learning, and Learning based on challenges through forums and group activities. Augmented Reality through apps, such us Animal 4D+ , Space 4D+ , Humanoid 4D+ and QuiverVision 3D Augmented Reality. Finally, Space Makers using Tangible User Interfaces through RFID tags integrated with physical objects.

3 Method

This research work was based in Design Thinking methodology, defined as an analytic and creative process that engages a person in opportunities to experiment, create and prototype models, gather feedback, and redesign. The five stages of Design Thinking are as follows: Empathise, Define (the problem), Ideate, Prototype, and Test [9].

3.1 Participants

The sample of the study was selected from the Master's Degree in Education, Innovation and Educational Leadership of the Universidad Tecnológica Indoamérica, located in the city of Ambato, Ecuador. We worked with n = 60 students. The participants were divided into two groups of 30 each in the Educational Infopedagogy module. The average age of the students was 35.5 years, 44.4% corresponded to the range of 36 to 40 years, followed by 18.5% in the range of 41 to 45 years, 14.8% between 26 to 30 years, 11.1% in the range of 31 to 35 years and the rest in ages under 25 and over 50. In relation to gender 90% corresponded to the female. The training of 75% of teachers was concentrated in basic education, followed by 20% in secondary education and 5% in higher education.

3.2 Materials

The materials used in the research are grouped into three categories: classroom materials, technology and software. In the first category we used paper, cardboard, colors, markers, watercolors, plasticine and aluminum foil. In the second category we used laptops, tablets, smartphone, RFID sensors and Makey Makey circuit board. In the third category we used Scratch, AppInventor, QR Generator, Hi-Q, Animal 4D+, Space 4D+, Humanoid 4D+ and QuiverVision 3D Augmented Reality.

3.3 Procedure

We worked with two groups of 30 students in the module Educational Infopedagogy, which aimed at the introduction of technology in the classroom and how it is integrated with the pedagogy. Five weeks were worked with each group, with an eight-hour session each week.

Following the Design Thinking methodology the teachers worked in groups related by the affinity of each teacher. In the empathise stage, teachers used the platform in six areas: (1) STEAM, (2) language and literature, (3) social problems, (4) learning strategies, (5) Educational Leadership and (6) Innovation and entrepreneurship. In each area, teachers defined the problem and needs of their educational centers through a survey and forums in the framework. Next they ideate the design of the pedagogical strategy according to the technologies used in class. After teachers design a prototype with the materials and technology selected. Finally, teachers test an assessment of learning using interactive and co-evaluation methods with their students.

4 Design of Hybrid Interfaces

In Computer Science an hybrid interface is the interface getting embedded in the objects of the physical world. The latter is perceived as an augmentation of the physical world [9, 10]. This section shows some the hybrid interfaces designed by the groups of teachers. We have divided into three categories: Interactive Books, Educational Board Games and Tangible Educational Resources.

4.1 Interactive Books

Within the category of interactive books, teachers created innovative interfaces for children. The participation and interaction by the reader was made with QR codes. The main idea of this interface was to introduce narrative an questions thought QR codes. First with the Hi-Q app recorded audio in MP3 format. Subsequently teachers generated QR codes of different sizes, which were pasted them in a physical book or they designed a sheet of cardboard with their own story as the Fig. 1 shows. The audio files and pictures were integrated in App Inventor application. In this interface children can use a smartphone to listen a story or answer a question that generates the QR code.

Another interface designed by other group were large sheet drawings that were pasted on the walls of the classroom, each sheet had a Qr code. The objective was the children search with smartphones te QR codes and listen the story or questions.

Finally, a group designed an Educative Twister, a game of physical skill. Children use a smartphone to scan QR codes. Each QR code generates a question and a challenge on the board of twister.

In the stage of testing, teachers evaluated these hybrid interfaces with their students achieving interest and motivation in the learning process. With this technology teachers could design their own educational technology resources to create multimedia reading experience [11].

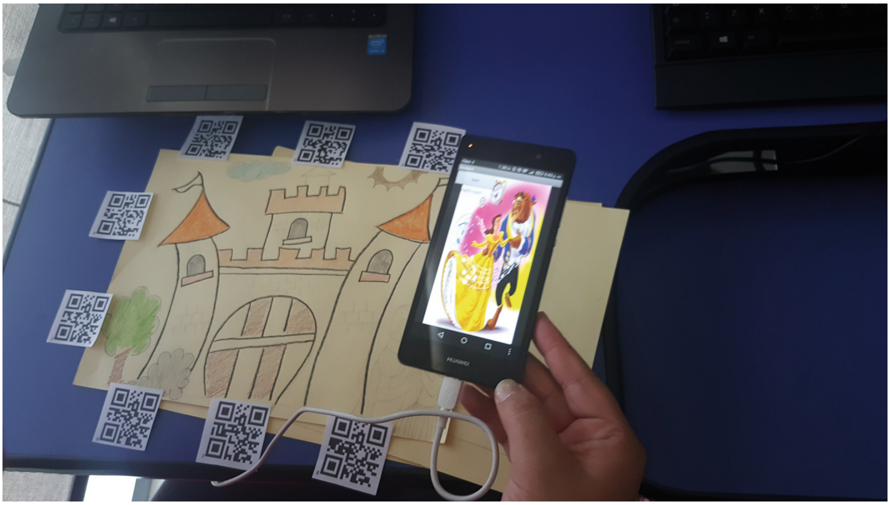


Fig. 1. Story drawn on cardboard with QR codes.

Another idea was based on Leapfrog's book which have enhanced the book surroundings, permitting pen interaction with the book [12]. To simulate this interface, teachers used NFC (Near Field Connection) tags and QR codes, which generate a sound, narrative or question throughout history. Figure 2 shows an example of the interactive book "The Little Prince".

Fig. 2. Interactive book with NFC and QR codes.

The use of NFC tags facilitated the interaction of children with special needs, unlike the QR codes that need more precision with the camera of the smartphone.

4.2 Educational Board Games

In this category teachers designed a new concept of a board game "The inclusive board game" that aims to enable children with especial needs play with other members of their family through activities supported by technology. The board game has traditional elements and in certain cases it is integrated with a tablet to see elements of the game with augmented reality, or smartphones that read NFC tags. For example, a card with Braille codes seen through a tablet can generate an augmented reality with its meaning. Other cards have QR codes that provide additional information or questions on the theme of the game. Teachers designed the hybrid interface "The route of the volcanoes" in which players can see a volcano in augmented reality and information that helps the entertainment and learning. We believe that it is an innovative idea that uses disruptive technology rarely seen in our environment and applicable to a vulnerable and little forgotten population.

The idea to design an inclusive board game was based on The Explorers, a board game of Unicorn Games. Figure 3 shows the idea of this board game.

Fig. 3. Inclusive board game.

The inclusive board game has the classic dynamics of a board game in terms of the distribution of turns, use of dice to advance and interaction between players around the board. The board game joins the logic of the traditional with the world of applications and carries the Tangible User Interface (TUI). The board game utilises traditional board game play and combines it with mobile and app technology that uses video games and card scanning. The corresponding app includes mini-game challenges, treasure hunts and AR objects.

4.3 Tangible Educational Resources

For tangible educational resources teachers were trained in the use of Makey-Makey and Scratch. Makey-Makey is a circuit board that allows users to connect some objects to computer programs using alligator clips and a USB cable. This board simulates the actions of the arrow of a keyboard or mouse buttons instead of pressing the keys or buttons. For the graphical interface, teachers used the Scratch application, which is an introductory programming language that enables young children to create their own interactive stories, collages, and games [13]. One of the hybrid interface considers the laterality learning with dance-dance revolution music. Figure 4 shows the hybrid interface with cardboard and foil.

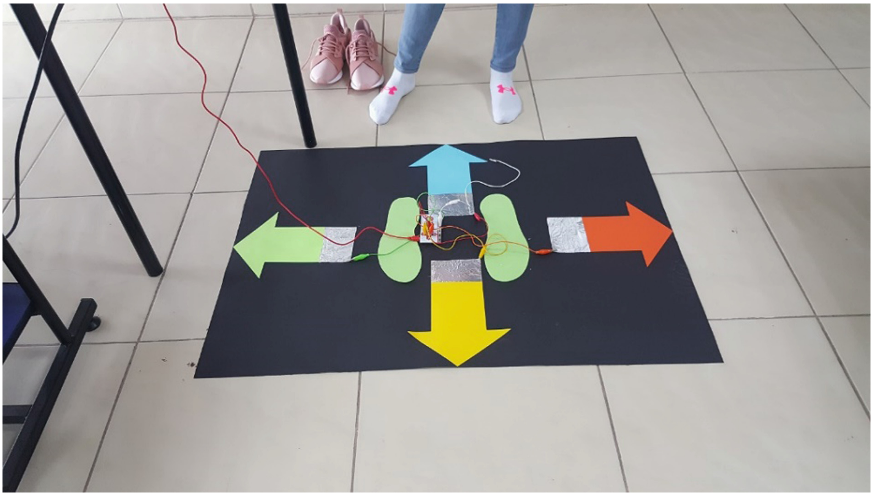


Fig. 4. Dance dance revolution to teach laterality

In language and literature area teachers designed a board built with Scratch and Makey-Makey to teach phonological awareness. The child creates some object with plasticine connected with a computer through Makey-Makey. On the other hand, Scratch shows a pictogram and the sound of the phoneme associated with the clay object. The application of this hybrid interface with children with special needs can improve the reading and spelling [14]. The board is shown in Fig. 5 below:

In the area of natural sciences a group designed a similar board with organs of the human body made with plasticine. When children touch an organ the information is showed on screen with Scratch.

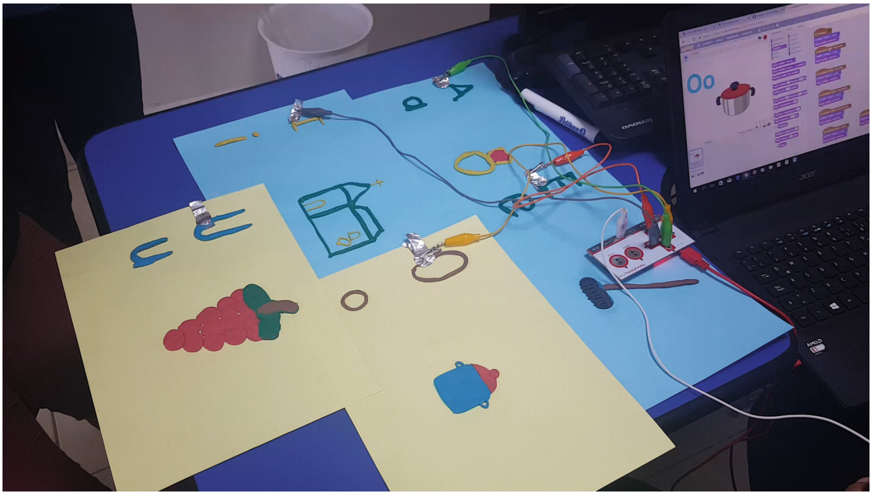


Fig. 5. Language board with Makey-Makey, Scratch and clay

5 Results and Discussion

The inclusion of children with disabilities is still a problem in Ecuador, in the educational, social and even family environment. In this sense, strategies to develop creativity and innovation through design thinking and learning based on challenges allowed to generate prototypes of educational resources supported by technology. This experience was very rewarding, since the participants were highly motivated and developed creative and innovative prototypes.

Another problem in Ecuador is the lack of technological resources in classrooms and the lack of knowledge on the part of the teachers. The introduction of technology to the classrooms is a challenge, but if it is done in an easy way, it allows the teachers create new educational strategies.

6 Conclusions

From the experience gained during the design process and the final evaluation of hybrid interfaces, teachers were motivated to design their own educational resources. The use of the hybrid interfaces allowed that young children and children with special needs were motivated.

With the use of basic apps and traditional resources, teachers can design new educational technologies resources selecting the best digital strategies to allow students enrich learning with activities that incorporate devices. The design of digital content allows the use of interactive tools that keep students' attention more easily and in this way achieve meaningful learning [15].

Technology is an essential factor in motivating users to collaborate with their learning process [16]. For future work, we are planning testing the hybrid interfaces in a real environment, in order to improve and customize the current interfaces to the particular needs. In the future we are planning to design hybrid interfaces for children with disabilities, such as blind or low vision children, deaf children, children with motor and cognitive disabilities.

Acknowledgments. The authors thank Universidad Tecnológica Indoamérica for funding this study through the project "TIME- ICCHIT" and students of MEILE-3.

References

1. Vosniadou, S.: How children learn. International Academy of Education (2001)
2. Ramos, C., Bolaños, M., Paredes, L., Ramos, D.: Tratamiento Neuropsicológico del TDAH en Preescolares: Entrenamiento de la Función Ejecutiva. Revista Ecuatoriana de Neurología **25**(1–3), 61–69 (2016)
3. Ramos-Galarza, C., Pérez-Salas, C.: Moderator role of monitoring in the inhibitory control of adolescents with ADHD. J. Attention Disord. 1–11 (2018)
4. Price S., Duffy S., Gori, M: Developing a pedagogical framework for designing a multisensory serious gaming environment. In: Proceedings of 1st ACM SIGCHI International Workshop on Multimodal Interaction (MIE'2017), Glasgow, UK, 8 p., November 2017 https://doi.org/10.1145/3139513.3139517
5. Cid, M.J., Camps, M.: Multisensory stimulation in the snoezelen room: concept and fields of application. Revista Española sobre Discapacidad Intelectual **41(4)**(236), 22–32 (2010)
6. Cerezo, E., Marco, J., Baldassarri, S.: Hybrid games: designing tangible interfaces for very young children and children with special needs. In: Nijholt, A. (ed.) More Playful User Interfaces. Gaming Media and Social Effects. Springer, Singapore (2015)
7. EduTrends 2017: Observatorio de Innovación Educativa, Reporte Edu Trends. Monterrey. Tecnológico de Monterrey, México (2017)
8. Jadán-Guerrero, J., Ramos-Galarza, C.: Metodología de Aprendizaje Basada en Metáforas Narrativas y Gamificación Un caso de estudio en un Programa de Posgrado Semipresencial. Hamut´ay **5**(1), 84–104 (2018). https://doi.org/10.21503/hamu.v5i1.1560
9. Razzouk, R., Shute, V.: What is design thinking and why is it important? Rev. Edu. Res. **82** (3), 330–348 (2012). https://doi.org/10.3102/0034654312457429
10. Reimann, D., Herczeg, M., Winkler, T.: Gaining computational literacy by creating hybrid aesthetic learning spaces. In: Proceedings 3rd IEEE International Conference on Advanced Technologies (2003). https://doi.org/10.1109/icalt.2003.1215135
11. Grasset R., Duenser A., Seichter H., Billinghurst M.: The mixed reality book: a new multimedia reading experience. In: CHI 2007 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2007), 1953–1958. ACM, New York (2007). https://doi.org/10.1145/1240866.1240931
12. Leapfrog https://www.leapfrog.com. Accessed Jan 2019
13. Strawhacker, A., Lee, M., Caine, C., Bers, M.U.: ScratchJr demo: a coding language for kindergarten. In: Proceedings of the 14th International Conference on Interaction Design and Children (IDC 2015). ACM, Boston (2015)
14. Fan, M., Antle, A.N., Hoskyn, M., Neustaedter, C., Cramer, E.S.: Why tangibility matters: a design case study of at-risk children learning to read and spell. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, USA, içinde (ss. 1805–1816) (2017). https://doi.org/10.1145/3025453.3026048
15. Kara, N.: Design, development and use of a smart toy for preschool children: a design and development research (Yayımlanmamış doktora tezi). Orta Doğu Teknik Üniversitesi, Ankara (2015)
16. Linke, R., Kothe, T., Alt, F.: TaBooGa: a hybrid learning app to support children's reading motivation. In: Proceedings of the 2017 Conference on Interaction Design and Children, USA, içinde (ss. 278–285) (2017). https://doi.org/10.1145/3078072.3079712

Implementation of Controls for Insertion of Accessible Images in Open Online Editors Based on WCAG Guidelines. Case Studies: TinyMCE and Summernote

Sandra Sanchez-Gordon[1](✉), Tania Calle-Jimenez[1], Jorge Villarroel-Ramos[1], Janio Jadán-Guerrero[2], César Guevara[2], Patricio Lara-Alvarez[2], Patricia Acosta-Vargas[3], and Luis Salvador-Ullauri[4]

[1] Department of Informatics and Computer Science, Escuela Politécnica Nacional, Quito, Ecuador
{sandra.sanchez, tania.calle, jorge.villarroel}@epn.edu.ec

[2] MIST, Universidad Tecnológica Indoamérica, Quito, Ecuador
{janiojadan, cesarguevara, patolara}@uti.edu.ec

[3] Intelligent and Interactive Systems Laboratory, Universidad de Las Américas, Quito, Ecuador
patricia.acosta@udla.edu.ec

[4] Centro de Educación Continua, Escuela Politécnica Nacional, Quito, Ecuador
lsalvador@cec-epn.edu.ec

Abstract. It is important to raise awareness among people who create web content about barriers that exist for people with visual disabilities to perceive images. This study had for purpose the implementation of features and controls for insertion of accessible images in two open online editors: TinyMCE and Summernote. TinyMCE is used in content management systems such as WordPress, Joomla and Drupal. Summernote is an open online editor distributed under an open MIT license. The accessibility of TinyMCE was improved in the insertion of images by adding entry fields for the type of image to be inserted, long description and title. As for Summernote, in the option to insert images, entry fields were added for the type of image to be inserted, alternative text, long description and title. Controls were implemented to validate mandatory and optional fields according to the type of image. The purpose is that open online editors comply with Part B of the Authoring Tool Accessibility Guidelines. A discussion is presented about how the original code of the two open online editors was written, how image accessibility was improved and how evaluation was performed via test scenarios, automated tools and usability surveys to both content authors and blind users. The HTML code generated with both editors showed compliance with the Web Content Accessibility Guidelines. Blind users stated that they were able to perceive the images when interacting with web pages created with both editors.

Keywords: Blindness · Images accessibility · Web content accessibility guidelines · WCAG · Authoring tool accessibility guidelines · ATAG · Online editors · TinyMCE · Summernote

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 315–326, 2020.
https://doi.org/10.1007/978-3-030-20040-4_29

1 Introduction

The Web is a very important resource for every person in several aspects, such as education, employment, government services, commerce, health and entertainment. The Web must be accessible to provide equal access and equal opportunities to people with disabilities [1]. The World Wide Web Consortium (W3C) states that web accessibility means that people with some type of disability, e.g. visual, auditory, cognitive, motor, can make use of the Web to perceive, understand, browse and interact with web content [2]. The software used to create web content has an important role in this aspect since it must guide content authors to generate accessible web pages so that people with disabilities can use them.

The open online editors are widely used by content management systems. Unfortunately, as reported in [3], most of these open online editors suffer from lack of features and controls to guide content authors in the creation of accessible web pages. Two of the most popular open online editors are TinyMCE [4] and Summernote [5]. Online editors can be improved by implementing features and controls for the insertion of accessible images based on the Web Content Accessibility Guidelines (WCAG) [6] so they comply with the Authoring Tool Accessibility Guidelines (ATAG) [7].

The rest of this article is structured as follows: Sect. 2 details background and relevant published research work. Section 3 presents the method used in this study. Section 4 presents the results and discussion. Finally, Sect. 5 presents conclusions and future work.

2 Background

Accessibility of images means that the information conveyed by an image can be perceived by users with and without disabilities in the same way. Accessibility of images also improves the searchability, since the alternative textual information provided is visible to search engines [8].

Currently, the WCAG are an internationally accepted reference for web accessibility, including accessibility of images. The Web Accessibility Initiative (WAI) of the W3C published version 1.0 of the WCAG in 1999. The WCAG 1.0 had a total of 15 guidelines [9]. In 2008, the WCAG 2.0 guidelines were approved as an official W3C recommendation [10]. In 2012, WCAG 2.0 was accepted as the standard ISO/IEC 40500:2012 [11]. Unlike the WCAG 1.0 guidelines that were developed exclusively for the web technologies that existed in 1999, e.g. HTML, CSS, JavaScript, the WCAG 2.0 guidelines were developed in a technologically neutral way. WCAG 2.0 consists of four principles, 12 guidelines and 61 success criteria. There are three compliance levels: A, AA and AAA. The WCAG 2.0 four principles are: perceptible, operable, understandable and robust [10]:

- Principle 1. Perceptible. Information and components of the user interface must be shown to the users in ways that they can understand.
- Principle 2. Operable. Components of the user interface and navigation must be manageable.

- Principle 3. Understandable. Information must be clear.
- Principle 4. Robustness. Information formats and components of the user interface must be well interpreted by a wide variety of user agents, including assistive technologies.

In 2017, the first working draft of the WCAG 2.1 was published. After more than one year of work, in June 2018, the definitive version of the WCAG 2.1 became a W3C official recommendation. The aim of this version of the WCAG is to improve the accessibility guidelines for three specific groups of users: people with cognitive or learning disabilities, people with low vision and people with disabilities who access the Web from mobile devices. These guidelines also help make content more usable for older people, who see their skills reduced because of age. In the WCAG 2.1, 17 new criteria have been included: five of level A, seven of level AA and five of level AAA. The WCAG 2.1 has an additive approach designed to show that, if a web page complies with the WCAG 2.1 guidelines, it is also complying with the WCAG 2.0. Therefore, all the compliance criteria of the WCAG 2.0 guidelines are included in the WCAG 2.1 guidelines. This is very important so that websites that are updated to the WCAG 2.1 guidelines do not lose their compliance with the WCAG 2.0 [12]. The relevant WCAG guidelines for the accessibility of images are detailed in Table 1, as identified in [13].

Table 1. WCAG Guidelines for Image Accessibility

WCAG guidelines	Success criteria and levels
Guideline 1.1 Text Alternatives. Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language	Success criterion 1.1.1 Non-text Content. (Level A). All non-text content that is presented to the user has a text alternative that serves the equivalent purpose
Guideline 1.4 Distinguishable. Make it easier for users to see and hear content including separating foreground from background	Success criterion 1.4.5 Images of Text. (Level AA). If the technologies being used can achieve the visual presentation, text is used to convey information rather than images of text
	Success criterion 1.4.9 Images of Text (No Exception). (Level AAA). Images of text are only used for pure decoration or where a particular presentation of text is essential to the information being conveyed

In the literature review, researchers found a limited number of published studies on evaluating image accessibility in web content. Much less literature has been produced regarding solutions specifically aimed to help web content authors to comply with image accessibility guidelines.

In [14], authors define a method for evaluating the accessibility of online content editors using WCAG 2.0 and ATAG 2.0 Part B. The method includes accessibility features that should be met by the images, headings and tables inserted through an online content editor. The results of this study provide criteria for those people who have the responsibility of selecting an accessible online content editor.

In [15], authors explain that screen-reader users access images on the Web using alternative text delivered via synthetic speech. These authors identify the WCAG guidelines relevant to image accessibility. They also present the results of an experiment that compares audemes, a type of non-speech sounds, with alternative text delivered using synthetic speech.

In [16], authors state that authoring tools facilitate the generation of contents. These contents should comply with accessibility features. This study present the results of a comparative analysis among four authoring tools, based on accessibility evaluations that were carried out in the content that were created in each of them, which allowed identifying the issues that the tools have regarding accessibility.

In [8], authors present the results of an evaluation of the accessibility issues of Studio, the authoring component of edX platform, based on ATAG 2.0. Authors explain that ATAG Part B establishes that authoring tools must support and guide content authors in the production of web content that is accessible and conforms to the WCAG. An ATAG compliant authoring tool should prompt the content author for accessibility-related information and provides ways to verify that the generated content is accessible. Authoring tools include online editors, multimedia authoring tools, blog sites, wiki sites, social networking sites. ATAG's main intended audience is authoring tool developers.

3 Method

In this study, the SCRUM framework was used to implement and test the improvements in two open online editors: TinyMCE and Summernote. Ten user stories and three sprints were planned and executed.

Table 2. Scrum product backlog

ID	Story	Sprint
US1	User interface design for TinyMCE	1
US2	Controls implementation for TinyMCE	1
US3	User interface design for Summernote	2
US4	Controls implementation for Summernote	2
US5	Manual testing and automated testing with accessibility evaluation tools for TinyMCE	3
US6	Manual testing and automated testing with accessibility evaluation tools for Summernote	3
US7	User testing of TinyMCE	3
US8	User testing of Summernote	3
US9	Blind/blindfolded users testing of content generated with TinyMCE	3
US10	Blind/blindfolded users testing of content generated with Summernote	3

The goal of the first sprint was to develop the new design of the TinyMCE interface for inserting images and implement the controls according to the type of image to be inserted.

The goal of the second sprint was to develop the new design of the Summernote interface for inserting images and implement the controls according to the type of image to be inserted.

The goal of the third sprint was to carry out a four step validation: manually executed test scenarios, tests with automated accessibility evaluation tools, a usability survey with users of the editors and a user experience survey with a focus group of blind and blindfolded users.

Table 2 presents the product backlog.

4 Results and Discussion

In this section, two study cases are presented. For each study case, there is a comparison of the original interface for inserting images and the improved interface. At the end of the section, the evaluation results are presented.

4.1 Study Case # 1. TinyMCE

TinyMCE is an open online editor that works completely in JavaScript and is distributed free of charge under the LGPL license created by Ephox Corporation. Being based on JavaScript, TinyMCE is independent of the platform and runs on any internet browser. TinyMCE is used in content management systems such as WordPress, Joomla, Drupal, Plone and Silverstripe [4].

Figure 1 shows sample input data in the interface of the option that is displayed when selecting the insert image button in the TinyMCE editor. This interface contains two tabs: General and Advanced. The General tab contains input fields for image URL address, image title and image dimensions. In the Advanced tab, there are fields for style, vertical and horizontal spaces and border. It is worth mentioning that the title field in this interface is in fact the alternative text that will contain the image when inserted.

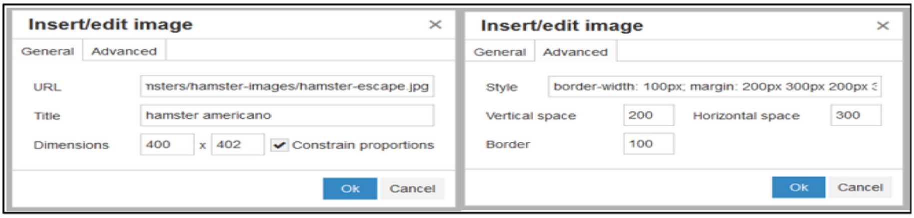


Fig. 1. Original interface for inserting images – TinyMCE

Figure 2 shows the HTML generated code which includes the following attributes for the *img* tag: style, margin, src, alt, width and height. It is worth noting that the input data from field Title is used as parameter of the alt attribute, which is the only accessibility metadata provided for TinyMCE original interface.

Fig. 2. HTML code generated through original interface – TinyMCE

Figure 3 shows the improved interface of the option that is displayed when selecting the insert image button in the TinyMCE editor. It can be seen that it contains only the General tab, the reason why the Advanced tab is not visible is because our focus was the accessibility of the inserted image. The General tab contains the following attributes: image URL, image type, alternative text, long description, title and dimensions fields.

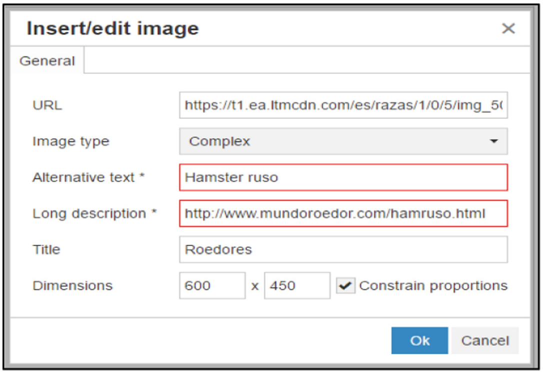


Fig. 3. Improved interface for inserting images - TinyMCE

Figure 4 shows the improved HTML generated code which includes the following attributes for the *img* tag: title, src, alt, width, height and longdesc.

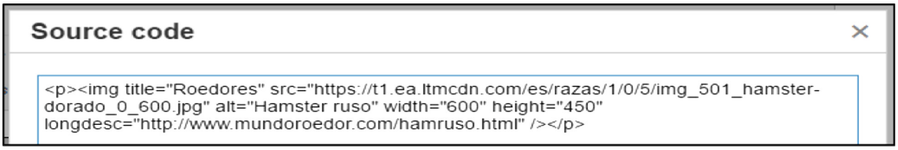


Fig. 4. HTML code generated through the improved interface – TinyMCE

4.2 Study Case # 2: Summernote

Summernote is an open online editor based on Bootstrap and jQuery that works entirely in JavaScript and is distributed under an MIT license created by Summernote Team [5].

Figure 5 shows the interface of the option that is displayed when selecting the insert picture button in the Summernote editor. It contains the Select from files field to enter an image file and Image URL field for the image.

Fig. 5. Original interface for inserting images – Summernote

Figure 6 shows the HTML code generated which includes the following attributes for *img* tag: src and style. No accessibility metadata is present in this code.

Fig. 6. HTML code generated through the original interface – Summernote

Figure 7 shows the improved interface of the option that is displayed when you select the insert picture button in the Summernote editor. The form contains the following fields: Image URL, Type Image, Alternative text, Long description and Title.

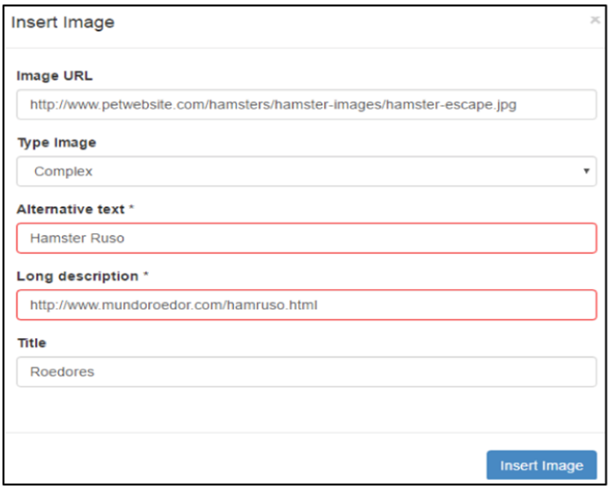


Fig. 7. Improved interface for inserting images – Summernote

Figure 8 shows the improved HTML code which includes the following attributes for the *img* tag: src, alt, longdesc, title and style.

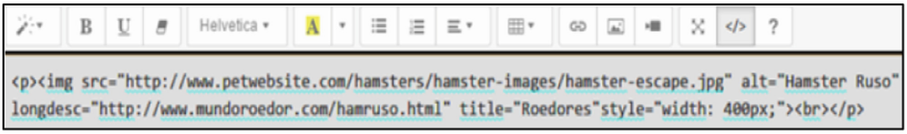


Fig. 8. HTML code generated through the improved interface – Summernote

4.3 Validation

Figure 9 shows the results of a sample of the tests executed to validate the controls on the improved interface of TinyMCE. A set of eight test scenarios were defined. These scenarios are an extension of the list of warnings presented in [8] for conditions potentially wrong regarding the alternative text:

1. If the user does not select the source of the image, the following notification is displayed: "You must input an image source."
2. If the user does not select image type, the following notification is displayed: "You must select a type of image."
3. If the user selects Decorative type, the alternative text field is marked as mandatory and the other fields are optional.
4. If the user selects a Simple type, the alternative text field is mandatory, long description and title are optional, the following notification is displayed: "You must input an alternative text for a simple image. If this image does not convey information, change the image type to Decorative."
5. If the user selects Complex type, the alternative text and long description fields are required, the title field is optional, the following notification is displayed: "You must input an alternative text and a long description for a complex image."
6. If the user selects Functional type, no field is activated and the following notification is displayed: "For a functional image, you must use the specific toolbar button to add functionality."
7. In the alternative text, check the chain to verify and notify the following anomalies:
   a. Verify that the alternative text is not very short according to a threshold defined in 6 characters minimum, the following notification is displayed: "Alternative text is too short to be acceptable."
   b. Verify that the alternative text is not very long, according to a threshold defined in maximum 144 characters, the following notification is displayed: "Long alternative text."
   c. Verify that the alternative text does not contain one or more invalid words such as jpg, png, gif, photo, image, and logo. The use of the XML file makes possible that in the future more words are included that must be considered automatically, the following notification is displayed: "Suspicious alternative text, Alternative text is the image file name, e.g. alt = "file.jpg", or contains generic words such as "photo", "image", and "logo"".

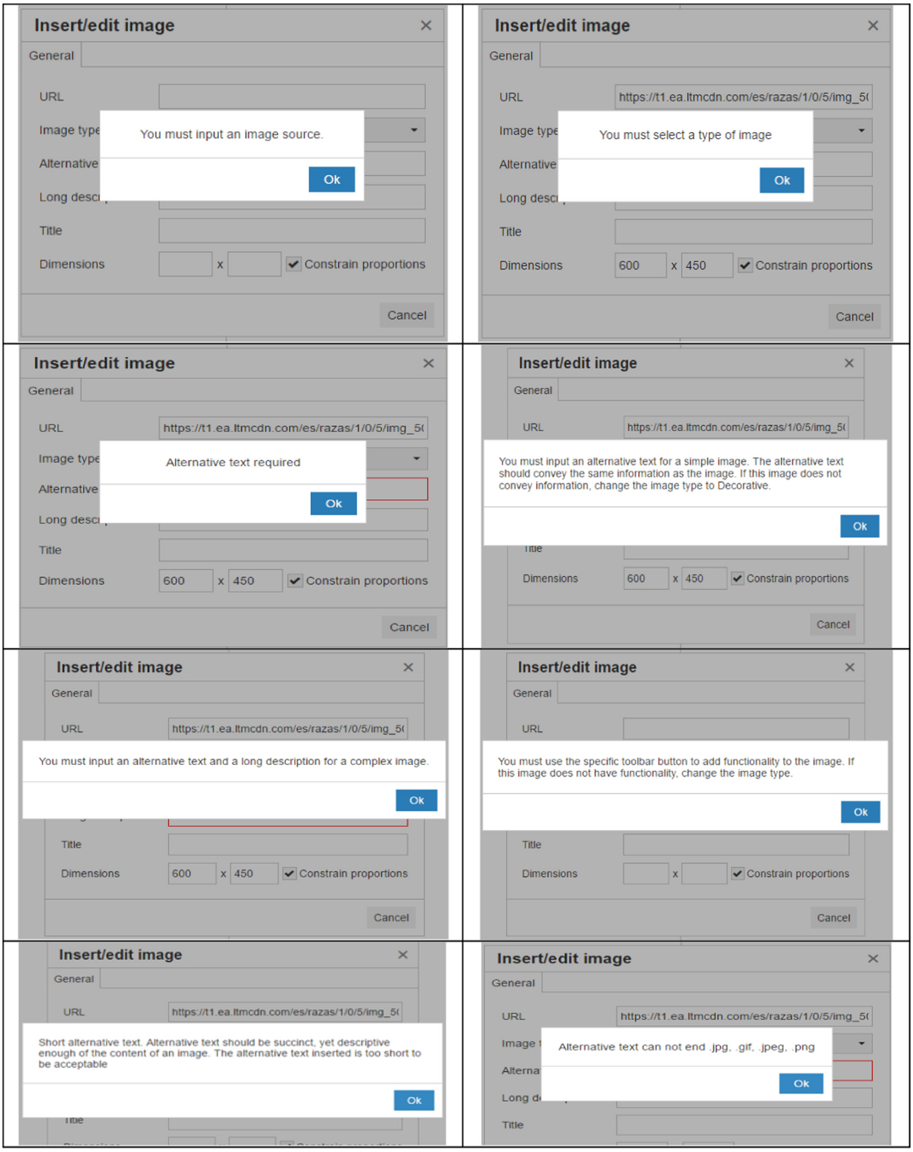


Fig. 9. Controls in the improved interface for inserting images - TinyMCE

8. In the long description, check the field to verify and notify the following anomalies:
   a. If the string entered in longdesc does not end in HTML or HTM, the following notification is displayed: “The long description must be valid URL ending in HTM or HTML.”

b. If the string entered in longdesc has 2 or more blank spaces, the following notification is displayed: "The long description must be a valid URL without blank spaces."
c. If the string entered is not a valid URL. That is, it is tested on the internet and that page does not exist, the following notification is displayed: "The long description must be an existing URL."

Positive results were presented when evaluating the HTML code of a web page generated with both editors using automated accessibility evaluation tools. The automated evaluation indicated that the generated HTML code complied with WCAG guidelines of web accessibility for images in both cases. Figure 10 shows the results presented for the tool WAVE stating that the generated HTML code has no errors.

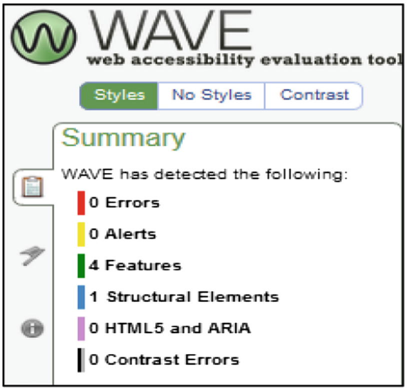


Fig. 10. WAVE accessibility evaluation

For the evaluation with content authors, a focus group of 12 participants responded to the question: Did the warning and error messages help to enter an accessible image correctly? Figure 11 shows the comparative results for both online editors. To the left, for TinyMCE, responses were: Strongly disagree: 0%, Disagree: 0%, Agree: 33.3%, Strongly agree: 66.7%. To the right, for Summernote, responses were: Strongly disagree: 0%, Disagree: 0%, Agree: 16,7%, Strongly agree: 83.3%.

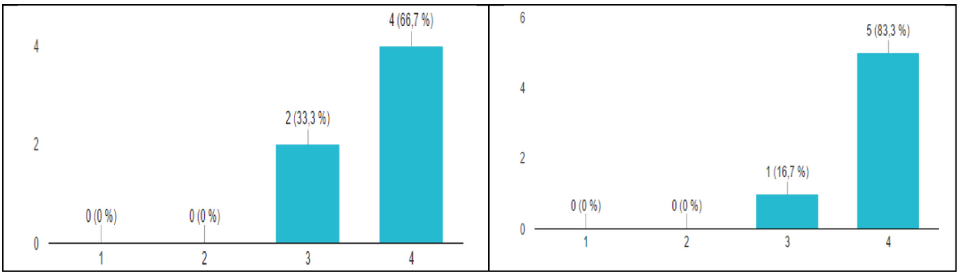


Fig. 11. Comparative results of content author's evaluation

The survey to evaluate the user experience of the HTML file generated with the online editors was carried out with the support of a second focus group that consisted of 14 participants, two blind people and 12 people who were blindfolded. A 92.9% of respondents said that they could distinguish the decorative, simple and complex images within the web page. When asked to explain what they understood of a decorative image presented, most guessed that it was a logo, because it contained an adequate alternative text. Similarly, the alternative text, title and long description of the complex image helped the users to correctly perceive the image.

5 Conclusion

There is a shortage of tools that help authors making their web contents more accessible, so that people with disabilities can have greater access to information. Through the implementation of features controls for the insertion of images in TinyMCE and Summernote editors, it was possible to achieve compliance with the WCAG guidelines on the web pages generated.

The implementation of features and controls for the insertion of accessible images for the case studies was solved in different ways because the development of the source code for the editors was different. While the editor TinyMCE has a well-defined structure, Summernote has all its coding in a single file.

The alternative text field in decorative images was defined as mandatory by blind users, although WCAG does not establish it as mandatory. Hence, the opinion of the final beneficiaries was considered, but in turn the semantic difference between a decorative image and a simple image was maintained.

When interacting with a web page that contains images, the screen reader used by users with visual disabilities ignores that there is an image in the absence of alternative text. It is recommended to keep the alternative text simple, short and concise so that users with visual disabilities can easily perceive and understand images of any type that are included in web pages.

When interacting with the web page elaborated with the improved interfaces of the TinyMCE and Summernote editors, users with disabilities stated that they could perceive the content of the image through the alternative text, title and long description included by the author of the content.

For future work, authors plan to design and implement additional controls and features to increase de accessibility of the web content generated by both TinyMCE and Summernote.

Acknowledgments. The authors thank Jenny Quiguango and Christian Chiguano for developing and testing the code to implement the improvements in both open online editors.

References

1. W3C: Accessibility (2018). https://www.w3.org/standards/webdesign/accessibility
2. WAI: Introduction to Web Accessibility (2018). https://www.w3.org/WAI/fundamentals/accessibility-intro/

3. Villarroel-Ramos, J., Sanchez-Gordon, S., Luján-Mora, S.: Architectural metamodel for requirements of images accessibility in online editors. In: Proceeding of International Conference on Information Systems and Computer Science (INCISCOS), pp. 312–319. IEEE (2018)
4. Tiny: TinyMCE (2019). https://www.tiny.cloud/
5. Summernote: Summernote Super Simple WYSIWYG Editor in Bootstrap (2019). https://summernote.org/
6. W3C: Web Content Accessibility Guidelines (WCAG) Overview (2018). https://www.w3.org/WAI/standards-guidelines/wcag/
7. W3C: Authoring Tool Accessibility Guidelines (ATAG) 2.0 (2015). https://www.w3.org/TR/ATAG20/
8. Sanchez-Gordon, S., Estevez, J., Luján-Mora, S.: Editor for accessible images in e-Learning platforms. In: Proceedings of Web for All Conference (W4A), pp. 1–2. ACM (2016). https://doi.org/10.1145/2899475.2899513
9. W3C: Web Content Accessibility Guidelines 1.0 (1999). https://www.w3.org/TR/WCAG10/
10. W3C: Web Content Accessibility Guidelines (WCAG) 2.0 (2008). https://www.w3.org/TR/WCAG20/
11. ISO: ISO/IEC 40500:2012 (W3C) Information technology - W3C Web Content Accessibility Guidelines (WCAG) 2.0 (2012). https://www.iso.org/standard/58625.html
12. W3C: Web Content Accessibility Guidelines (WCAG) 2.1 (2018)
13. Calle-Jimenez, T., Sanchez-Gordon, S., Rybarczyk, Y., Jadán, J., Villarreal, S., Esparza, W., Acosta-Vargas, P., Guevara, C., Nunes, I.: Analysis and improvement of the web accessibility of a tele-rehabilitation platform for hip arthroplasty patients. In: Nunes, I. (ed.) Advances in Human Factors and Systems Interaction, AHFE 2018. AISC, vol. 781, pp. 233–245. Springer, Cham (2019)
14. Acosta, T., Acosta-Vargas, P., Salvador-Ullauri, L., Luján-Mora, S.: Method for accessibility assessment of online content editors. In: Proceedings of International Conference on Information Theoretic Security, pp. 538–551. Springer, Cham (2018)
15. Thapa, R.B., Ferati, M., Giannoumis, G.A.: Using non-speech sounds to increase web image accessibility for screen-reader users. In: Proceedings of the 35th ACM International Conference on the Design of Communication. ACM (2017)
16. Sánchez Morales, D.V., Tabares Morales, V., Londoño Rojas, L.F., Duque Méndez, N.D.: Comparison of authoring tools based on accessibility evaluations. J. Virtu@lmente **5**(1), 7–23 (2017)
17. Sanchez-Gordon, S., Luján-Mora, S.: How could MOOCs become accessible? The case of edX and the future of inclusive online learning. J. Univ. Comput. Sci. (J. UCS) **22**(1), 55–81 (2016)

Development of a Programming Code for Image Processing of Nodular Cast Iron

Victor Hidalgo(✉), Carlos Díaz, Anibal Silva, José Erazo, and Esteban Valencia

Departamento de Ingeniería Mecánica, Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito 170517, Ecuador
victor.hidalgo@epn.edu.ec

Abstract. This study focuses on the development of a code to perform an appropriate analysis of nodular cast iron metallography. The platform developed was written in Python programming language and used the Open Source Computer Vision library (OpenCV) for the image processing. The OpenCV tool was applied in order to convert the color photo of the metallography to a grayscale image and hence enable the segmentation of the gray phases to calculate the percentage of carbon within the cast iron test specimen. The categorized microstructural phases were perlite, ferrite and graphite. For validation of the platform and the methodology, the obtained results were contrasted with the Architecture Street Furniture (ASF) ductile iron chart, from there the percentage of differences between the model developed and the baseline specimens were among 2 to 18% for ferrite, 0.4 to 2.2% for pearlite and 2.1 to 12.1% for graphite. Furthermore, the obtained nodularity from the study cases were compared using examples from the ASTM A247 norm and the differences were between 1.4 to 8.1%.

Keywords: Nodular cast iron · Digital image processing · Python · OpenCV · Nodularity

1 Introduction

The nodular cast iron or ductile iron is an alloy based on iron and carbon, which presents in its structure graphite particles dispersed in the form of spherical nodules. These nodules impede alloy embrittlement and also provide properties of hardness, elongation, resistance to fatigue and wear resistance [1]. In consequence, ductile iron is used in the oil and hydroelectric industries for the construction of pipes and fittings of piping, gears and gearboxes, crankshafts, disc brake calipers, among others [2, 3].

Nevertheless, its ample spectrum of implementation contradicts the antique methods used to characterize microstructure, which mainly depend on experience and ability of a technical specialist [4, 5]. Regarding this last point, image processing applied to metallographic analysis has gained special interest in this field of study, since it allows to analyze multiple images simultaneously and is capable of extracting specific details of a metallography in very short timeframes [6, 7].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 327–334, 2020.
https://doi.org/10.1007/978-3-030-20040-4_30

Nowadays, there is a variety of software for processing images, that can be used in the metallurgical industry such as: Adobe Photoshop, PAX-it, Matlab, Stream Essentials, among others [8]. However, these programs are licensed and present a black box structure, which limits its use for research. In this context, the present study focuses on developing an open source code for processing images of the spheroidal graphite iron. The code is written in Python programming language and it will allow to identify aspects such as: nodularity, percentage of ferrite and pearlite, size of nodules, as well as graphite types.

2 Methodology

For a better understanding, the applied methodology was divided on: digital image processing, and image processing with Python – OpenCV.

2.1 Digital Image Processing

Digital image processing is a method to process digital pictures with computer algorithms [4], and the used scheme for the present research, it is indicated in Fig. 1.

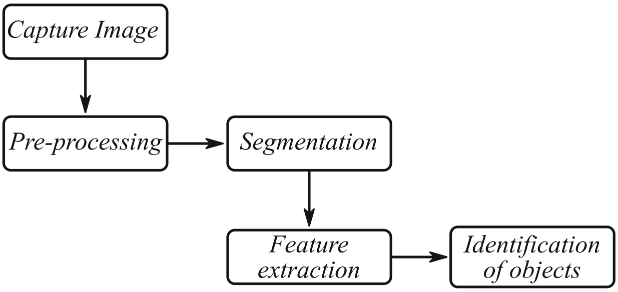


Fig. 1. Scheme of the digital image processing method.

Capture Image. It means to take picture of a metallography cast iron, which is going to be processed by a computer algorithm. Characteristics of the captured images depend on the type of application. For that, a metallographic microscope with a 100x magnification was used. The digitalization consists of decompose the image to an array of M by N, where each element has a proportional value to its gray level and takes a position of pixels as indicated in Fig. 2.

Pre-processing. In this step the pixels of the image are filtered for data optimization and the used digitalization are: Fast Fourier transform to remove noise, Gaussian dispersion, convolution for smoothing contours, and deconvolution to decrease fogging.

Segmentation. The image is dived into regions or segments that can be manipulated in grayscale. For example, regions that have similar gray levels are a segment, instead

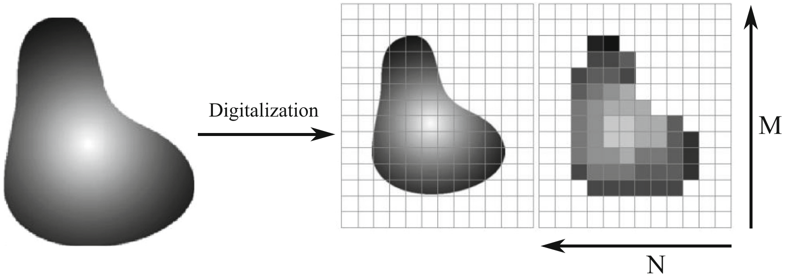


Fig. 2. Process of digitalization of an image.

there are abrupt changes in the gray level creates another segment, detecting the borders or edges of the image in the processing tasks with greater difficulty, because this depends on the interpretation of results.

Feature Extraction. It is a process of description or recognition of the elements found in the segmentation, to obtain information from these elements with features, such as: their geometrical characteristics, size, shape, etc., and for differentiate them from others that are also part of the set. At this stage the appropriate characteristics for the identification of desired objects are extracted.

Identification of Objects. It is an automated process of recognition or interpretation, where the characteristics of the image are compared with a database or decision-making algorithms, designed to run the final task of image processing, which in this case is the identification of nodularity of the sample and determine the percentages of ferrite and perlite.

2.2 Implementation Using Python and OpenCV Library

Grayscale. It is the process to change from color image to gray image based on that each pixel should take a value among 0 and 255 as indicated in Fig. 3. For that, the OpenCV library of python has been used with a buckle to compare the pixels with referential values (a, b), and to obtain a new image with the selected metallography phase as a matrix of pixels, which is used to compare with the original ones and to obtain the percentage of microstructural phase. The selected phases are: pearlite, ferrite and graphite, and the last one can be dispersed or concentrated as a graphite nodule.

Analysis of Percent Perlite, Ferrite and Graphite. The percentages of microstructure analysis are obtained using the digitalization and adding a counter vector. The pixel within the established range will be added to this vector. Then the elements of vector are counted. Finally, the percentage is determined by dividing the number of elements that are in the vector for the number of total pixels from the image. This process is shown in the Fig. 4. To perform a successful analysis, it is necessary to consider two values of thresholds in the digitalization of metallography. First, threshold

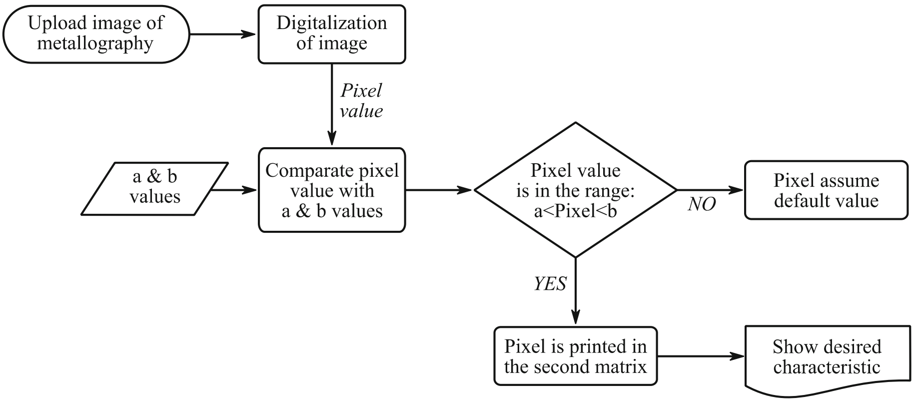


Fig. 3. Process of digitalization of an image.

values are needed to identify graphite and perlite in the metallography. Figure 5 shows an example of the proposed code, which is applied to a metallography.

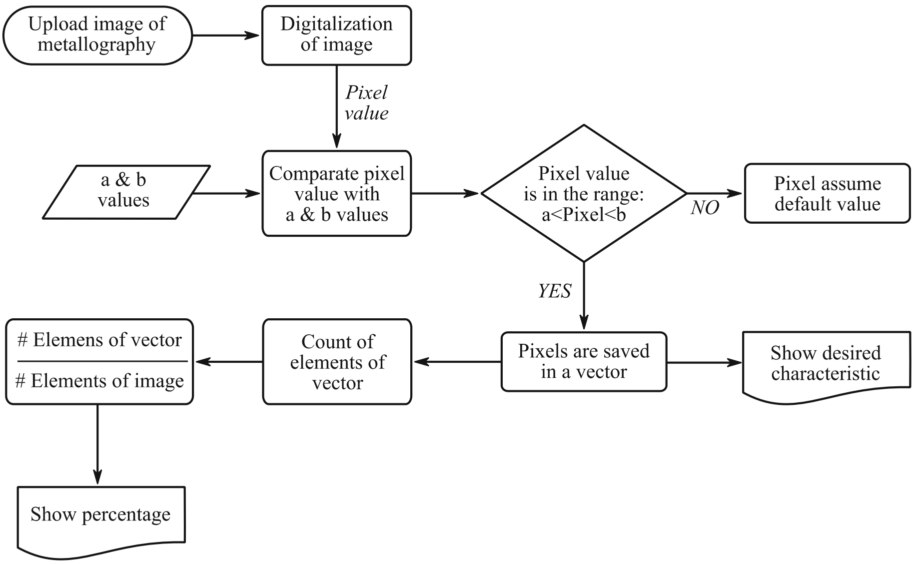
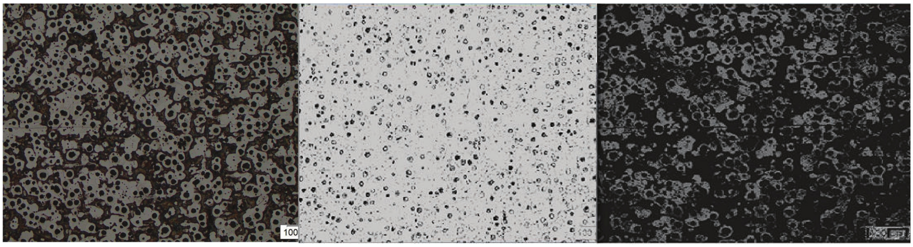


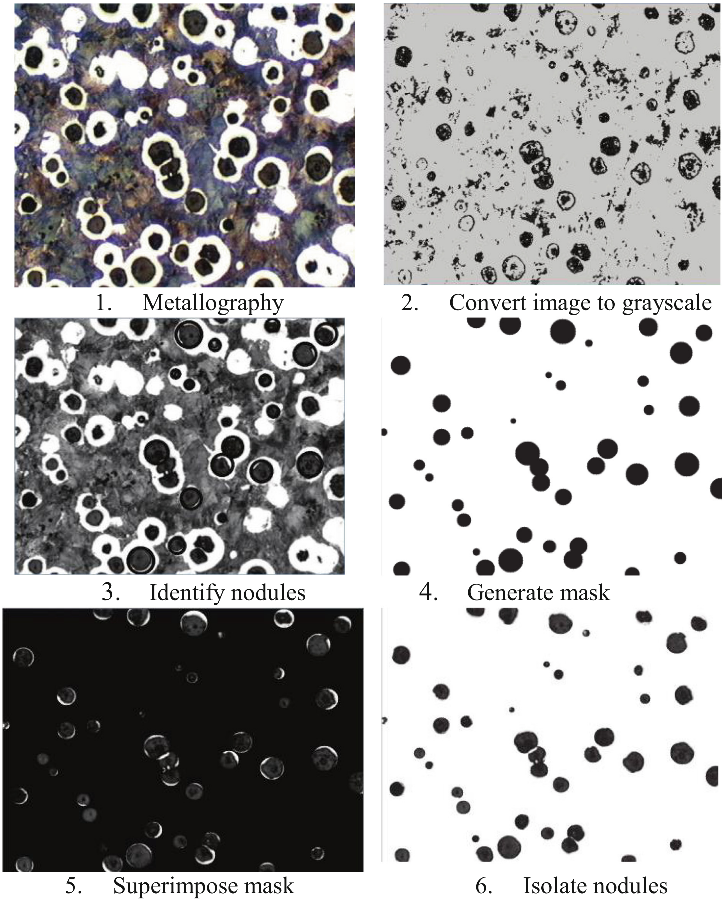
Fig. 4. Process to determinate percent of microstructural phases.

Nodularity. It is determined applying the steps of Fig. 6, and nodules are identified by selecting each one with circles to generate a mask, and after the mask is superimposed to the original image.



a) Metallography b) Graphite percentage c) Ferrite Percentage

Fig. 5. Example of the code proposed applied to a metallography.



1. Metallography
2. Convert image to grayscale
3. Identify nodules
4. Generate mask
5. Superimpose mask
6. Isolate nodules

Fig. 6. Steps to determine the nodularity of a metallography.

3 Experimental Production of Metallography Microstructures

Metallography photos were obtained from four experimental cases, which carried out using the metallography laboratory of Mechanical Engineering Faculty in Escuela Politécnica Nacional at Ecuador with norms of The American Society for Testing and Material ASTM E3 – 11 and ASTM E407 – 07, which are for preparation of metallography specimens of metals and alloys, metallographic examinations, chemical attack and a proper choice of specimen location. The photos were taken using a microscope with 100x of magnification and a camera with 12 megapixels without filters. With the specimen list, was to carry out a chemical attack to reveal the microstructure of the alloy with nital 2%, namely nitric acid diluted in water.

4 Results and Discussion

Results for the four study cases using the prosed python code, and the proposed methodology are indicated in Table 1.

Table 1. Results of the metallography structures obtained by the image processing.

Name	Graphite percentage (%)	Ferrite percentage (%)	Perlite percentage (%)
Case 1	12,63	45,75	45,25
Case 2	9,79	32,61	61,32
Case 3	11,21	31,43	63,27
Case 4	13,54	21,26	71,7

From Tables 2, 3 and 4, the percentage of graphite, ferrite and perlite are compared with the chapter of classification of microstructures of cast iron from the laboratory. In this context, the ferrite is the phase with the best detection, and the graphite is shown as the worst estimated phase by using the code with values close to 36%. The error can increase due to the dispersion of the graphite in the microstructure.

Table 2. Percentage of graphite in the analyzed metallography.

Name	Graphite Percentage (%)-Python code	Graphite Percentage (%)-reference	Difference	Error (%)
Case 1	12.63	10	2.63	26.3
Case 2	9.79	10	0.21	2.1
Case 3	11.21	10	1.21	12.1
Case 4	13.54	10	3.54	35.4

Table 3. Percentage of ferrite in the analyzed metallography.

Name	Ferrite percentage (%)-Python code	Ferrite percentage (%)-reference	Difference	Error (%)
Case 1	45.75	45	0.75	1.67
Case 2	32.61	30	2.61	8.7
Case 3	31.43	27	4.43	16.41
Case 4	21.26	18	3.26	18.11

Table 4. Percentage of perlite in the analyzed metallography.

Name	Perlite percentage (%)-Python code	Perlite percentage (%)-reference	Difference	Error (%)
Case 1	45.25	45	0.25	0.56
Case 2	61.27	60	1.32	2.2
Case 3	63.27	63	0.27	0.43
Case 4	71.7	72	0.30	0.417

The nodularity analysis is indicated in Table 5, and results shown that the error can reach the 20% such as the case 4. However, the other cases presented an error less than 10%. In this context, the obtained error from the metallography of case 4 can be related to mistakes in the experimental part too.

Table 5. Percentages of nodularity in the analyzed metallography.

Nodularity	Norm ASTM 247-16	Python code	Difference	Error (%)
Case 1	40	43,25	3.25	7.51
Case 2	90	92,63	2.63	2.92
Case 3	70	70,95	0.95	1.36
Case 4	10	12,35	2.35	19.02

5 Conclusions

A Python code was proposed to study microstructure phases of nodular cast iron, and its percentage of nodularity. Results shown that the developed code can be used to carry out the metallographic analysis of the iron. In addition, it can be used to obtain the percentages of microstructures in the boundaries. The mistakes to detect the graphite part and nodularity can be related to errors during to take the data from experiments.

Acknowledgments. The authors gratefully acknowledge the support provided by the Non-destructive testing laboratory of Escuela Politécnica Nacional for the development of this work, which is part of the research project PIJ 17-13.

References

1. Madtha, L.S., Babu, B.R.N.: Experimental behavioural study of ductile cast iron microstructure and its mechanical properties. Int. J. Eng. Res. **3**(3), 6 (2013)
2. Souza, T.N.F., Nogueira, R.A.P.S., Franco, F.J.S., Aguilar, M.T.P., Cetlin, P.R.: Mechanical and microstructural characterization of nodular cast iron (NCI) with niobium additions. Mater. Res. **17**(5), 1167–1172 (2014)
3. Nicoletto, G., Collini, L., Konečná, R., Riva, E.: Analysis of nodular cast iron microstructures for micromechanical model development. Strain **42**(2), 89–96 (2006)
4. Godbole, M.S., Jayashree, V.: Microstructure analysis of spheroidal graphite iron (SGI) using hybrid image processing approach. Int. J. Adv. Res. Comput. Eng. Technol. **3**(7), 6 (2014)
5. Fuentes, C., Alejandro, M.: Software para la caracterización de aceros al carbono mediante el procesamiento digital de imágenes metalográficas. Thesis, Universidad Central "Marta Abreu" de Las Villas (2012)
6. Sarojadevi, H., Shetty, A.B., Murthy, A.K., Shetty, P.B., Mukunda, D.P.G.: Digital image processing technique for microstructure analysis of spheroidal graphite iron. **1**(4)
7. Imasogie, B.I., Wendt, U.: Characterization of graphite particle shape in spheroidal graphite iron using a computer-based image analyzer. J. Miner. Mater. Character. Eng. **03**(01), 1–12 (2004)
8. McCaslin, S., Kesireddy, A.: Metallographic image processing tools using mathematica manipulate. In: Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering, pp. 357–363 (2015)

Security and Crises Management

Influence of Dynamic Automation Function Allocations on Operator Situation Awareness and Workload in Unmanned Aerial Vehicle Control

Yulin Deng[1], James Shirley[1], Wenjuan Zhang[1], Na Young Kim[1], and David Kaber[2](✉)

[1] Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695, USA
{ydeng2, jdshirle, wzhang28, nkim3}@ncsu.edu
[2] Department of Industrial and Systems Engineering, University of Florida, Gainesville, USA
dkaber@ise.ufl.edu

Abstract. The functional capabilities of unmanned aerial vehicles (UAVs) have dramatically expanded, placing substantial attentional and information processing demands on UAV operators. This study utilized a high-fidelity UAV flight simulation to explore the potential for DFAs in UAV control to reduce operator workload and support overall situation awareness. Three levels of UAV automation (LoAs) were compared, including DFA and static high and low level of automation. This research extended a preliminary investigation by Zhang et al. (2018). The present research addressed the limitations of the preliminary study by increasing the sample size and comparing effects of LoAs during 'easy to hard' and 'hard to easy' task difficulty transitions. Results of this study demonstrated the presence of "out-of-the-loop performance" issues under high LoA. Results also showed some support for use of DFAs to address out-of-the-loop problems in UAV operations. Findings of this study provide some guidance for design of DFAs in UAV control.

Keywords: UAV · Dynamic automation · Situation awareness · Workload

1 Introduction

Recent technological advances have dramatically expanded the functional capabilities of unmanned aerial vehicles (UAVs). Consequently, UAV control interfaces have become increasingly complex, in order to display system information and ensure operator awareness of system states. The complexity of system designs places substantial attentional and information processing demands on UAV operators. To reduce operator workload and assist task performance, automation has been widely applied to UAV functions, which allows pilots to focus on higher-level decision-making tasks and may serve to reduce the current manpower requirements for UAV operations [1].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 337–348, 2020.
https://doi.org/10.1007/978-3-030-20040-4_31

However, some researchers are concerned that higher level of automation (LoA) may result in severe reductions in situation awareness (SA) [2, 3], which can lead to delayed decision making and operation errors [4]. This negative side-effect of high LoA was observed by Endsley and Kris [2] in use of full automation in automobile navigation tasks, and they referred to it as the "out-of-loop" (OOTL) performance problem. The OOTL problem is exacerbated when a highly automated system fails to provide explanations for system performance and outcomes, which is common in automation applications across many domains [2, 5, 6]. In this case, it is less likely for operators to understand or question the basis for certain system actions. Therefore, they can become over reliant on the automation system and experience degraded situation awareness.

Compared to full automation, intermediate levels of automation have been found to be to reduce task workload without compromising operator situation awareness [2]. To address OOTL issues, some researchers have called for investigation of the utility of dynamic function allocations (DFAs) [7, 8]. Applications of DFA shift responsibility for tasks between a human operator and automation over time in order to achieve optimal system performance [9]. According to de Visser and Parasuraman [10], in various information processing tasks, only when the human and automation work together like this can optimal task performance be achieved. Earlier evidence suggests that DFAs can be beneficial in terms of reducing operator workload while maintaining situation awareness. In an air traffic control-like simulation, Kaber et al. [7] found that adaptive automation could effectively support low-level information processing functions, for example, information acquisition and basic actions. In a piloting and image analysis task simulation, Calhoun et al. [11] also observed improved mission performance and reduced pilot cognitive workload with adaptive automation. A recent investigation in the use of dynamic automated assistance in simulated UAV and unmanned ground vehicle (UGV) tasks also indicates that dynamic automation can increase operator SA and reduce workload [12]. Along the same lines of research, Afergan et al. [13] manipulated adaptive system behavior by increasing or decreasing simulated UAV control task difficulty. They utilized brain signals as an indicator of workload. Their results indicated adaptive system behavior leads to fewer vehicle collisions and improved error detection rates.

Collectively, the above research provides some evidence of the potential for critical effects of adaptive automation on UAV operational tasks. However, much of the current literature pays particular attention to applying DFAs to information analysis tasks rather than decision-making tasks. There remains a need to apply adaptive automation to potentially riskier system decision-making functions in UAV control.

This study explored the potential for dynamic function allocations (DFAs) in a UAV control simulation to reduce operator workload and support overall system/situation awareness (SA). An experiment was conducted with a high-fidelity UAV control simulation requiring participants to complete six simulated, military-style reconnaissance scenarios. This research extended a preliminary investigation of the benefits of DFAs in UAV control decision tasks by Zhang et al. [14]. The present research addressed the limitations of the preliminary study by increasing the sample size and comparing effects of LoAs during 'easy to hard' and 'hard to easy' task difficulty transitions.

2 Method

2.1 Participants

A total of 24 participants (23 male, 1 female; mean age = 20.3 yrs., S.D. = 2.5) were recruited from the local Reserve Officers' Training Corps (ROTC) program. The majority of participants (22) reported no prior experience with or knowledge of ground control stations (GCSs) software. Two participants reported limited experience with GCS. All participants had 20/20 vision with no color vision impairment. The study was approved by the university's institutional review board.

2.2 Apparatus

The study used a UAV simulation developed with the JustinMind prototyping tool (San Francisco, USA). The prototype was based on the ArduPilot Mission Planner software. The GCS features of this software are similar to those included in military UAV systems (e.g., primary flight display, digital flight instruments, multi-function control and display units, etc.). Slight modification was made to the prototype in order to deliver additional in-flight tasks.

The UAV control interface consisted of a (1) primary flight display (PFD), (2) mission map, (3) emergency controls, (4) flight parameter display, (5) waypoint display table, and (6) mission task bar (see Fig. 1). The required in-flight tasks were displayed to participants at various times during a mission.

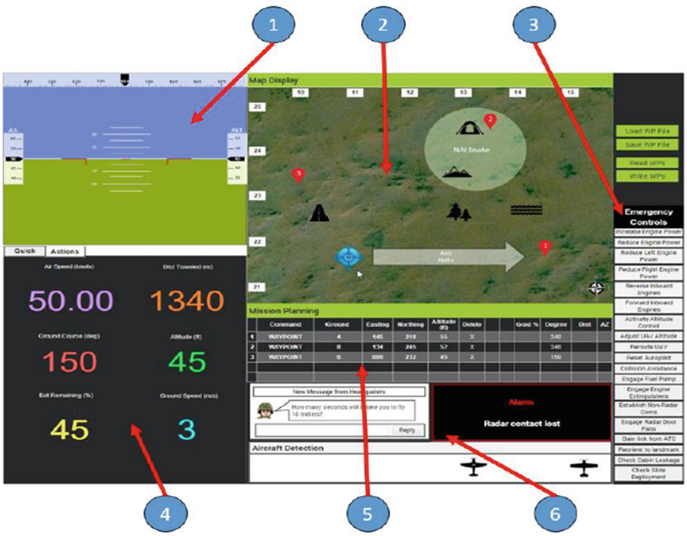


Fig. 1. Prototype interface - (1) primary flight display (PFD), (2) mission map, (3) emergency controls, (4) flight parameter display, (5) waypoint display table, (6) mission task bar.

All experiments were conducted in a controlled laboratory environment and an embedded office-like cubicle space. A monitor displayed the UAV control interface and participants interacted with the UAV simulation using a mouse and keyboard. Participants were also provided with pen and paper, which they were allowed to use only for flight arrival time calculations.

2.3 Scenario Structure

Each participant completed six simulated, military-style reconnaissance scenarios. Three types of decision-making tasks were embedded in each flight simulation scenario (see Sect. 2.4), including hostile aircraft identification, emergency alarm handling, and flight arrival time calculation. These tasks were presented with two levels of difficulty (easy vs. hard) across scenarios. This manipulation was intended to create critical workload conditions (e.g. overload under hard tasks) and SA-critical conditions (e.g., underload with easy tasks). Three levels of automation (LoAs) assistance were applied to the alarm handling tasks, and each participant was exposed to all three LoAs, including: (1) low static automation, (2) DFAs, and (3) high static automation.

Each mission scenario consisted of three blocks of tasks with different difficulty levels (see Fig. 2). The first and third block contained hard tasks and the second block contained easy tasks. Each block consisted of two "phases of operation". Each phase contained three different tasks presented in randomized order. Only one task was presented at a time and participants were given 20 s to complete each task. At the close of each phase, the simulation was suspended for 60 s, during which time the participant answered three SA queries displayed on the monitor. After each block, participants were asked to provide a subjective workload rating in addition to responding to SA questions. In this case, the simulation was suspended for an additional 20 s, resulting in a total of 80-second suspensions between blocks.

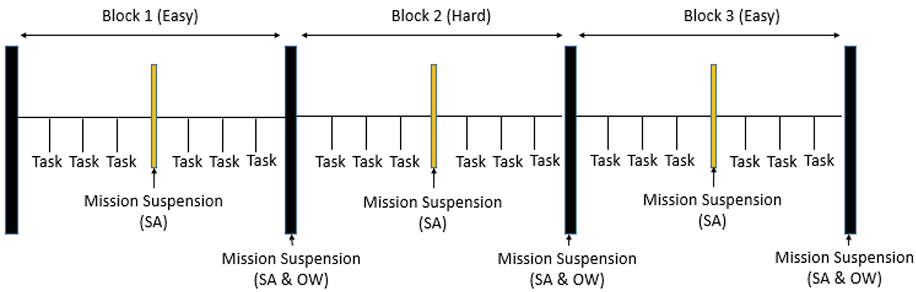


Fig. 2. A simulated UAV mission trial. ("Task" refers to one of the three types of decision making tasks posed to participants during trials.)

2.4 Tasks

Aircraft Detection Task. Prior to the experiment, participants were given sufficient time to memorize eight aircraft icons (four friendly icons, four hostile). During the experiment, participants were presented with random combinations of aircraft icons and they were asked to select any icons representing hostile aircraft and avoid selecting friendly aircraft. The level of difficulty of the task was manipulated by number of icons. Two icons were presented under the 'easy' level (one friendly and one hostile). Six icons (three friendly and three hostile) were presented under the 'hard' level of task difficulty.

Flight Arrival Time Calculation Task. The flight arrival time calculation task required participants to provide a time estimate (in seconds) for the UAV to fly a specified distance. This task measured reserve attentional capacity [15] and, therefore, was also considered an objective indicator of overall workload for participants. In order to complete the task, participants had to divide the specified distance by the vehicle's current ground speed, which was provided on the flight parameter display. For example, if a task message asked, "How many seconds will it take you to fly 12 m?" and the vehicle's current ground speed was 3 meters per second (m/s), then their answer would be 4 s. The 'easy' level task was a two-digit division problem with no remainder (e.g., 15 divided by 3). The 'hard' level task was three-digit division problem with a remainder (e.g., 101 divided by 4), and participants would need to round answers to the nearest single decimal place. Previous studies used a similar method to manipulate difficulty of division problems [16, 17].

Alarm Handling Task. The alarm management task required participants to react to off-nominal events by selecting an appropriate control action in the emergency control menu. During training, participants were presented with a table of all possible alarms and their solutions (i.e., associated emergency control actions). They were given adequate time to understand the logic of the emergency control solutions but advised that it was not necessary to memorize the solutions. When an alarm was triggered, participants had 20 s to select an appropriate control action. The emergency controls represented a set of responses for manned aircraft incidents based on the Aviation Safety Reporting Systems (ASRS) database.

During experiment trials, two LoAs were applied to the alarm handling task. The 'high' LoA condition represented a 'rigid system' alternative [see 18], where the computer recommended a limited set of actions for participants. When an alarm was presented under the 'high' LoA condition, the computer highlighted the three most relevant emergency control actions, and participants were instructed to choose from those suggestions. No decision-making assistance was provided with 'low' LoA.

SA Queries and Workload Ratings. Each SA query represented one of three different levels of SA (e.g., perception, comprehension, and projection), as defined by Endsley [19]. The SA questions tested participant knowledge of vehicle flight parameters, flight trajectory, and geographical features presented on the map display. All queries occurring during a scenario were unique, one from another.

During the simulation suspensions at the end of each block, participants were presented with another grey screen prompting them to provide a subjective rating of their overall cognitive workload (OW) on a scale from 0 to 100 [20]. They had 20 s to provide a response.

2.5 Experiment Design

The experiment followed a within-subjects design with two controlled manipulations: (1) in-flight task difficulty (easy vs. hard) and (2) mode of task automation (low LoA, high LoA, and DFA conditions). The subtask difficulty manipulation was included in order to increase or decrease pilot workload such that the task automation could be evaluated in terms of potential for mitigating workload under the harder difficulty condition. Each participant was exposed to all automation modes with one replication, producing 6 trials in total. The automation modes were presented in random order. With low LoA, no automation assistance was applied to the alarm task. With high LoA, automation assistance was applied to the alarm task in all scenario blocks. With the DFA mode, automation assistance for the alarm task was applied only under the hard level of difficulty.

Four response measures were collected to determine the effectiveness of the various modes of task automation. These measures included aircraft detection accuracy and flight arrival time calculation accuracy. SA query accuracy was calculated as the ratio of correct answers to the total number of queries posed during each trial block. The SA query accuracy and aircraft detection accuracy were considered objective assessment of operator SA. The flight time calculation accuracy and OW rating were used to measure operator cognitive workload.

2.6 Procedure

Upon arrival, participants were presented with informed consent form. After consenting, participants completed a demographic survey on their age, gender, visual acuity, and experience with UAV control interfaces.

Participants were then given a training session, in which they were familiarized with the UAV control interface and the three subtasks. They were allowed sufficient time to read a list of possible alarms and corresponding emergency control actions and memorize icons representing friendly and hostile aircraft. Subsequently, participants completed a training trial to become familiar with the UAV in-flight tasks. The training trial included all three subtasks, and both easy and hard levels of task difficulty were presented.

Following the training session, participants had the opportunity to review the friendly and hostile aircraft icons, to which they did not have access during test trials. Participants were instructed to complete 7 test trials. However, the first trial was in fact a "pseudo-trial", which did not involve recording of data for analysis purposes. The purpose of the "pseudo-trial" was to further reduce potential participant learning effects on response measures. The training and pseudo trials lasted about 10 min. Between trials, participants were given a 3-minute break.

2.7 Hypothesis

The DFA condition was expected to result in greater overall operator SA, indicated by higher SA query accuracy (H1) and higher aircraft detection task accuracy (H2) than the high LoA.

When transitioning from the hard to easy level of task difficulty, the DFA condition was expected to result in greater overall operator SA, indicated by a greater increase in SA query accuracy (H3) and aircraft detection accuracy (H4) than the high LoA.

The DFA condition was expected to result in lower overall operator cognitive work-load, indicated by lower subjective workload rating responses (H5) and higher flight arrival time calculation accuracy (H6) than the low LoA.

When transitioning from the easy to hard level of task difficulty, the DFA condition was expected to result in lower overall operator cognitive workload, indicated by a smaller increase in subjective workload responses (H7) and smaller decrease in flight arrival time calculation accuracy (H8) than the low LoA.

3 Results

3.1 Situation Awareness

ANOVA results identified no significant effect of the mode of automation on SA query accuracy (counter to H1). However, statistical analysis revealed a significant effect ($F(25, 406) = 12.62$, $p = 0.0447$, $1 - \beta = 0.6004$) of automation mode on aircraft detection task accuracy in-line with hypothesis (H2). Post-hoc analysis (Tukey's HSD test) indicated the high LoA condition led to less accurate hostile aircraft identification than the low LoA (see Fig. 3).

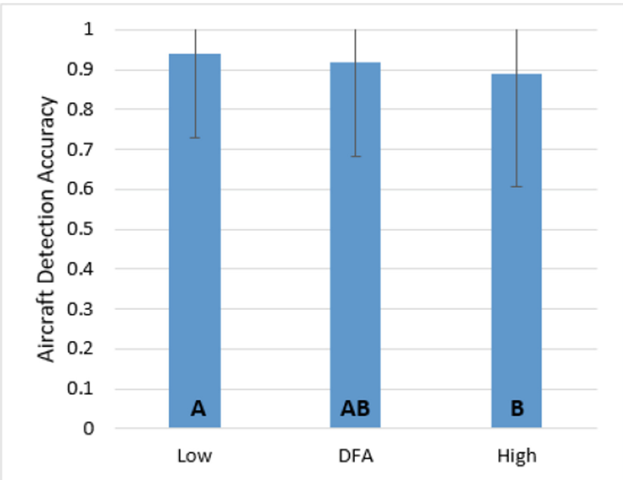


Fig. 3. Aircraft detection accuracy for different modes of task automation

Automation mode also had a significant effect ($F(24, 71) = 1.65$, $p < 0.0001$, $1 - \beta = 0.998$) on SA query accuracy change when transitioning from the hard to easy difficulty block (H3). According to Tukey's HSD tests, participants using the high automation condition improved in SA response accuracy, whereas the DFA condition resulted in a slight decrease in SA response accuracy (see Fig. 4). Automation mode did not have a significant effect on aircraft detection accuracy during transitions in task difficulty (counter to H4).

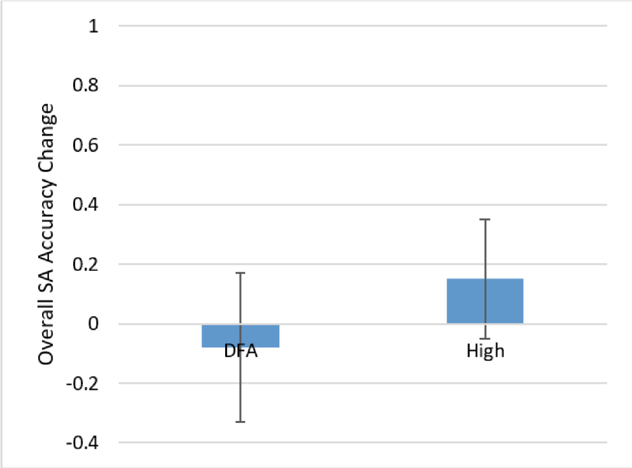


Fig. 4. Mean change in SA query accuracy when transitioning from hard to easy

3.2 Workload

No significant effect of automation mode was revealed on the subjective workload response ratings (counter to H5). However, statistic analysis revealed a significant effect of automation mode ($F\ (25, 400) = 2.59$, $p = 0.0163$, $1 - \beta = 0.7319$) on flight time projection task accuracy (H6). As revealed by Tukey's HSD tests, participants were less accurate under the high automation than the low automation or DFA conditions (see Fig. 5). No significant difference was identified between the low LoA and DFA conditions.

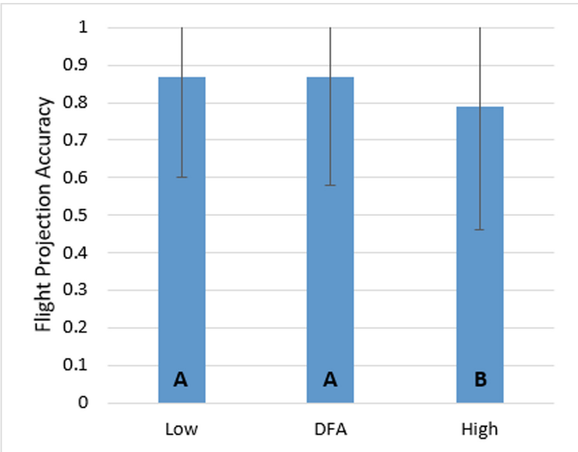


Fig. 5. Flight time projection accuracy for different modes of task automation

ANOVA results revealed no significant effect of the LoA on the degree of subjective workload changes during transitions in task difficulty (counter to H7) and no effect on the flight time calculation accuracy during transitions (counter to H8).

4 Discussion

4.1 Situation Awareness

H1 was not supported by the current findings. SA query accuracy was not significantly influenced by the decision-making assistance provided by the DFA, which was inconsistent with prior findings [9, 12, 17]. This result is possibly due to the DFA only being applied to the alarm handling task in the present study; whereas, participant SA may be influenced by other subtasks and monitoring of simulation environment.

The second hypothesis (H2) was also not supported, as results revealed no significant difference between the DFA and static automation modes. Findings differ from those of Parasuraman et al. [12], who demonstrated benefits of DFA in terms of situation awareness. Our findings confirm the presence of OOTL performance problems under the high static automation condition, which match results obtained by Endsley and Kaber [1995]. The trend of results suggests the potential for DFAs to reduce OOTL performance problems. However, the experiment and task design could be improved in order to make the benefit of DFAs more pronounced in terms of response measures.

Contrary to expectation (H3), SA query accuracy increased during task difficulty transitions under the high automation condition but decreased under the DFA condition. This finding also differs from previous research that observed improved SA with adaptive automation over high LoAs [9]. One likely explanation is that the high automation condition freed up additional attentional resources for participants to allocate to environment and mission status monitoring.

The fourth hypothesis (H4) was also unsupported. No significant difference in aircraft detection accuracy was found among automation modes. Again, this may be due to the automation manipulation being limited to the alarm detection task, which may not have been sufficient to cause significant differences in operator SA.

4.2 Workload

No difference in subjective workload ratings was found among automation modes. This finding also differed from our expectation (H5). Since automation assistance was only applied to one subtask (i.e., alarm handling), the difference in cognitive workload might not be subjectively noticeable. Moreover, as the workload rating followed the SA queries, it is possible that participants partially based their workload evaluation on their perceived success in responding to the queries, although they were instructed not to.

H6 was also not supported by results. No difference in flight arrival time calculation accuracy was identified between the DFA condition and low automation. The deteriorated calculation accuracy with high automation demonstrates an OOTL performance problem, which was not present under the DFA condition.

During transitions in difficulty, the DFA condition did not result in a smaller increase in subjective workload responses than the low automation condition, as was expected (H7). This result is logical given the outcome for H5, where there was no significant overall difference among conditions.

It was hypothesized that the DFA condition would result in a smaller decrease in flight arrival time calculation accuracy during transitions in difficulty, as compared with the low automation condition (H8). However, no significant difference was observed between the DFA and low automation conditions. Again, this may be due to the limited manipulation of automation under the DFA condition for participants to benefit in terms of workload.

5 Conclusion

This study used a high-fidelity UAV simulation to explore the potential for DFAs in a UAV control tasks to reduce operator workload and support overall system/situation awareness. The study focused on decision making tasks, which had not been previously explored by other research on adaptive automation by increasing the sample size and comparing effects of LoAs during ‘easy to hard’ and ‘hard to easy’ task difficulty transitions, the present study extended the preliminary investigation by Zhang et al. [14] and addressed the limitations of the preliminary study.

Results of the present study demonstrated the presence of OOTL performance issues under high-level static automation. While evidence of benefits of DFA was not as strong as expected, the study does partially substantiate use of DFAs to address OOTL performance problems in UAV operations, specifically in supporting aircraft detection tasks.

A major limitation of this research was limited utilization of the UAV control automation in decision-making tasks. Automated assistance was only applied to the alarm handling task, thus the effect of automation on operator SA and workload was not as prominent as expected. It is recommended for future investigations that automation assistance be applied to multiple aspects of UAV control, in order to make potential benefits of DFAs more pronounced in terms of operator SA and cognitive workload. Furthermore, subjective workload reporting took place after task completion, which might have affected the accuracy of workload responses. Real-time physiological measures of workload, like electroencephalography (EEG) or heart rate variability monitoring, may provide more sensitive and direct indicators of cognitive demands during UAV control. Therefore, future studies may consider using such real-time measures. Higher fidelity and military-style simulations may also be helpful in future investigations.

References

1. Monfort, S.S., Sibley, C.M., Coyne, J.T.: Using machine learning and real-time workload assessment in a high-fidelity UAV simulation environment. In: Next-Generation Analyst IV, vol. 9851, p. 98510B. International Society for Optics and Photonics (2016)
2. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. Hum. Factors **37**(2), 381–394 (1995)
3. Wiener, E.L., Curry, R.E.: Flight-deck automation: promises and problems. Ergonomics **23** (10), 995–1011 (1980)
4. Endsley, M.R., Onal, E., Kaber, D.B.: The impact of intermediate levels of automation on situation awareness and performance in dynamic control systems. In: Proceedings of the 1997 IEEE Sixth Conference on Human Factors and Power Plants, 1997. Global Perspectives of Human Factors in Power Generation, p. 7. IEEE (1997)
5. Porat, T., Oron-Gilad, T., Rottem-Hovev, M., Silbiger, J.: Supervising and controlling unmanned systems: a multi-phase study with subject matter experts. Front. Psychol. **7**, 568 (2016)
6. Chen, J.Y., Barnes, M.J., Harper-Sciarini, M.: Supervisory control of multiple robots: human-performance issues and user-interface design. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **41**(4), 435–454 (2011)
7. Kaber, D.B., Perry, C.M., Segall, N., McClernon, C.K., Prinzel III, L.J.: SA implications of adaptive automation for information processing in an air traffic control-related task. Int. J. Ind. Ergon. **36**(5), 447–462 (2006)
8. Hou, M., Zhu, H., Zhou, M., Arrabito, G.R.: Optimizing operator–agent interaction in intelligent adaptive interface design: a conceptual framework. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **41**(2), 161–178 (2011)
9. Kaber, D.B., Endsley, M.R.: The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theor. Issues Ergon. Sci. **5**(2), 113–153 (2004)
10. de Visser, E., Parasuraman, R.: Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. J. Cogn. Eng. Decis. Making **5** (2), 209–231 (2011)
11. Calhoun, G.L., Ward, V.B., Ruff, H.A.: Performance-based adaptive automation for supervisory control. In: Proceedings of Human Factors Ergonomics Society Annual Meeting, vol. 55, no. 1, pp. 2059–2063. SAGE Publications, Los Angeles (2011)
12. Parasuraman, R., Cosenzo, K.A., De Visser, E.: Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. In: Military Psychology, vol. 21, no. 2, p. 270 (2009)
13. Afergan, D., Peck, E.M., Solovey, E.T., Jenkins, A., Hincks, S.W., Brown, E.T., Jacob, R.J.: Dynamic difficulty using brain metrics of workload. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 3797–3806. ACM (2014)
14. Zhang, W., Shirley, J., Deng, Y., Kim, N.Y., Kaber, D.: Effects of dynamic automation on situation awareness and workload in UAV control decision tasks. In: International Conference on Applied Human Factors and Ergonomics, pp. 193–203. Springer, Cham, July 2018
15. Osburn, W.J.: Levels of difficulty in long division. Elementary Sch. J. **46**(8), 441–447 (1946)
16. Wickens, C.D., Gordon, S.E., Liu, Y., Lee, J.: An introduction to human factors engineering (1998)

17. Murata, A., Iwase, H.: Evaluation of mental workload by fluctuation analysis of pupil area. In: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 6, pp. 3094–3097. IEEE (1998)
18. Endsley, M.R., Kaber, D.B.: Level of automation effects on performance, SA and workload in a dynamic control task. Ergonomics **42**(3), 462–492 (1999)
19. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Hum. Factors **37** (1), 65–84 (1995)
20. Vidulich, M.A., Tsang, P.S.: Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. In: Proceedings of Human Factors Ergonomics Society Annual Meeting, vol. 31, no. 9, pp. 1057–1061. SAGE Publications, Los Angeles (1987)

Pilots' Role in the Critical Infrastructure of Aviation

Aysen K. Taylor(✉) and T. Steven Cotter

Engineering Management and Systems Engineering, Old Dominion University, Norfolk, VA 23529, USA
aysenafk@cox.net

Abstract. The US government considers aviation to be a critical infrastructure as reflected in government publications developed after the 9/11 terrorist attacks. The Patriot Act of 2001 defined the transportation systems sector as one of the assets essential to the physical and economic security of the nation. Aviation represents a large part of the US economy and pilots are integral to its operations. Some airline operators have declared bankruptcy citing the inability to hire enough qualified pilots. Concerns about pilot shortages have grown because the expected growth in air traffic in the coming decades will require many more pilots. Pilot unions are concerned that low pilot pay and the cost of training to earn a commercial license is making it unaffordable for prospective pilots. Cyber security vulnerabilities are likely to increase as the NexGen air traffic control system comes online and more automation is used in the aviation system. GPS jammers are not hard to obtain and might impact safety while a broad scale attack on the GPS satellite infrastructure would devastate an automated transport system. Cockpit crew composition has shrunk as technology was developed to reduce the number of crew required. This increased the criticality of the crew that remains when automated navigation equipment fails, reaches its design limits, or acts in a manner the crew is unprepared for a dilemma is looming where the aviation industry to both working to reduce the necessity of human pilots by increasing use of automation while at the same time developing new programs and incentives for people to make piloting a career. Will perspective pilots see this trend "too" soon and cause a drastic drop in available pilots before the industry has no need for them? With an unknowable timeline, the process leading to a pilotless airliner will continue until fully autonomous commercial passenger transports are a reality. Aviation, and humanity, will be forever changed.

Keywords: Critical infrastructure · Aviation · Pilots

1 Introduction

The Patriot Act of 2001 defined critical infrastructure as assets that are essential to the physical and economic security of the country. The George W. Bush administration developed the National Strategy for the Physical Protection of Critical Infrastructures and Key Assets (NSPP) in the context of its homeland security plans. Loss of these systems could have a severe impact on the security, safety, or economic viability of the

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 349–360, 2020.
https://doi.org/10.1007/978-3-030-20040-4_32

United States [1]. The Department of Homeland Security (DHS) lists 16 areas that comprise America's critical infrastructure as shown in Table 1 [2].

Table 1. Sixteen sectors that comprise critical infrastructure in the United States

Chemical Sector	Commercial Facilities Sector	Communications Sector	Critical Manufacturing Sector
Dams Sector	Defense Industrial Base Sector	Emergency Services Sector	Energy Sector
Financial Services Sector	Food and Agriculture Sector	Government Facilities Sector	Healthcare and Public Health Sector
Information Technology Sector	Nuclear Reactors, Materials, and Waste Sector	Transportation Systems Sector	Water and Wastewater Systems Sector

The transportation system sector is one of the largest and the DHS and Department of Transportation (DOT) provide oversight in this domain. The goal of these Federal agencies is to ensure the transportation system moves people and cargo within the United States and overseas in a safe and efficient manner. Aviation is a significant part of the transportation system and constitutes five percent of America's gross national product. Professional airline pilots employed by U.S. air carriers transport an average of nine hundred million passengers each year [3]. Pilots represent some of the soft system elements required for the aviation system to function [4]. Highly complex and costly technologies comprise many of the hard elements. Economic considerations include Open Sky route access agreements between countries, trade agreements [5], tax policies, as well as the controversial topics of state subsidies provided to foreign carriers and calls to privatize the Air Traffic Control (ATC) system. Over the last few years, the role of pilots in support of the aviation component of the Transportation systems sector has received more attention. Advances in cockpit automation have made some question the role human pilots will have in the future [6, 7]. Media reports of looming pilot shortages are met with calls for better pay from pilot unions to attract more people to the profession [3]. Unmanned Aerial Systems (UAS) will be integrated into the National Aviation System (NAS) but these also require trained pilots to operate them from ground stations as few fully autonomous aircraft have been flown outside of a research setting [8]. Professional piloting, as a career, is under stress from many changes in regional and global economics, the regulatory environment, and advances in automation technology. Most forecasters predict a rapid expansion of aviation worldwide in the next 20 years with the greatest growth coming in Asia. Less clear is the question of pilot supply versus demand and when partially or fully autonomous commercial passenger traffic will become a reality.

Cyber security concerns in aviation will increase as the new air traffic control system known as NextGen is deployed. Among the many changes coming with NextGen will be automated plane-to-plane communication to allow for reduced separation minimums during most phases of flight, increasing capacity and efficiency. NextGen will utilize vulnerable GPS satellite signals to improve navigation accuracy. China and other countries have tested anti-satellite weapons that could cripple any guidance system dependent on GPS. Pilots have a vital role in providing resiliency to the NextGen system in the event of a system fault, system destruction, or hacking attacks from individuals, terrorist organizations, or state actors. In August of 2016, Delta airlines lost electrical power to an operations center in Atlanta, resulting in the cancellation of 2000 flights and a financial loss of 150 million dollars for the company [9]. No lives or aircraft were lost in this incident but it serves as a window into what could happen should a large-scale infrastructure attack be successful in disrupting commercial flight operations in the United States.

General aviation is defined as all aspects of the industry outside of military and air carrier operations. Over 140 million passengers are transported each year by 630,000 certified general aviation pilots. General aviation provides most pilots with their basic initial training and first flight experience through work in charter operations, emergency medical services, law enforcement, firefighting, as well as corporate business services.

Accident rates in commercial aviation are low but without further improvements in this rate, increased traffic over the next twenty years may result in many more deaths and hull losses in absolute terms. Automation has reduced human error in some aspects of flying but also created new areas of concern relating to interactions between humans and automation [10]. The National Transportation Safety Board (NTSB) often cites pilots as the primary cause of commercial aircraft crashes according to accident investigations conducted over the last 40 years. How human pilots enhance safety in highly automated airliners may be underestimated. The human potential to solve novel problems by analyzing core facts is exemplified in the incident of an Airbus A380 operated by Qantas Airways that suffered severe mechanical damage to a wing after an explosion in one of its four engines. The flight control system was among the components damaged and the Captain sent one of the crew (pilot) to the plane's passenger cabin to visually assess the damaged engine and wing. The crew conducted an assessment of their aircraft systems status and performed landing calculations and, after a discussion amongst the crew, determined a landing was feasible at their airport of origin. The A380 landed with no loss of life or injuries. [11]. This outcome was due to the expert decision making of the human crew to manage a complex and novel challenge that current automation technologies or remote piloting schemes would have been less likely to handle as well.

Can future fully autonomous passenger aircraft manage similar scenarios successfully? These questions will need to be examined as the process of moving people and freight through the air is further automated on the path to full automation. Military research efforts are leading the development of optionally piloted aircraft (OPA) and could provide a basis to develop autonomous civilian aircraft [8].

Pilots are often viewed as the cause of accidents but they could be viewed as the last factor in a chain of circumstances. Hart [12] visualizes this as a box with rotating discs, each having one hole. As a light is projected from the left, light will only exit the right side of the box if all the holes line up on the discs as seen in Fig. 1.

Hart suggests accident investigations erroneously focus primarily on the last disc (the pilots) and cautions against only fixing the last disc without examining the interactions of the preceding discs. These precursor discs represent many components of the system the pilots have no control over [12].

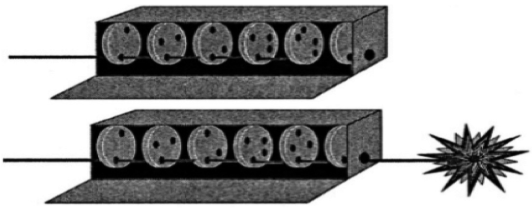


Fig. 1. Hart [12]

The 9/11 terror attacks had great economic impact on the domestic airline industry and caused many pilots to be furloughed because of a dramatic drop in air travel. This cost the industry 3 years of growth [13]. The 2008 financial crisis similarly hurt the transport economy and cost 2 years of growth [14].

Women have seen growth in workplace opportunities in most industries over the last 30 years but they are significantly under represented in the airline industry [15]. Measures to bring more women into the pilot profession could help meet expected increases in pilot demand in the medium and long term.

Improvements in automation have influenced crew composition by steadily reducing the number of humans needed in the cockpit and the natural extension of this will lead to fully autonomous commercial transports. The timetable for this is difficult to know and a transition phase where partial responsibility for flight control is held by ground based crewmembers is expected.

2 Perspectives on a Pilot Shortage

In 2016, Republic Airways, a regional carrier supporting Delta Airlines, United Airlines, and American Airlines cited a pilot shortage as the cause of its bankruptcy filing. They claimed that this shortage forced them to ground so many planes they were in danger of becoming economically unviable. They continued partial operations while under bankruptcy protection and hoped to reorganize their company. The airline employs just over 2000 pilots. Most flights carry 76 or fewer passengers. Republic Airways, like many regional carriers, is the sole operator providing air service to many small cities that cannot support large, legacy carrier operations. Republic increased pay for its pilots and offered signing bonuses but still had difficulty meeting its recruitment goals. Training requirement changes mandated by the FAA that reflected regional carrier pilot competency concerns following the crash of Colgan Air flight 3407 in 2009 increased minimum required flight experience hours for First Officers from 250 to 1500. This slowed the flow of new pilots into regional carrier programs. The large carriers Republic Airways contracts with started suing Republic because they cancelled flights for lack of pilots [16].

The increase in minimum flight hours required to become a First Officer increased the costs associated with becoming an airline pilot substantially. The FAA provided exceptions to the 1,500-hour rule for former military pilots who have 750 h and trainees who complete 4-year aviation schools with 1,000 h. The 1,500-hour rule resulting from the Colgan Air crash also increased simulation training requirements regarding how to respond to a stall event and increased rest periods between shifts. Regional carriers complain this rule is an arbitrary mandate for a given quantity of flight training and regulations should instead emphasis training in particular areas rather than total flight time. An industry sponsored study called Pilot Source Study 2015 [17] found that pilots with less than 1,500 total flight hours had *fewer* non-completions and *less* extra training events than those with more than 1,500 h. This study noted other variables that impacted the performance of new First officers to regional carriers and this research has been used to undermine the necessity of the FAA's enhanced training requirements put in place since the Colgan air crash.

Potential airline pilots accumulate flight time through several channels. Some fly while in military service but increased retention efforts by the armed forces have reduced the number of pilots who choose to transition from the military to commercial aviation. Many new pilots work full or part time jobs as a Certified Flight Instructor (CFI) to amass more flight time. College level aviation programs convey bachelor's degrees and commonly employ senior students as flight instructors for newer students in the same program. Many other niche fields such as banner towing, pipeline patrol, and aerial photography are low paying but can be useful in attaining flight hours. Commercial pilots require an Air Transport Pilot (ATP) license and must have 1,500 total hours to include 500 h of cross-country flight time, 100 h of night flying, and 75 h of Instrument Flight Rules (IFR) time. The IFR flight hours can be in a simulator. Other ATP requirements include being 23 years old, having a separate instrument rating, and a comprehensive understanding and use of English in written and spoken form. ATP certification also requires passing a rigorous written examination. A lessor rating known as Airline Transport pilot – restricted (ATP-r) is available for those who do not meet the more stringent requirements of an ATP. This rating only requires 1,500 total hours with 200 being cross-country. An ATP-r holder can only serve as a First officer in a two pilot crew. They can thus acquire the other qualifications to get an unrestricted ATP license and do not need to undergo further examination [18]. The total number of pilots (including all ratings) has declined over the past several decades with 827,000 registered pilots in 1980 decreasing to 590,039 in 2015. While the total number of registered pilots went down, commercial ATP license holders increased from 107,732 in 1990 to 152,933 in 2015. All pilots except those with Sport or Recreational ratings are required to undergo periodic medical exams; those holding ATP licenses have the most stringent exam requirements.

The International Civil Aviation Organization (ICAO) revised the maximum age for certain pilots in international operations from 60 to 65 in 2006. In 2007, the Fair Treatment for Experienced Pilots Act became law and increased the mandatory retirement age for airline pilots under FAA regulation from 60 to 65 [19]. This bought some time before a wave of forced retirements results in a significant decline in ATP license holders against a backdrop of increasing pilot demand.

A review of the data for ATP license holders reveals a decrease of pilots in age groups 40–44 and 45–49 between 2011 and 2016. And those numbers almost stayed same for the age group 50–54. Although there is an increase in demand for air transportation, this information shows a decline of mid-career pilots to support the industry [20]. Underpaying pilots makes this career less attractive for the upcoming generations. Newer pilots have less manual flying experience in their training and learn to rely on automated control systems earlier in their career than previous generations of pilots. The FAA only "recommends" for air carriers to increase their pilots' manual flying time but provides no mandates for how many hours and how often they need to fly manually (Figs. 2 and 3).

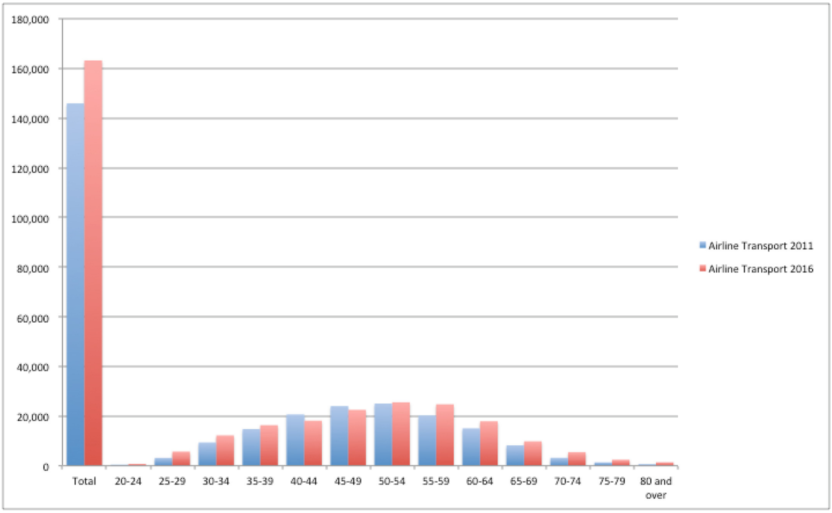


Fig. 2. Comparison on ATP License holders between 2011–2016

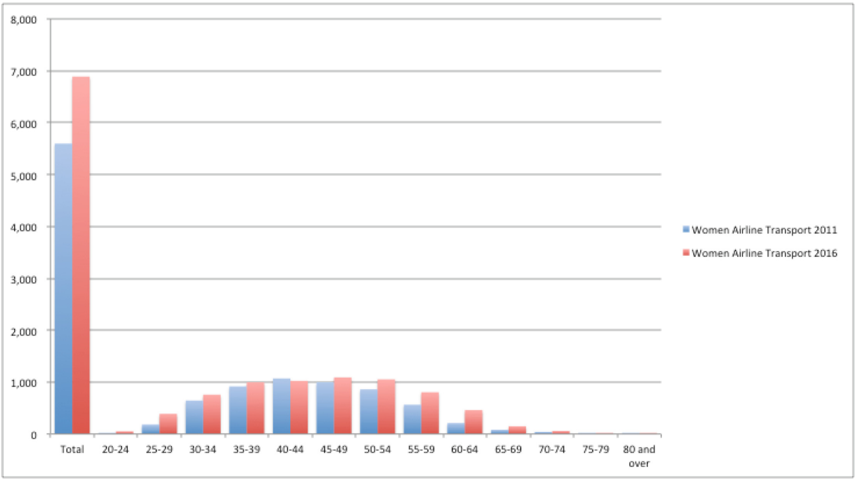


Fig. 3. Comparison on Female ATP License holders between 2011–2016

Pilot unions and professional associations have long maintained that there is a shortage of pay for pilots, not a shortage of pilots. The Airline Pilot Association (ALPA) believes that low pay has prevented sufficient numbers of new pilots from entering the profession.

ALPA has an extensive outreach program that sends volunteers into schools and colleges to advocate for choosing a career in commercial aviation. While supportive of the 1,500 h rule for First officers, ALPA laments the high cost of flight training and recognizes this as a barrier for some to choose aviation as a profession. ALPA has lauded regional airlines that have increased pay for new pilots and provided avenues to migrate to larger carriers through their code share partners; ALPA has also criticized some regional carriers that have first year salaries below $30,000. They also assert that service to small communities is not being impacted by a pilot shortage but by macro economic realities and that some of these small markets have seen an increase in service. Some airlines have started transition programs for military rotary wing (helicopter) pilots to cross train as fixed wing (airline) pilots and this approach could potentially supply a meaningful number of pilots for their fleets [21].

3 Elimination of the Flight Engineer

In the 1980's, advances in computerization of the cockpit made it possible to eliminate the role of flight engineer in the traditional 3-person cockpit crew. Advances in technology allowed the captain and first officer to adjust and monitor the systems and also perform their original duties. This advance was directly linked to new autopilot and auto throttle control systems. This was the last major change regarding standard crew configuration of commercial aircraft and may provide insight into how future disruptions in the human component of flight control might look like. John Tracy (Boeing) said "some freight customers are asking for autonomous planes today" [6]. Issues include certification procedures, regulatory requirements, and public perception. There are no formal programs from Boeing or Airbus on building pilotless transports and the FAA states that no one has approached them about certifying such designs. NASA has a program for UAV transport systems but FAA does not allow them to conduct tests without a pilot inside the vehicle yet.

4 Path to Fully Autonomous Transports

Figure 4 shows a possible scenario of development leading to autonomous passenger transports.

Technological advances have prompted the reduction of human crewmembers over time, leading to the current two pilot crew configuration. It is notable that accident rates have steadily declined against the backdrop of declining numbers of humans in the cockpit. NextGen will provide a framework in which further crew reductions are tenable. Military research is advancing rapidly in support of a one crew cockpit configuration using either a robot First Officer or a ground based human backup. The next logical step is a semi-autonomous (pilotless) plane with ground crew support. Pilotless

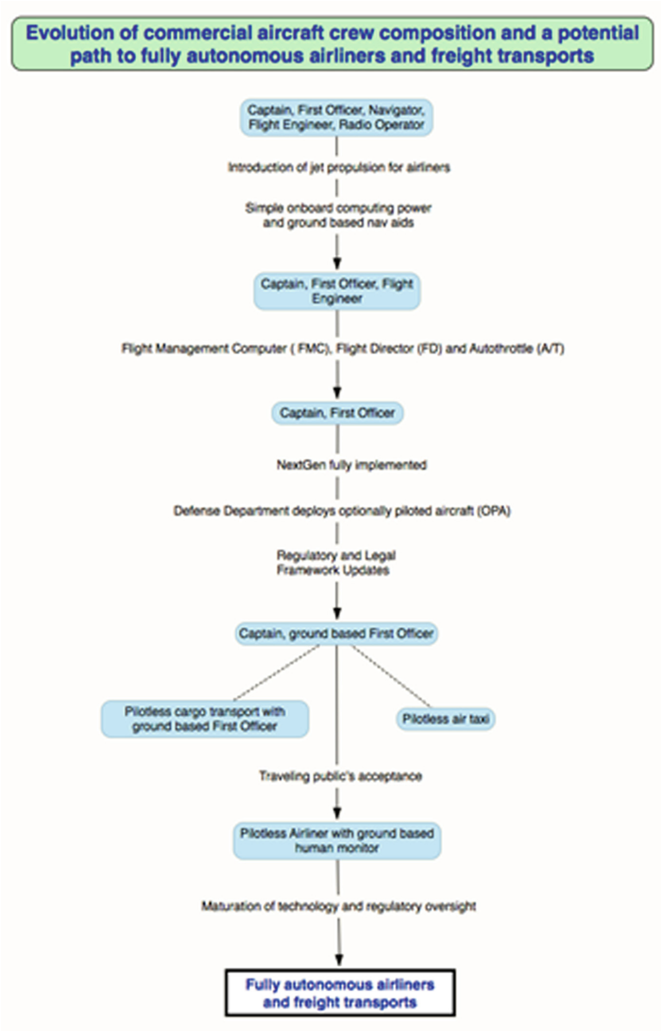


Fig. 4. Path to fully autonomous systems

cargo transports may be the first to use that system while small air taxi's introduce the public to the notion of flying without a human pilot onboard. Regulatory and legal standards for semi and fully autonomous transports will need to be defined before any reduction to the current crew configuration is possible. The flying public will need to be convinced that safety will not be compromised when no human pilots are onboard their aircraft. At some point in this process, the challenge of recruiting people to be professional pilots while at the same time reducing the need for onboard pilots could cause a crisis in the aviation industry.

5 Pilots' Role in Safe(R) Air Transportation

Although the importance of human pilots is being deemphasized in commercial aviation, several incidents illustrate how pilots are critical assets of these systems. Flight management computers rely on an array of sensors to obtain data from the environment and adjust control surfaces and throttle settings to maintain a safe and efficient flight path and reduce fatigue for the human pilots. These systems allow for high altitude and long haul flights that would be difficult and less safe with only humans at the controls. What happens when sensors fail? Either automated systems will have no or reduced data and transfer full control to the human pilot instantly (Air France 447), or if there is erroneous data, this data will create inappropriate output (Qantas 72, Turkish Airlines 1954) and human pilots will have to take manual control; oftentimes not understanding why their aircraft is behaving the way it is and in a time critical situation [22–24]. The effectiveness of pilot responses in these scenarios can vary and influence the outcomes of such inflight emergencies. If the pilot was not trained effectively and over relied on the automated systems, the result is a catastrophic loss such as when Air France 447 crashed over the Atlantic Ocean 5 min after the autopilot disconnected. The A330's air speed sensors failed causing a reversion to manual control that the crew was not able to manage. Other examples demonstrate the essential value of human pilots when automation fails as seen in the Qantas 72 incident that occurred in 2008 [23]. In this case, a failed air data inertia reference unit (ADIRU) caused the autopilot to command sudden $-8°$ pitch down, causing significant injuries amongst the passengers and crew. A highly experience Captain took manual control of the aircraft and safely landed at a diversion airport. The final report noted that the extensive testing and verification work performed on the automated system failed to detect this possible failure scenario. Airbus altered their design and published directives for crew procedures in case of similar emergencies and this reflects the reactive nature of the industry to events in which human crews save their plane when automation fails. What might have happened in an aircraft utilizing a ground-based crew or in a fully autonomous airliner? Turkish Airways 1951 is an example of a competent crew losing control of their aircraft in part because of opaqueness in the logic of automated flight control systems [24]. A faulty radio altimeter caused the autopilot to reduce thrust as if the aircraft was a few feet above the runway when it was actually several hundred feet above ground. The crew's attempt to go around came too late and failed.

Captain Sullenberger successfully ditched an A320 aircraft that suffered bird strikes in both engines soon after take off. Again, this demonstrates the resiliency human pilots can provide to highly automated aircraft when rare or unique challenges manifest for a particular flight. The human piloting crew will continue to lose emphasis in future designs with the expectation that improvements in automation technology can allow it to recover from the uncommon emergencies discussed in this paper. A transitional path to pilotless airliners is revealed in an Airbus patent, US 20140180508 A1, for a crew placement design that removes the pilots from the traditional cockpit and places them in the plane's cargo section, relying on cameras to see outside the aircraft.

On 29 October, 2018, a 737 MAX 8 operated by Lion Air as flight 610 crashed into the Java sea 13 min after takeoff, killing all aboard. The MAX 8 version of the 737 is the first to feature what Boeing calls the Maneuvering Characteristics Augmentation System (MCAS), which automatically pushed the aircraft's nose down if it detects a stall. One of the angle of attack sensors was faulty on this flight. While the preliminary report listed no probable cause due to it being an incomplete investigation, they stated the pilots counteracted the MCAS system's nose down command 26 times [25]. Boeing has been known for their pilot centric design philosophy but the MCAS system indicates a shift in philosophy that increases automation's authority at the expense of pilot authority. The MCAS system can be disabled by the crew but pilot union representatives expressed dismay after the crash that many pilots were never warned of the potential need to disable this system in the event of an angle of attack sensor failure [26].

6 Conclusion

Commercial passenger and cargo transports are a vital part of the global transportation infrastructure. A sufficient number of qualified pilots are required for the safe and efficient operation of aviation and many more pilots will be required over the next twenty years to accommodate the expected growth in traffic. Carriers are concerned about a lack of qualified pilots to replace retiring pilots and to expand their flight operations to take advantage of efficiencies the NextGen ATC system will provide.

Pilot unions cite low pay and the high cost of training as causal factors in any shortfall of new pilots entering the system. Automation has helped reduce the crew size for modern airliners to a captain and first officer and this trend will continue to the logical conclusion of fully autonomous transports. The timetable for this transition is difficult to estimate but a long interim period has been initiated with new technologies such as NextGen. This system will provide many benefits but the GPS signals it uses can be interfered with and plane to plane communications are also vulnerable to hacking and loss through unforeseen circumstances.

Pilots are often blamed after crashes but their role in mitigating accidents may be underestimated. While automation has made flying safer, persistent concerns about the interface between the automation and the human monitor have arisen. Pilots will be in the cockpit for a long time to come and improvements in computer based flight control systems will be needed to reduce the accident rate in the context of a greatly expanding aviation market. While the current loss rates are admirably low, the great increase in expected traffic would result in many deaths in absolute terms if safety rates are not improved further. Economic fluctuations, either as part of a normal economic cycle or in response to dramatic events like 9/11 or the financial crisis of 2008 can significantly impact the profession of piloting and cause economic hardship across the industry. Women are under represented in aviation and this should be explored further. Military helicopter pilots could possibly provide more pilots for fixed wing commercial transports and some carriers are promoting transition programs to recruit this subset of experienced pilots.

Fully autonomous transports will require a transition period with one human crew in the cockpit and one on the ground in support. Regulatory changes and an evolution of public perception regarding the suitability of using airliners without any piloting crew on board will be needed before we see this foundational shift in how we travel around the world.

References

1. Moteff, J., Parfomak, P.: Critical infrastructure and key assets: definition and identification (Order Code RL32631). Government Printing Office, Washington, DC (2004)
2. Critical Infrastructure Sectors (2017). https://www.dhs.gov/critical-infrastructure-sectors
3. Airline Pilots Association. We keep America flying (2017). https://www.alpa.org/~/media/D376EB61112D48E38350A58FE899C3BA.ashx
4. Taylor, A., Keating, C., Cotter, T.: Cockpit in systems engineering lenses. In: Coperich, K., Cudney, E., Nembhard, H. (eds.) Conference Proceedings of the 2017 Industrial and Systems Engineering Conference, May 2017
5. US DOT reveals EU's inability to prevent flags of convenience in aviation (2017). https://www.eurocockpit.be/news/us-dot-reveals-eus-inability-prevent-flags-convenience-aviation
6. Reiner, A.: Towards the end of pilots (2016). https://www.theatlantic.com/technology/archive/2016/03/has-the-self-flying-plane-arrived/472005/
7. Moskvitch, K.: Would you fly in a pilotless airliner? (2016). http://www.bbc.com/future/story/20160912-would-you-fly-in-a-pilotless-airliner
8. RPA vector: vision and enabling concepts 2013–2038 (2014). https://www.globalsecurity.org/military/library/policy/usaf/usaf-rpa-vector_vision-enabling-concepts_2013-2038.pdf
9. Jansen, B.: Delta outage a reminder of fragility of airline computers (2017). https://www.usatoday.com/story/news/2017/01/30/delta-outage-airline-technology-problems/97250834/
10. Taylor, A., Cotter, T.: Human opinion counts-making decisions in critical situations when working with highly automated systems. In: Long, S., Ng, E.H., Downing, C., Nepal, B. (eds.) Proceedings of the 2016 International Annual Conference American Society for Engineering Management, October 2016
11. ATSB transport safety report AO-2010-*089* (2013). https://www.atsb.gov.au/media/4173625/ao-2010-089_final.pdf
12. Hart, C.A.: The global aviation information network (gain). In: Johnson, C.W., Palanque, P. (eds.) Human Error, Safety and Systems Development, pp. 17–30. Kluwer Academic Publishers, New York (2004)
13. The impact of September 11, 2001 on aviation. (n.d.). http://www.iata.org/pressroom/Documents/impact-9-11-aviation.pdf
14. Goyal, R., Negi, D.: Impact of global economic crisis on airline industry. Int. J. Commer. Bus. Manag. **3**, 297–301 (2014). http://www.iracst.org/ijcbm/papers/vol3no22014/7vol3no2.pdf
15. Maksel, R.: Why are there so few female pilots? (2015). https://www.airspacemag.com/daily-planet/why-are-there-so-few-female-pilots-180954115/
16. Briggs, J.: Why Republic airways filed for bankruptcy even though it's profitable (2016). https://www.usatoday.com/story/travel/flights/todayinthesky/2016/02/27/why-republic-airways-filed-bankruptcy-even-though-s-profitable/81035522/
17. Pilot source study (2015). https://www.pilotsourcestudy.org/executive-summary
18. Pilot certification and qualification requirements for air carrier operations (2013). https://www.faa.gov/regulations_policies/rulemaking/recently_published/media/2120-AJ67.pdf
19. Fair treatment of experienced pilots act (2012). https://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/info/all_infos/media/age65_qa.pdf

20. U.S. civil airmen statistics (2017). https://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics
21. Rotor transition program (2018). https://www.envoyair.com/pilots/rotor-transition-program/
22. BEA final report AF 447 (2012). http://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf
23. ATSB transport safety report AO-2008-070 (2011). https://www.atsb.gov.au/media/3532398/ao2008070.pdf
24. Turkish Airways 1951 Final Report (2010). http://www.onderzoeksraad.nl/uploads/items-docs/1748/Rapport_TA_ENG_web.pdf. Accessed 18 Mar 2014
25. Lion Air Flight 610 Preliminary Report (2018). https://reports.aviation-safety.net/2018/20181029-0_B38M_PK-LQP_PRELIMINARY.pdf
26. Gregg, A., Halsey, A.: Pilots unions accuse Boeing of withholding safety information. The Washington Post, 13 November 2018. https://www.washingtonpost.com/business/2018/11/13/pilots-unions-criticize-boeing-withholding-safety-information/?noredirect=on&utm_term=.c422a8051e30

Improving Usability with Think Aloud and Focus Group Methods. A Case Study: An Intelligent Police Patrolling System (I-Pat)

Tania Calle-Jimenez[1(✉)], Doris Tutillo-Sanchez[1], Sandra Sanchez-Gordon[1], Janio Jadán-Guerrero[2], Cesar Guevara[2], Patricio Lara-Alvarez[2], Patricia Acosta-Vargas[3], Luis Salvador-Ullauri[4], and Isabel L. Nunes[5]

[1] Escuela Politécnica Nacional, Quito, Ecuador
{tania.calle,doris.tutillo,sandra.sanchez}@epn.edu.ec
[2] Universidad Tecnológica Indoamérica, Quito, Ecuador
{janiojadan,cesarguevara,patolara}@uti.edu.ec
[3] Universidad de Las Américas, Quito, Ecuador
patricia.acosta@udla.edu.ec
[4] Centro de Educación Continua, Quito, Ecuador
lsalvador@cec-epn.edu.ec
[5] Universidad de Nova de Lisboa, Lisbon, Portugal
imn@fct.unl.pt

Abstract. This study proposes the use of Think Aloud and Focus Group methods for evaluating the usability of the Intelligent Police Patrolling System (I-Pat). The study was conducted with twenty-one students of computer engineering. The study included two evaluation methods. First, the application of Think Aloud using audio recordings, image capturing and questionnaires. Second, the application of a Focus Group for brainstorming. The total number of the usability problems identified was fifteen. Comprehensiveness (46%) and layout (43%) problems were the most frequently found. The study showed that the problems encountered were due to the lack of understanding of the system's functions, so it is recommended increasing the users' levels of knowledge about the system. The application of these methods caused the students to find a greater number of errors than when applying a single method, allowing them to generate a report with usability improvements according to the reported errors.

Keywords: Think Aloud · Focus Group · Usability · Patrolling · I-Path

1 Introduction

The International Organization for Standardization in ISO/IEC 9126-1 defines usability as the ability of the software product to be understood, learned, used and be attractive to the user, when used under specific conditions [1]. However, this concept does not receive the appropriate importance, which implies confusion, error, delay or absolute failure to complete some task by the user [2].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 361–373, 2020.
https://doi.org/10.1007/978-3-030-20040-4_33

Nowadays, this is changing, promoting an evaluation of the initial phases of the software development process and a more active participation by the user in the later phases of the cycle, including usability that is considered a crucial non-functional requirement for economic success in competitive markets [3, 4].

This research evaluates the usability of the Intelligent System of Police Patrols (I-Pat) developed by the Ministry of the Interior in Ecuador. This system covers one of the fundamental needs for the Ecuadorian community, that is security with technology, patrols and surveillance of cities to identify crimes and criminal groups. I-Pat is a model that generates dynamic routes applying artificial intelligence through the K-means algorithm to identify critical points when patrolling a circuit within the Ecuadorian territory. In addition, I-Pat has incorporated the Google Maps API tool to design routes and verify means of transport. The usability evaluation of I-Pat allowed the measurement of the capacity of the system to be understood, learned and used by police officers. The Think Aloud (TA) and Focus Group (FG) evaluation methods were used to propose an evaluation that emphasizes aspects such as the global structure of the site, focusing on content and navigability, designed to offer clear paths from the origin to the destination, without complex terminology and with the mission of guiding the user. Finally, the consistency of the contents, as well as the visual format, was evaluated. The disposition of elements in the pages offer homogeneous environments that help to promote an effective communication of the message. At the same time, the consistency helps the user to form a mental model of the site [5].

TA allows collection of direct verbal information from the user about their thoughts. This includes cognitive processes and strategies while participating in a task. This method ranges from open and unstructured questions to structured interviews that can be adjusted according to the responses and behavior of the user [6]. In addition, TA is a technique used to understand cognitive processes based on tasks requested from users, both within usability tests and in the broader study of human-computer interaction. In this study, the evaluation was applied to twenty-one users; users interacted with the system's interfaces and described the actions and tasks they performed. Users reported inconsistencies, confusion, concepts not understood, unfinished actions, as well as those aspects that seemed good about the interaction.

FG is a focused discussion in which a moderator addresses a group of participants through a set of questions about a particular topic. This method consists of a planned and designed discussion about the interfaces. In this research, the discussion was guided and facilitated by a moderator. The moderator followed a predefined structure so that the discussion remained focused. Two groups were formed and participated in separate discussion sessions. The activity that was used was brainstorming, in order for the participants to exchange opinions. This research presented the results of the usability evaluation in a technical report that contains recommendations to improve the usability level of the I-Pat system.

2 Literature Review

In the literature review, authors have identified the use of TA and FG as methods to measure usability of various software products, [7] defines TA as a usability evaluation method used to gain insight into how people work with a product or interface, [8] argues that usability professionals currently use variations of TA as the main method to identify usability problems. This study compares three TA Protocols for using in testing. The first protocol, called Instruction, is where there are no instructions for varying amounts of instruction or practice sessions, then Intervention: different types of intervention with a variety of probes or prompts, and Prompting: varying rates of the test with the administrators administering a prompt. This study shows variations of TA with these protocols that usability practitioners currently use, the practitioners who want measures that reflect unaided user behavior can choose between two methods traditional or speech communication that do not differ significantly from each other in terms of their effects on user performance. In the [9] study the concurrent think-aloud method was used. This is a highly valued technique that is used to understand peoples' task-based cognitive processes, within both usability testing and the broader study of human-computer interaction. In addition, [10] showed that a relaxed think-aloud, which included evaluator interventions, increased test time and led to changes in behavior at the interface, [11] has other approaches to gathering verbal data; these include constructive interaction, where users work together in pairs. Constructive interaction can take a number of forms: users can work together on an equal footing [12] or alternative formats may be used such as teach-back scenarios, in which the user teaches another person about the focus system. In [13] comments users said that they may not give reasons for their actions during think-aloud, only a description of the actions themselves, [14] proposes some tools to capture the observations during the sessions using TA. Other studies have focused more on tactical questions of usability evaluation, for example [15] presents an example of a results report that should be presented after the evaluation with TA.

The literature suggests that TA is a widely used method for evaluating usability by taking its initial version proposed by Ericson and Simon or with different versions obtaining good results, as shown by [16] where it was discovered that think-alouds were classified among the five main methods used by the usability professionals in Sweden over prototyping, interviews, field studies and scenarios.

On the other hand, FG have been used for decades as mentioned in [17]. Discussion is frequently used as a qualitative approach to gain an in-depth understanding of social issues. The method aims to obtain data from a purposely selected group of individuals rather than from a statistically representative sample of the broader population. FG are less threatening to many re- search participants, and this setting is useful for participants to discuss perceptions, ideas, opinions and thoughts. For example, [18] proposes a series of steps to maximize the reliability of the data and recommends some criteria like reliability, internal validity, transferability and confirmability to ensure that important aspects of qualitative research are reported, which increases the quality of the published studies of FG.

[19] specifies a framework to capture and analyze the information derived from FG in relation to the type of information generated, quasi-statistical methods and conversation analysis. In addition, [20] proposes that FG can provide a valuable and complementary empirical experience quickly and at low cost, [21] uses the method to identify usability characteristics with respect to an institutional website.

However, as mentioned by [22], although focus groups can be a powerful tool in software development, it should not be used as the only source of usability data, [23] reviews some of the characteristics of the techniques for usability and considers TA as a complement in the application of FG. Consequently, the present research will use TA and FG as usability evaluation methods for the I-Pat system.

3 Methods

When considering safety as a global society need, this study applies the TA and FG evaluation methods to the police patrol system developed by the National Police of Ecuador. One of the objectives of the application of these methods is to obtain usability errors according to the characteristics of their participants, such as their behavior, gender, age and knowledge of mobile systems. The I-Pat system, the participants in the evaluation and the evaluation process are described in detail below.

3.1 I-Pat System

I-Pat is the acronym for Intelligent Police Patrolling. This system covers one of the fundamental needs for the Ecuadorian community, i.e. safety. With the use of technology, patrols and surveillance of cities will be facilitated by identifying crimes and criminal groups, helping the work of the national police. In this context I-Pat was developed for applications that use the Android system in order to perform intelligent patrolling for the National Police of Ecuador. For this purpose, it uses information about the crimes reported in three circuits during a period of four years [24], supported by a model that generates dynamic routes applying artificial intelligence with the K-means algorithm to identify critical points when patrolling a circuit within the Ecuadorian territory. In addition, the Google Maps API tool is used to design the routes and means of transport that the police officer will use. The I-Pat system has historical modules detailing the current information available regarding theft from people, theft from houses, theft from banks, theft of vehicles, theft of motorcycles, theft of accessories and theft on roads. It also has a complementary section with functions which include requests for police reinforcements patrols, predictive maps, schedules of the agents and civil unrest. Figure 1 shows a sample of interfaces of the patrol, civil unrest and vehicle theft modules.

3.2 Participants' Test

In this research, the usability of the I-Pat System was evaluated with the participation of students from the Escuela Polit´ecnica Nacional of Ecuador. Participants in this study consisted of two groups, including testers and facilitators. Testers included twenty-one

Fig. 1. I-Pat system interfaces

computer engineering students in their sixth and seventh semester who are trained to do the quality assessment of different systems, since students have approved the software quality course. Testers had knowledge about designing and analyzing computer systems as well as the principles and concepts of programming languages but no knowledge and experience using I-Pat. Before starting the evaluation, all the users were verbally familiar with the study's aims, and their informed consent was obtained. For the first TA assessment method, the facilitator reminded users to express their thoughts aloud. While, for the second FG method, the facilitator took the role of moderator, directing the groups to obtain the greatest number of ideas regarding the usability of I-Pat.

3.3 Evaluation Process

Figure 2 shows the process followed to apply TA and FG to evaluate the usability of the I-Pat system. To start with, the evaluation users were asked to install the application on their mobile phones with the Android operating system. Some testers used an emulator installed on a desktop machine, because they did not have a mobile with this operative system. Once testers had the application started, an introduction to the system's functions was done to familiarize the users.

The first session used TA, the method that allows the collection of information about the users' cognitive interaction with the system [25]. The users were asked to express what they saw, thought and felt with respect to the use of the I-Pat interfaces. After this introduction, the TA ran for 30 min and, during this session the facilitator reminded the users to express their thoughts aloud and audio recorded them for later analysis.

The second session worked with the FG method. Two groups were organized, the moderator brainstormed with them with the aim of obtaining original ideas about the evaluation of the usability of the system, through the free exchange of ideas between each of the testers. The moderator explained the rules of the session, such as avoiding criticism and expressing all the ideas that the users have. The moderator also mentioned that all thoughts were important even if those thoughts did not appear to be common sense because those ideas could serve as inspiration for the development of new proposals. Testers expressed their ideas to the rest of the group.

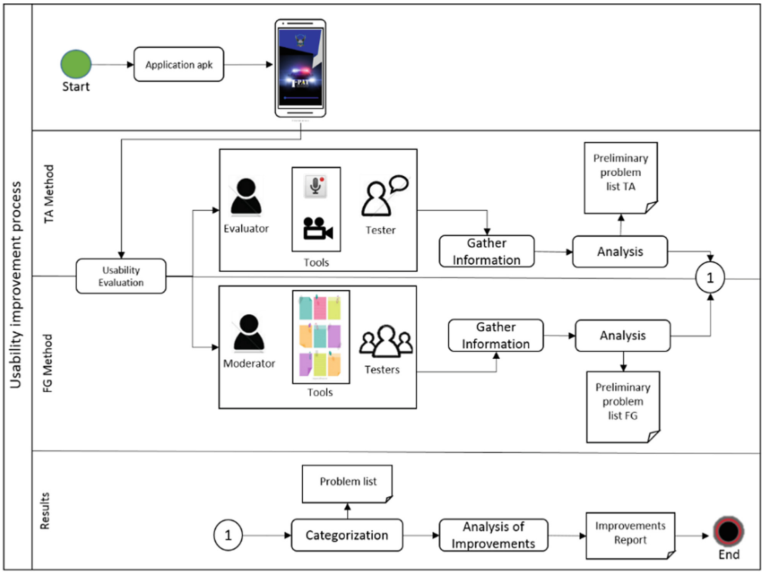


Fig. 2. Usability evaluation

3.4 Questionnaires

In addition, users filled out a questionnaire consisting of twenty questions to measure usability. This questionnaire was based on QUIS [26], SUMI [27] and SUS [28].

The questionnaire applied in this research consisted of four sections. The first contained questions about demographic details, such as age, gender and education. It also focused on the participants' experience in working with mobile devices. The second section was focused on learning about the system, and the third section worked on topics related to system capacity. The final section worked with usability and user interfaces. The items are presented on a scale from zero to nine, zero being confusing and nine being clear.

4 Results

Finally, the audio, images and questionnaire applied for TA and the brainstorming obtained with the FG were available for analysis. This study was conducted to analyze the usability problems of the I-Pat System using TA and FG methods. Participants included thirteen (62%) males and eight (38%) females with age ranges of between 18–22 (66.7%), 23–27 (28.6%) and 28–32 (4.8%). Eighteen errors were found with the TA method; however, errors were reported several times by different users.

Table 1. TA and FG results

Item	Problems	TA	FG
1	Sections without help	x	
2	Non-intuitive terms	x	x
3	Buttons without information/labels	x	x
4	Zoom in maps not working	x	x
5	Button close session is wrongly located	x	x
6	Buttons without functions	x	x
7	Patrol did not work	x	
8	Incorrect menu distribution	x	
9	Maps without help	x	
10	Personalization was not allowed		x
11	Messages were not displayed		x
12	Without user manual		x
13	Buttons without help	x	x
14	Unsuitable colors in interfaces		x
15	The menu did not work well		x

Table 1 shows the errors identified in the TA and FG sessions, the greater part of the errors found in TA were related to incorrect or unintuitive terms, the lack of textual aids or a user manual. In addition, functionality errors were reported. On the other hand, thirty-four errors were reported with the FG method, with this method the errors of functionality and the lack of help found with the TA are maintained and interface errors are added. To summarize, fifty-two errors reported by twenty-one users were found but, after the analysis and categorization, these were summarized as fifteen errors. The errors found were classified according to the Haak method [12], this method establishes four main categories described in Table 2.

Table 2. Haak method categories

Item	Category	Description
1	Layout problems	The participant fails to spot a particular element on a screen
2	Terminology problems	The participant does not comprehend part(s) of the terminology
3	Data entry problem	The participant does not know how to interact with the search
4	Comprehensiveness problems	The system lacks information necessary to use it effectively

In addition, the Nielsen method [29] was used to identify the severity of the problems found according to the classification in Table 3.

Table 4 shows the classification and severity of the errors which users encountered in the assessment process. According to the table, comprehensiveness problems were the most frequent with seven (46%) of cases. Around 43% of the problems in this category had a severity of 2 and 3. In this category, the main problems reported were the malfunction of the zoom in the maps (3), functions such as the patrol not working (12), non-descriptive icons (18), functions whose purpose was unknown (6). Then six (40%) layout problems were identified by the users. In this category, four problems were rated as Severity 2, the problems related to lack of help in the icons (18), the absence of messages (6), inappropriate colors (5). These were the ones which recurred most often.

The main results, after the application of the questionnaire, were regarding the users' experience with the operation of the system. According to this, fifteen (71%) users had no problem using the application and four (19%) cases had one to three interruptions without stopping the evaluation process.

Table 3. Severity classification

Item	Problems
0	No usability problem
1	Cosmetic problem
2	Minor usability problem
3	Major usability problem
4	Usability catastrophe

Table 4. Errors found

Item	Problems	Category	Severity	Method
1	Zoom in maps not working	Data entry problem	3	TA/FG
2	Incorrect menu distribution	Layout problems	1	TA
3	Icons without description	Comprehensiveness problems	2	FG
4	Buttons without functions	Comprehensiveness problems	3	TA/FG
5	Titles on the buttons made no sense	Comprehensiveness problems	3	TA
6	Map without help	Comprehensiveness problems	3	TA
7	Misplaced buttons	Layout problems	1	TA/FG
8	Unsuitable colors	Layout problems	1	FG
9	The menu did not work well	Data entry problem	3	FG
10	Personalization was not allowed	Layout problems	1	FG
11	Messages were not displayed	Data entry problem	3	FG
12	Non-intuitive functions	Comprehensiveness problems	3	FG
13	There was no help for the user	Comprehensiveness problems	2	TA/FG
14	Patrol did not work	Comprehensiveness problems	3	TA
15	Non-intuitive terms	Terminology problems	2	TA/FG

Furthermore, two (9%) problems remained as unresolved after the user's review, and the evaluation process continued without that task. After the evaluations, 90% of the users made proposals for improvements to the system. Most of the suggestions were related to better design (57%) and improvements with the functions that Google maps use (21%).

The main results after the application of the questionnaire, it was regarding the users experience with the operation of the system. According this 15 (71%) users had no problem using the application and 4 (19%) cases had one to three interruptions without stopping the evaluation process. Furthermore, 2 (9%) problems remained as unresolved after the users review, and the evaluation process continued without that task. After the evaluations, 90% of the users made proposals for improvements to the system. The most suggestions were related to the better designing (57%) y improvements with the functions that Google maps use (21%).

4.1 Technical Report

As result of this research, a technical report was prepared with the proposed improvements to guarantee a higher level of usability of the I-Pat system. Table 5 shows the errors found with the respective technical improvement. The improvements were prioritized taking into account the criteria of functionality as priority 1, with the aim that the system is fully functional avoiding the problem of the user being unable to perform their tasks. Priority 2 took into account errors related to clarity and ease of use of the system, i.e. the implementation of user aids, intuitive terms, descriptive icons, help and validation messages, thus avoiding confusion for the user. Finally, taken into account were errors related to interfaces, which are errors that do not stop the operation of the systems, however they must be considered so that the user is comfortable with the use of the system.

Table 5. Technical Improvements

Problems	Improvement
Zoom in maps not working	Check the Google API integration in the application, also enable the zoom functions within the maps
Patrol does not work	Review the implementation of the k-means algorithm to correctly perform patrolling
Buttons without functionalities	Assign the methods correctly to each button within the application
The menu does not work well	Verify that the classes and methods assigned to the menu perform the tasks according to the requirements In addition, it must verify that the corresponding task is enabled
Messages are not displayed	Verify that the confirmation and validation messages are implemented to facilitate the tasks performed by the user within the application
Non-intuitive functions	Implement messages in each of the functions detailing their main objective

(*continued*)

Table 5. (*continued*)

Problems	Improvement
Map without help	The maps must include aids that allow entering the current location, destination and the time it will take for a certain route. In addition it must allow the visualization of alternative routes
Titles on the buttons without sense	The words on the buttons should contain descriptive names, allowing the user to understand what the purpose of that function is. Or, in turn, a link to the user manual must be included within the application
Non-intuitive terms	Should use terminology that the user can understand. It is recommended to change the labels to more descriptive terms
There is no help for the user	It is recommended to have a user manual, place tool tip text on the buttons, displaying the contact information to support users
Icons without description	This problem is related to points 8, 9 and 10. It is also recommended that the buttons, menus and images contain a label with a descriptive name
Misplaced buttons	The buttons must be located correctly, without blocking access to other functions within the application
Unsuitable colors	It is recommended to follow some basic principles for handling colors such as theory, models, color schemes that allow, through a study, to determine the colors that help the user feel comfortable with the application
Incorrect menu distribution	It is recommended to prioritize the order of the functions in the menu, you must place the functions with which you work continuously until you reach the lowest level with little-used functions
Personalization is not allowed	It is recommended to implement a personalization of aspects such as: colors, tones, font types and language with which the user feels more comfortable and can more easily access the application

5 Discussion

As can be seen from the evaluation results, fifty-two errors reported by twenty-one users were found. After the analysis and categorization, these were summarized to fifteen errors in Table 4. The errors that had more incidence were those of comprehension since there were functions of the system that were not correctly implemented, hindering the navigability and completion of the tasks. Secondly, there were the problems related to the interfaces specifically as regards colors, help in the icons and correct distribution of the elements. As can be seen in Table 1, greater errors were obtained using FG. This was because users felt more comfortable expressing their ideas in a group because they could verify that the other users expressed the same or similar problems. It should be noted that TA is a method that seeks users to express their thoughts aloud. However, it is an activity that users are not used to and it caused some discomfort, in spite of the evaluator, even though he was limited to reminding the users to keep talking.

In addition, there was a questionnaire that had to be completed outside the test site, the questions were aimed at fine-tuning the results obtained in the user test in aspects related to the utility, ease, speed, effectiveness and satisfaction of the system. What users valued most positively was the ease of use of the system, the clarity of the vocabulary and the availability of the application. On the other hand, what the users found most negative was the lack of clarity of the icons, the texts of the menus and the structure of the screens and, finally, the absence of a help page. Regarding the severity of the problems encountered, both with the interface and with comprehension, the highest index is categorized as severity 3, since no problem was identified that would prevent the use of the application from continuing. This, to a certain extent, was influenced by the experience of the users in computer applications.

The literature review suggests the use of TA and its variations for the evaluation of usability [16, 20]. On the other hand, [24]'s research proposes the use of FG as a qualitative research technique widely disseminated in various fields of research in psychology and other scientific disciplines. The present research combines the advantages of both methods to find a greater number of usability problems. As you can see in the results, 5, there were problems that were reported with both methods and others that corresponded only to TA and FG respectively. Due to the nature of each method, the results will be influenced by how the user feels when expressing their thoughts and findings. Some prefer to internalize and analyze their experience before using the system and then express it and discuss it with more users, while others have the ability to instantly communicate what they think while using and testing the system. The purpose of the usability testing studies is to identify system usability problems and provide solutions to solve these problems. In this regard, the users made suggestions for solving these problems and improving system performance. Most of these suggestions were about the better design of some buttons, for example, the close session button, in addition to including aids or descriptive texts in the icons of the different functions. One limitation that was found in the investigation was the lack of users of the National Police, who had provided more objective suggestions on the usability of the system. This was a consequence of the nature of the project since I-Pat is a private government system.

6 Conclusions

Safety is an important necessity in the development of society. For this reason, the National Police of Ecuador sought a solution called I-Pat that, supported by technology, allows patrolling the areas with more conflicts, thus providing a degree of comfort to the population. Being a system of such importance, evaluating usability allows the user to use the system more easily, becoming more familiar with the functions, thus supporting the work of the national police. Usability testing by itself cannot develop a comprehensive list of defects, for this, it is important to use an appropriate mix of methods. In this context, this research allowed a new way of evaluating usability by combining TA and FG, methods that combined their advantages allowing them to cover a greater number of problems within the system. TA protocols were propagated and far more often used as a method for system usability testing, then they are used to gain insight into end users' thought processes in interaction with a system to support the development of usable systems by identifying system deficiencies The application of TA and FG in the early stages of the software life cycle will bring great advantages since the software programs, webs or consumer products can be designed with greater precision, not intuitively anticipating the actions of the client, but predicting them quite accurately. Using methods or techniques in a complementary manner as was observed in the research, was an adequate complement, which, when used prior to the debate with the FG, allowed us to obtain an accumulation of qualitative results, which contributed structural elements for the subsequent debate and, at the same time, captured trends in the concepts of the experts. It is noteworthy that FG expose the process, but do not impose specific tools. However, there is a diversity of resources that can be used in each of its phases, which make it versatile and flexible.

References

1. ISO Standards: 9126 standard (2018). https://www.cse.unsw.edu.au/cs3710/PMmaterials/
2. ISTQB: International software testing qualifications board, advanced level syllabus usability tester (2016). https://www.istqb.org/
3. Brhel, M., Meth, H., Maedche, A., Werder, K.: Exploring principles of user-centered agile software development: a literature review. Inf. Softw. Technol. **61**, 163–181 (2015)
4. Pilco, H., Sanchez-Gordon, S., Calle-Jimenez, T., Rybarczyk, Y., Jadán, J., Villarreal, S., Esparza, W., Acosta-Vargas, P., Guevara, C., Nunes, I.L.: Analysis and improvement of the usability of a tele-rehabilitation platform for hip surgery patients. In: International Conference on Applied Human Factors and Ergonomics, pp. 197–209. Springer (2018)
5. Maniega-Legarda, D.: Aplicación de criterios de usabilidad en sitios web: consejos y pautas para una correcta interpretación. Observatorio TIC: REBIUN Red de Bibliotecas Universitarias (2006)
6. User Experience Professionals' Association: The usability body of knowledge (2018). https://www.usabilitybok.org/think-aloud-testing
7. Guan, Z., Lee, S., Cuddihy, E., Ramey, J.: The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1253–1262. ACM (2006)
8. Olmsted-Hawala, E.L., Murphy, E.D., Hawala, S., Ashenfelter, K.T.: Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data dissemination web sites for usability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2381–2390. ACM (2010)
9. McDonald, S., McGarry, K., Willis, L.M.: Thinking-aloud about web navigation: the relationship between think-aloud instructions, task difficulty and performance. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, pp. 2037–2041. SAGE Publications, Los Angeles (2013)
10. Hertzum, M., Hansen, K.D., Andersen, H.H.: Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? Behav. Inf. Technol. **28**(2), 165–181 (2009)
11. Barnum, C.M.: Usability testing essentials: ready, set… test!. Elsevier, Amsterdam (2010)
12. Van Den Haak, M.J., De Jong, M.D., Schellens, P.J.: Constructive interaction: an analysis of verbal interaction in a usability setting. IEEE Trans. Prof. Commun. **49**(4), 311–324 (2006)

13. Makri, S., Blandford, A., Cox, A.L.: This is what im doing and why: methodological reflections on a naturalistic think-aloud study of interactive information behaviour. Inf. Process. Manage. **47**(3), 336–348 (2011)
14. Nørgaard, M., Hornbæk, K.: What do usability evaluators do in practice? An explorative study of think-aloud testing. In: Proceedings of the 6th Conference on Designing Interactive Systems, pp. 209–218. ACM (2006)
15. Molich, R., Ede, M.R., Kaasgaard, K., Karyukin, B.: Comparative usability evaluation. Behav. Inf. Technol. **23**(1), 65–74 (2004)
16. Gulliksen, J., Boivie, I., Persson, J., Hektor, A., Herulf, L.: Making a difference: a survey of the usability profession in Sweden. In: Proceedings of the Third Nordic-Conference on Human-Computer Interaction, pp. 207–215. ACM (2004)
17. Nyumba, T.O., Wilson, K., Derrick, C.J., Mukherjee, N.: The use of focus group discussion methodology: Insights from two decades of application in conservation. Methods Ecol. Evol. **9**(1), 20–32 (2018)
18. Plummer-D'Amato, P.: Focus group methodology part 2: considerations for analysis. Int. J. Ther. Rehabil. **15**(3), 123–129 (2008)
19. Onwuegbuzie, A.J., Dickinson, W.B., Leech, N.L., Zoran, A.G.: A qualitative framework for collecting and analyzing data in focus group research. Int. J. Qual. Methods **8**(3), 1–21 (2009)
20. Kontio, J., Lehtola, L., Bragge, J.: Using the focus group method in software engineering: obtaining practitioner and user experiences. In: Proceedings. 2004 International Symposium on Empirical Software Engineering, ISESE 2004, pp. 271–280. IEEE (2004)
21. Sunikka, A.: Usability evaluation of the helsinki school of economics website. Master's thesis (2004)
22. Nielsen, J.: The use and misuse of focus groups. IEEE Softw. **14**(1), 94–95 (1997)
23. Obeso, M.E.A., Lovelle, J.M.C., Prieto, A.B.M.: Metodología de medición y evaluación de la usabilidad en sitios web educativos. Universidad de Oviedo (2005)
24. Guevara, C., Jadan, J., Zapata, C., Martínez, L., Pozo, J., Manjarres, E.: Model of dynamic routes for intelligent police patrolling. In: Multidisciplinary Digital Publishing Institute Proceedings, vol. 2, p. 1214 (2018)
25. Habibi, M.R.M., Khajouei, R., Eslami, S., Jangi, M., Ghalibaf, A.K., Zangouei, S., et al.: Usability testing of bed information management system: a think-aloud method. J. Adv. Pharm. Technol. Res. **9**(4), 153 (2018)
26. Harper, B.D., Norman, K.L.: Improving user satisfaction: the questionnaire for user interaction satisfaction version 5.5. In: Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference, pp. 224–228 (1993)
27. Arh, T., Blazic, B.J.: A case study of usability testing-the sumi evaluation approach of the educanext portal. WSEAS Trans. Inf. Sci. Appl. **5**(2), 175–181 (2008)
28. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usabil. Eval. Ind. **189**(194), 4–7 (1996)
29. Nielsen, J.: Severity ratings for usability problems. Pap. Essays **54**, 1–2 (1995)

Disaster Management Support System Prototype Design Evolution Based on UX Testing

Mário Simões-Marques[1(✉)], Pedro Mendonça[2], Diogo Figueiredo[3], and Isabel L. Nunes[3,4]

[1] CINAV – Portuguese Navy, Alfeite, 2810-001 Almada, Portugal
mj.simoes.marques@gmail.com
[2] Laboratório de Óptica, Lasers e Sistemas, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal
pfmendonca@fc.ul.pt
[3] Faculty of Sciences and Technology, Universidade NOVA de Lisboa, 2829-516 Costa de 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 Costa de Caparica, Portugal

Abstract. The analysis of past disaster response operations led to recommendations for improving disaster management, namely on the areas of situational awareness, decision support, and organizational agility. Information technologies offer solutions with the potential to answer these recommendations. This was the reason for launching the THEMIS (disTributed Holistic Emergency Management Intelligent System) project, which followed the UCD approach. THEMIS is an Intelligent System that applies Artificial Intelligence methodologies in support of disaster managers and first responders. The paper addresses the activities regarding designing and testing the user interaction of a digital prototype developed for disaster managers, describing the procedures adopted, the test conditions, test population, and the evolution of prototypes based on test findings. The paper also addresses the use of eye-tracking to support the analysis of tests, helping on the identification of user errors or problems faced while performing the tasks established in a validated test script.

Keywords: Usability · Disaster management · Eye-tracking · THEMIS · Human-systems interaction

1 Introduction

According to the United Nations International Strategy for Disaster Reduction 'Disaster management' is the organization, planning and application of measures preparing for, responding to and recovering from disasters. Particularly during the disaster response stage, disaster management is a complex activity where decision-makers are faced with very demanding problems.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 374–385, 2020.
https://doi.org/10.1007/978-3-030-20040-4_34

The outcome of the analysis of past disaster response operations usually includes recommendations for improving disaster management, addressing the need for improved situational awareness and a common operating picture; improved decision support and resource tracking and allocation; and greater organizational agility for disaster management. Information technologies (namely through the use of intelligent systems) have the potential to provide solutions that give a robust answer to these recommendations offering decision support to disaster-managers. This was the context and the rational for the development of the THEMIS (disTributed Holistic Emergency Management Intelligent System) project. THEMIS is an Intelligent System that applies Artificial Intelligence methodologies in support of disaster management decision-makers and first responders. Different perspectives of the THEMIS project were discussed, for instance in [1–5].

The project followed the UCD approach. After the steps devoted to understanding and specifying the context of use, and to specifying the user and organizational requirements, the team engaged in designing and testing user interaction prototypes. The paper addresses such activities regarding the desktop user interface digital prototype developed for disaster managers, describing the procedures adopted, the test conditions, test population, and the evolution of prototypes as a consequence of the test findings. The paper also addresses the use of eye-tracking to support the analysis of the tests, helping the identification of the causes of errors or problems that users faced while performing the tasks established in a validated test script.

The used eye-tracker was Pupil from Pupil-Labs [6] and has 3 cameras: 1 world camera with a sensor able to acquire images of 1920 $\times$ 1080 pixels at 30 fps (frames per second) (or higher fps with lower resolution), located on the forehead of the user; and 2 other cameras, able to acquire images of 400 $\times$ 400 pixels at 120 fps, one assigned for each eye. The Pupil eye-tracker makes use of the “dark pupil” detection method [7, 8]. In this method, the subject eyes are illuminated with a surface mounted Infrared (IR) LED emitting at 860 nm wavelength, and IR images are acquired using an IR band-pass filter. Results of performance evaluation show, that under ideal conditions, Pupil can provide an average gaze estimation accuracy of 0.6 $\pm$ 0.08 degree of visual angle with a processing pipeline latency of only 0.045 s [7].

2 Interface Design

After an initial brainstorming process for the interface design based on drawings in paper and a first validation based on a *Wizard of Oz* approach, a first desktop prototype was developed using Balsamiq[1], since this is a quite user-friendly and easy to use prototyping environment. This first version was validated by an expert and subsequently subject to usability tests, after creating and validating the tests’ script. In these tests it was possible to use a Pupil Labs eye-tracker to further collect data such as the users’ gaze. After the first test round analysis changes were introduced in the UI and the tests were repeated with the improved prototype. Figures 1 and 2 illustrate an example

[1] Balsamiq is a low-fidelity wireframing software, accessible at https://balsamiq.com/.

of some of these UI changes including layout improvement (e.g., inclusion of icons, highlighting of command buttons).

The tests involved 20 users from the Naval Academy, which were divided in two groups, one engaged in each of the test rounds[2]. The test users' ages ranged from 22 to 28 years old, being 6 male and 14 female. Regarding education all users had a university degree (e.g., law, nursing, psychology, environment).

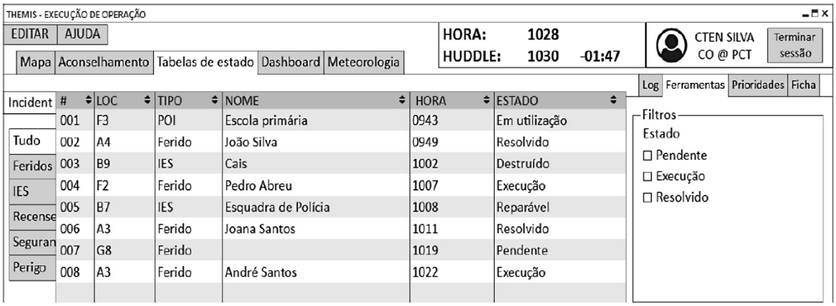


Fig. 1. First version of the desktop prototype

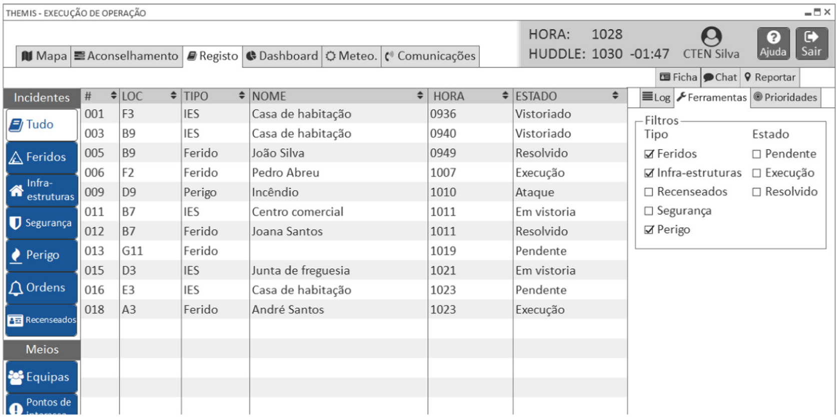


Fig. 2. Final version of the desktop prototype

2.1 Task Success

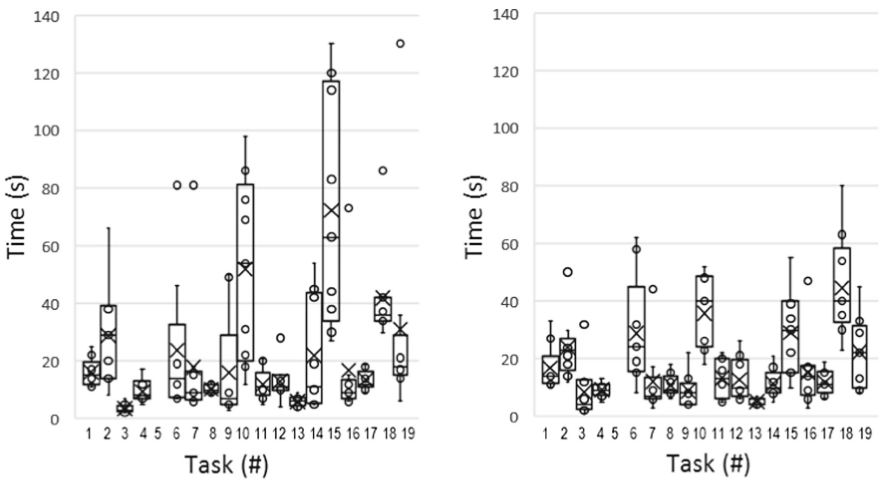
Almost all tasks in both rounds were successful. The first round had a 99.4% success in task completion while the second round had 99.5% success in task completion.

[2] In parallel with the tests discussed in this paper there were tests conducted to a mobile app prototype. The user groups swapped from prototype type from Round 1 to Round 2.

The lack in success can also be explained by the fact that the test users are not familiar with emergency management and disaster relief operations and exercises. Another explanation could be the tasks not being completely clear in what is its goal, causing the users to not understand what is expected that they do.

2.2 Efficiency (Task Time)

While the time taken to complete a task can't be compared between tasks, due to differences in task difficulty and complexity, it is possible to calculate the average time each task took to complete, and compare it between rounds of tests, to understand how the changes in the prototype affected the system's usability. The comparison of results synthetized in Fig. 3(a) and (b), allows to perceive the improvements reached in the average time taken to complete the tasks. The coming discussion will detail the evolution of the results regarding three tasks (#2, #10 and #15).



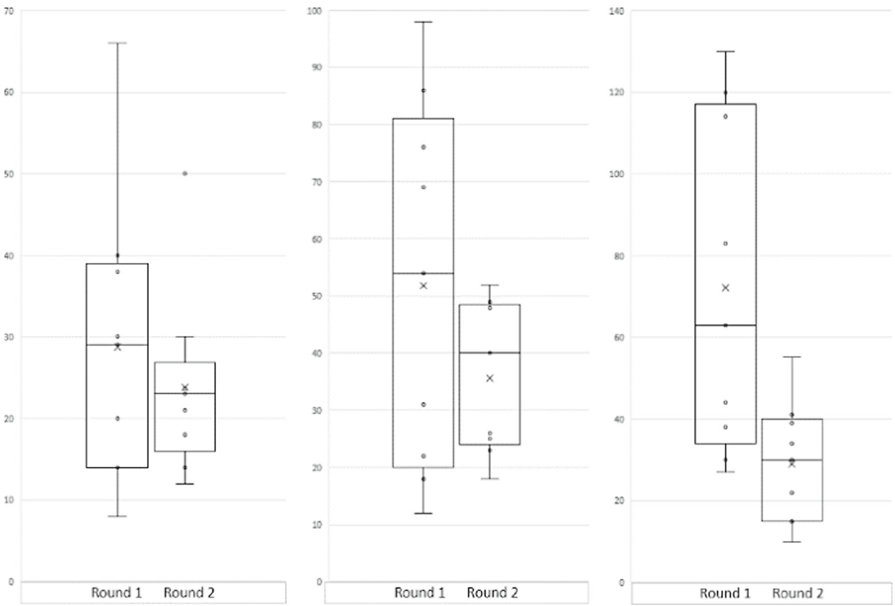
(a) Round 1 (b) Round 2

Fig. 3. Average time taken in each task on the test rounds of desktop prototype

In Task #2, where users were asked to "Import, from operation DISTEX_2018, the information regarding 'Disaster', 'Tasks/Priorities' and 'Communications Plan'", there was an improvement of 23.6%, from 28.8 s to 22.0, as it is possible to conclude by analyzing Fig. 4(a).

Regarding Task #10, where users were asked to "Count and identify the number wounded people shown in the map.", there was a more noticeable improvement of 35.3%, from 51.8 s to 33.5 s, as illustrated in Fig. 4(b). This can be due to the fact that the users in Round 2 had already experience with the prototype's terms and icons, which made it easier for them to recognize what was asked of them, as well as a more streamlined version of the UI.

In Task #15, where users were asked to “Check infrastructure #001 record. How many injured were found there?”, registered the biggest improvement of 59.8%, going from 72.1 s on average to 29.0 s, as seen on Fig. 4(c). This can be attributed to the fact that the UI in “Registo” has easier to see buttons to select the desired information.



(a) Task 2 (b) Task 10 (c) Task 15

Fig. 4. Comparison between average time taken to complete tasks in Round 1 and Round 2

3 Eye Tracker Tests

During the tests, Pupil Labs’ eye tracking hardware and software was used. As stated in the introduction, Pupil has 3 cameras: 2 pointing at each eye and a world camera located on the forehead of the user (as illustrated in Fig. 5). One example of the view of this world camera can be seen in Fig. 6 where the corners of the screen monitor used to run the test have 4 markers. These are 5-bit markers used to define the Area of Interest (called surface) within the monitor, to be analyzed by the Pupil Player software (also being executed on the Fig. 6). The calibration process for each user, after manually obtaining focus on the pupils, was performed by Pupil Capture and required the user to stare at each of a 5-target points (4 corners + center). At the end of the process deviation vectors at each of the 5-target points were drawn and further adjustments of the eye cameras were done if needed.

Gaze heatmaps were generated by Pupil Player for the defined surface, in this case the full screen in order to cover the complete application. The surface is divided into a

2D grid and colors are normalized such that cells with the highest amount of gaze are displayed in red and those with the least gaze are displayed in blue.

For each task below, a comparison between prototype versions is shown. The red circles indicate which UI elements users should have looked to complete said task.

(a) Set up (b) Hardware

Fig. 5. Eye tracking tests

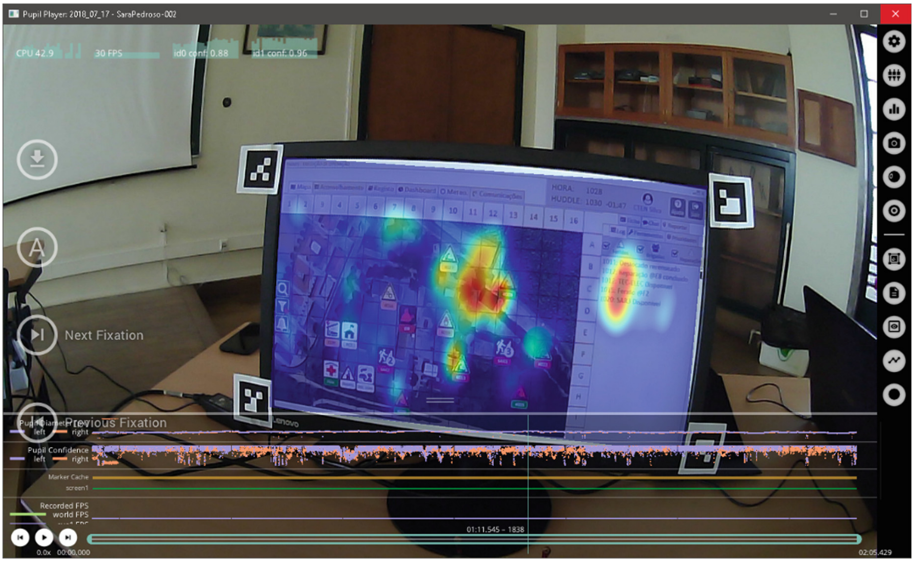


Fig. 6. Example of the field of view of the world camera where the 5-bit markers defining the Area of Interest (AOI or surface) can be seen and a gaze heatmap generated by the Pupil Player is shown for a specific task.

3.1 Task #2

In this task users were asked to import specific data from an already existing operation, which required them to find the import button, select the correct operation from a list, choose the information to import and then confirm the selection. Figure 7 illustrates part of the initial version of the UI involved in the first activity (find the import button). Figure 8 presents the second version, where icons were added to every button or interactable UI element, with the purpose of helping users identify the different functions. The user information and log out button were also changed to take less space on the UI, as well as the size of the team configuration area (on the right side).

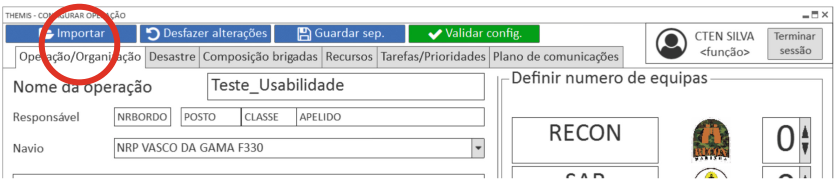


Fig. 7. First prototype version (the circle marks the UI element required to complete Task #2)

Fig. 8. Second prototype version (the circle marks the UI element required to complete Task #2)

Comparing the eye tracker heat maps, illustrated in Figs. 9 and 10, it is possible to observe less spread of the gaze in the second one, which is a more streamlined version of the interface.

Fig. 9. Heat map of Task #2's user interaction for the first version of the prototype

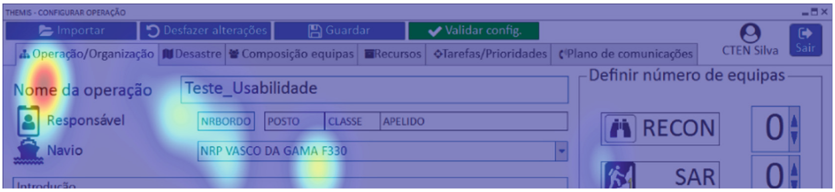


Fig. 10. Heat map of Task #2's user interaction for the second version of the prototype

This means the user looked at less UI elements than the first version, while searching for the element that would allow them to complete the task. In this case, the user was looking for the button "Importar". It is possible to see that in the first version the user looked for it in the tabs area, while in the second version the user mostly looked at the top left region of the UI, where the button is located.

3.2 Task #10

In this task users were asked to identify the number of incidents corresponding to injured people displayed in the map and the corresponding identification number. Considering the two versions displayed on Figs. 11 and 12, the buttons on the left side of the UI were reduced to the most critical functions so as not to pollute the map with unnecessary elements, and the remaining were also changed to comply with the standard appearance adopted for the UI elements throughout the whole prototype.

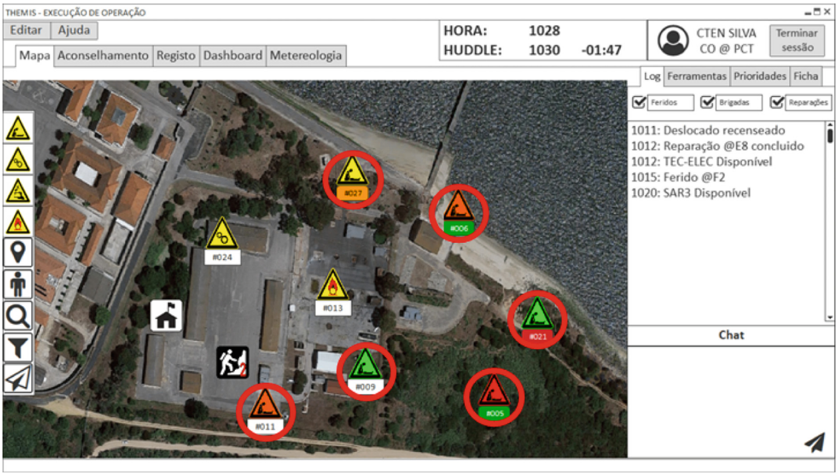


Fig. 11. First prototype version (circles mark the UI element addressed by Task #10)

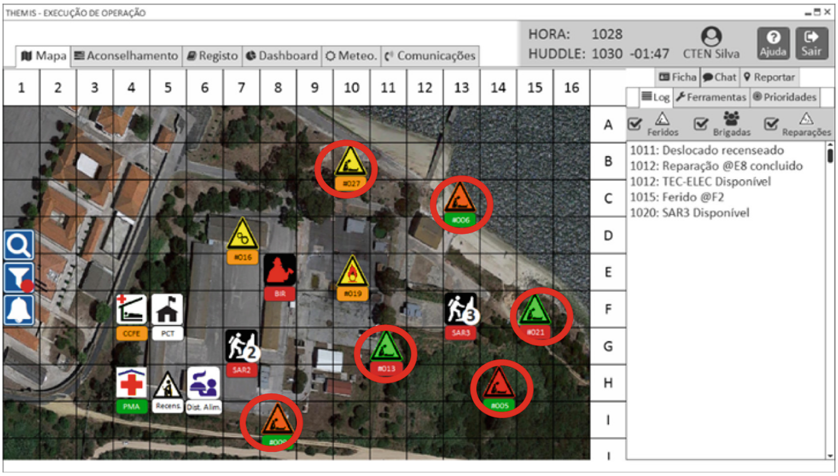


Fig. 12. Second prototype version (circles mark the UI element addressed by Task #10)

More icons were added to the map to better simulate the real scenario, and color codes were also added to their labels to help identify the response status of the incident. New functions were added to the tabs on the right side to make them easier to access and use, as well as the left side header becoming more streamlined and a new map grid.

After reducing the number of elements on the left command bar, the map becomes easier to browse, as becomes evident by comparing the heat maps illustrated in Figs. 13 and 14. Considering that users were asked to identify injured people in the map, it is possible to see that the first prototype version (Fig. 13) the user was distracted by the button bar to the left of the map, as opposed to the second version (Fig. 14) where the user only focused on the icons of the map, especially the injured icons, as asked.

Fig. 13. Heat map of Task #10's user interaction for the first version of the prototype

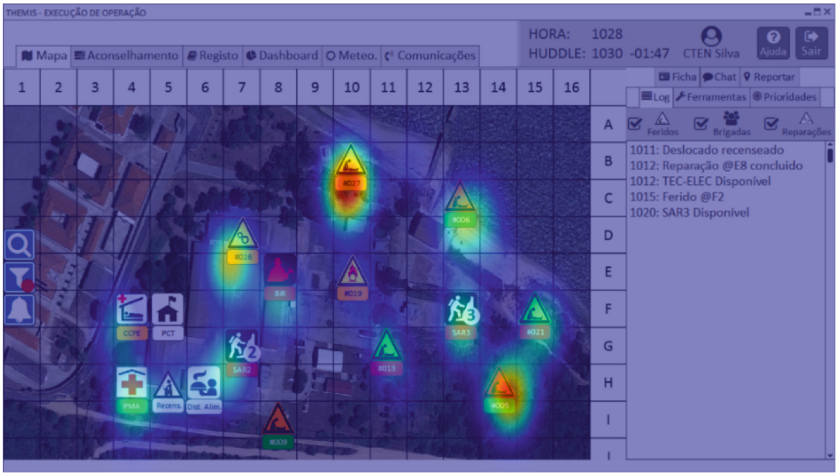
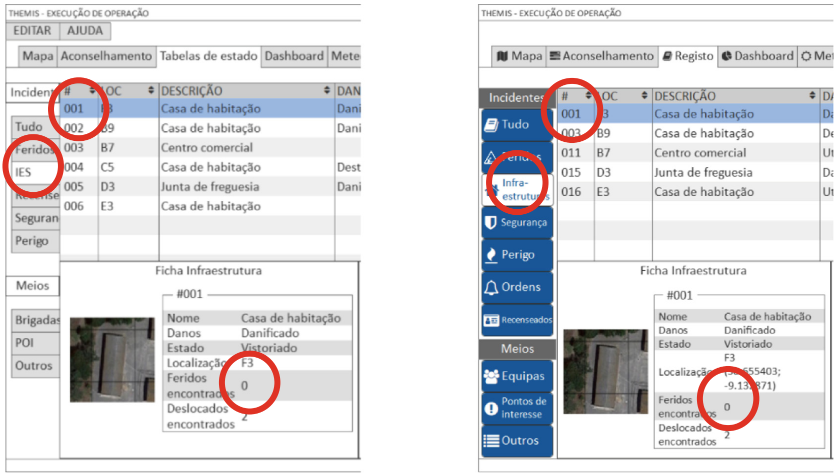


Fig. 14. Heat map of Task #10's user interaction for the second version of the prototype

3.3 Task #15

In this task users were asked to find the number of injured people found in a specific infrastructure. This required users to go to the registry tab, select the button that gives access to the information regarding infrastructures, find the correct infrastructure in a list and access to the corresponding information. Figure 15 illustrates part of the UI used for the last activity of the task.

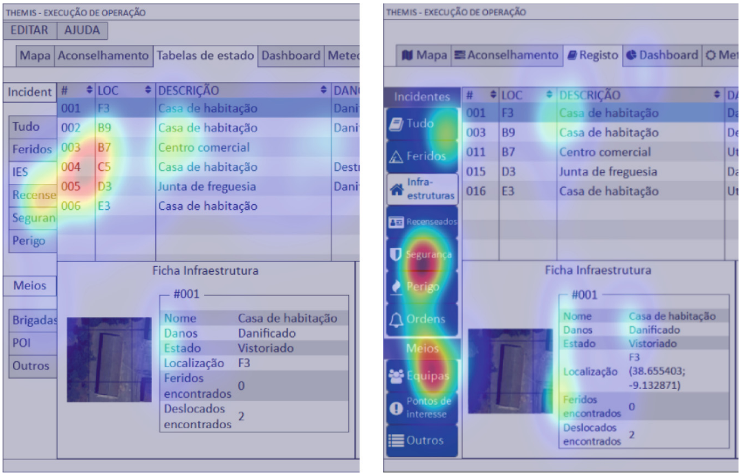
The initial version, illustrated in Fig. 15(a), was improved by the inclusion of icons in the top tabs to help in identifying their purpose, and by the replacement of the tabs to the left by buttons that better comply with the standard layout adopted for the whole prototype. The second version is illustrated in Fig. 15(b).



(a) First version (b) Second version

Fig. 15. Prototype versions (circles mark the UI element addressed by Task #15)

Regarding the first version it is possible to observe, in Fig. 16(a), that rather than focusing the buttons to the left that give access to the desired information, the user gazed persistently to a table containing information. In the second version (Fig. 16(b)) most gaze points are centered in the button area and on the information that the user was asked to retrieve.



(a) First version (b) Second version

Fig. 16. Heat map of Task #15's user interaction for the two versions of the prototype

4 Conclusions

The paper presents the THEMIS UI design evolution process performed based on the UCD approach, addressing the activities regarding improving the desktop digital prototype design. This included describing the procedures adopted, the test conditions, test population, and the evolution of prototypes as a consequence of the test findings. The paper also addresses the use of eye-tracking to support the analysis of the tests, helping the identification of the causes of errors or problems that users faced while performing the tasks established in a validated test script.

Between the initial and the second versions of the prototype significant improvements were reached in terms of Usability, assessed in terms of efficiency and effectiveness.

Furthermore, the use of eye tracking technology was key to gain better perception of users' behavior during the tests, identifying the problems faced while performing the script tasks.

Acknowledgments. The work was funded by the Portuguese Ministry of Defense and by the Portuguese Navy and by project UID/EMS/00667/2013. Pedro Mendonça acknowledges the financial support from the Portuguese Foundation for Science and Technology (FCT), PD/BD/52329/2013 and "CAMELOT - C2 Advanced Multi-domain Environment and Live

Observation Technologies" Project ID: 740736, funded under the European Commission grants H2020-EU.3.7.3. - Strengthen security through border management; H2020-EU.3.7.7. - Enhance standardization and interoperability of systems, including for emergency purposes.

References

1. Simões-Marques, M., Teodoro, M.F., Calhamonas, G., Nunes, I.L.: Applying a variation of Delphi method for knowledge elicitation in the context of an intelligent system design. In: Nunes, I. (ed.) Advances in Human Factors and Systems Interaction, AHFE 2019. AISC, vol. 959, pp. 386–398. Springer, Cham (2020)
2. Simões-Marques, M., Figueira, J.R.: How can AI help reduce the burden of Disaster Management decision-making? In: Nunes, I. (ed.) Advances in Human Factors and Systems Interaction, AHFE 2018. Advances in Intelligent Systems and Computing, vol. 781, pp. 122–133. Springer, Cham (2019)
3. Nunes, I.L., Lucas, R., Simões-Marques, M., Correia, N.: An augmented reality application to support deployed emergency teams. In: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (eds.) Proceedings of the 20th Congress of the International Ergonomics Association, IEA 2018. Advances in Intelligent Systems and Computing, vol. 822, pp. 195–204. Springer, Cham (2019)
4. Simões-Marques, M., Correia, A., Teodoro, M.F., Nunes, I.L.: Empirical studies in user experience of an emergency management system. In: Nunes, I. (ed.) Advances in Human Factors and Systems Interaction, AHFE 2017. Advances in Intelligent Systems and Computing, vol. 592, pp. 97–108. Springer, Cham (2018)
5. 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, AHFE 2017. Advances in Intelligent Systems and Computing, vol. 592, pp. 109–120. Springer, Cham (2018)
6. Pupil Labs Open source eye tracking. http://pupillabs.com. Accessed 23 Nov 2018
7. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. https://doi.org/10.1145/2638728.2641695. Accessed 26 Nov 2018
8. Pupil Capture and Pupil Player software. https://github.com/pupil-labs/pupil/releases/latest. Accessed 23 Nov 2018

Applying a Variation of Delphi Method for Knowledge Elicitation in the Context of an Intelligent System Design

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

[1] CINAV – Portuguese Navy, Alfeite, 2810-001 Almada, Portugal
mj.simoes.marques@gmail.com,
maria.alves.teodoro@marinha.pt
[2] CEMAT, Instituto Superior Técnico, Lisbon University, Av. Rovisco Pais, 1, 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. The effectiveness, efficiency and satisfaction are key indicators of the Usability of a system. The outcome of the knowledge management process associated with the development of intelligent systems is determinant in this regard. The paper describes the use of the Delphi method to identify the level of fitness that subject matter experts (SMEs) recognize on different types of response units to perform specific disaster response tasks, with the purpose of expliciting the knowledge required for the inference process of the THEMIS intelligent system. Since the SMEs involved in this elicitation process had different levels of experience on different domains of disaster management, to help reaching consensus the Delphi method was modified to weight the level of experience expressed by the SMEs. This variant of the Delphi method is presented, discussing the approach adopted in each step of the iterative process, the obtained results, and the followed validation process.

Keywords: Knowledge management · Disaster management · Fuzzy · THEMIS · Human-systems interaction

1 Introduction

Recognizing the commonly pointed gap regarding the limited exploitation of the potential offered by Intelligent Systems in support of disasters relief operations (DRO) the Portuguese Ministry of Defense funded the THEMIS project; whose aim is to create a distributed emergency management Intelligent System that can support disaster managers and first responders engaged in international DRO. Different perspectives of the THEMIS project were discussed, for instance in [1–5].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 386–398, 2020.
https://doi.org/10.1007/978-3-030-20040-4_35

'Intelligent Systems' is a generic designation to refer computer-based tools used to provide support to decision-making or problem-solving processes using Artificial Intelligence. Such tools are particularly relevant when the decision factors are complex, or the timeliness of the decision is critical, making the decision-makers' task difficult, for instance due to the amount of information to process, because of the involved uncertainty and vagueness, or because of the stressful pressure that results from the environment or the impact of the decision.

Intelligent systems present a generic architecture composed by four building blocks that perform the following functions:

- Knowledge Base - this component stores the knowledge required for the specific problem-solving application;
- Working Memory - this component stores data about the problem context to analyze;
- Inference Engine - this component runs the knowledge against the data, assessing the concrete situation and generating the conclusions and advise, as well as any explanations required by the user;
- User Interface - this component supports the human-computer interaction necessary to input data, insert requests and obtain system outputs.

THEMIS' Knowledge Base contains knowledge intrinsic to the response resource types (namely, organization, and skills offered by the response teams), intrinsic to the disaster incident types (namely, tasks required to handle the incidents, and skills required to perform the tasks) and relating resources to incident response tasks (response time, skills matching).

Eliciting the knowledge for implementing the THEMIS posed quite demanding challenges, particularly regarding the acquisition of tacit knowledge used by decision-makers in their reasoning processes.

The knowledge acquisition considering the specific problem addressed in this paper involved a group of SMEs. The Delphi method [6] was identified as a particularly suited approach to deal with the considered problem, since its goal is to consolidate the views of SMEs of a particular domain, based on an iterative process designed to converge to a commonly agreed opinion on a specific issue.

Since mid-20th century, there are numerous references in many distinct areas where the Delphi method was applied. For example, in industrial engineering, namely the construction of an index in automotive supply chain [7], transportation [8], in paper pulp production [9], education [10], natural sciences [11], etc. Paper [12] offers numerous descriptions of Delphi method applications. The main idea of this method can be summarized as: The Delphi method intends to study situations of lack of historical data, in which it is intended to define new ideas; this is done consulting a group of experts through an iterative questionnaire that can be circulated several times by the same team, the aim is to reach a consensus on the answers and a solid conclusion of a particular study; this method is based on the use of the knowledge and experience of the chosen panel, assuming that a group's opinion best portrays the reality better than the opinion of just one individual.

As discussed by Adler and Ziglio [13], the Delphi method is a structured process to collect and synthesize knowledge in a specific area, obtained from a group of experts, using an iterative process of questionnaires, providing feedback to them on the results obtained in previous iterations. As Wissema [14] points out, the Delphi method has the advantage of avoiding the problems associated with the traditional techniques of consensus of group opinions, namely the “Focus Groups”, that can create problems of bias in the answers, due to the presence of opinion leaders.

The goal of the study presented here was to elicit knowledge on the level of fitness that disaster managers recognize between different types of response units (e.g., search and rescue, technical, medical) and specific disaster response tasks (e.g., first-aid, stabilization of structures, rescue of entrapped victims, firefighting, restoring of basic services, feeding and sheltering), with the purpose of making explicit the knowledge required for the inference process of the THEMIS intelligent system.

Considering that all SMEs had previous experience in disaster management but in different technical areas of disaster response (i.e., different degrees of expertise regarding multiple cross-domain areas), and to help reaching convergence, the DELPHI method was applied considering an index of weighted opinions according with the level of experience and confidence on the answers (as stated by the SMEs). Some discussion about how obtain these weights can be found in [15].

Therefore, the variant of the Delphi method was used to establish the importance that SMEs assign to the selection of specific response teams to perform typical tasks required in DRO.

The scope of the Delphi questionnaires was defined in meetings with SMEs of the Navy, corresponding to the importance to be attributed to each team to perform specific tasks of emergency response in the process of resource allocation that supports the definition of orders for employment of the teams. The results of this process supported the development of THEMIS’ inference model necessary to prioritize the team selection to respond to pending incidents. The work had as starting point the definition of the organization to consider, which allowed the enumeration of the response teams, presented in Table 1.

Table 1. Mobile teams engaged in disaster response operations

Team
Reconnaissance (RECON)
Search and Rescue (SAR)
Urban Search and Rescue (USAR)
Search and Rescue - Structures (STRUCT)
Mechanical (MECHANIC)
Electricity (ELECTRIC)
Water and Sanitation (SANITATION)
Firefighting (FIRE)
Medical (MEDICAL)
Logistics (LOG)
Food (FOOD)

The established response organization is presented in the chart of Fig. 1.

Similarly, it was necessary to enumerate the set of typical tasks to be considered in the response to emergency situations. To this end, meetings were held with navy officers with experience in emergency cases, documents were collected and analyzed on exercises performed and through direct observation of a DISTEX exercise (simulation of humanitarian assistance in a catastrophe situation) carried out at the Portuguese Navy Naval Base. Subsequently, the data collected were validated by SMEs with experience in this type of operations.

In order to implement and analyze the questionnaires, several platforms that could support the Delphi method were analyzed and tested. Thus, we have tested: (i) the "SurveyMonkey.com" website, rejected due to the free version questions limit; (ii) the "Google Forms" platform, rejected due to the impossibility of submitting, in future questionnaires, the data on the questionnaires already carried out; (iii) the website "armstrong.wharton.upenn.edu" developed by J. Scott Armstrong in partnership with the International Institute of Forecasters, rejected because of the inability of respondents to review their answers in completed questionnaires. Therefore, it was decided to implement the questionnaires in Excel format, considering the ease of sending (e-mail), data processing, as well as the possibility of always maintaining the same platform throughout the several rounds that would need to be performed.

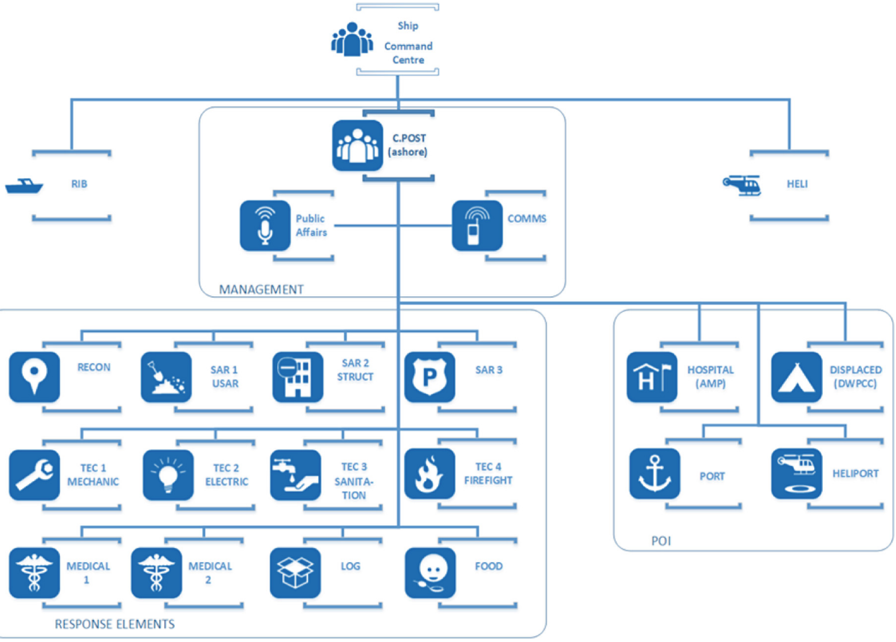


Fig. 1. Organizational chart of the deployed response structure

After the preparatory stage, several rounds of questionnaires were carried out to a group of SMEs identified as having adequate experience in the field. The conduction and analysis of the results from the several rounds of Delphi method are described in the next sections of this manuscript.

2 Applying the Delphi Method

2.1 Characterization of SMEs

The questionnaire used in the first round of questionnaires consisted of a first part, designed to characterize the population of responding SMEs, based on the following information: age, gender, rank, class, experience in response to disasters (real, training) and boarding time.

The questionnaires were sent to about two dozen SMEs, and responses were obtained from 12 SMEs, all males and all with 5 or more years of experience at sea. The ages ranged from 35 to 54 years old, being 33% in the age group [35–44] and 67% in the age group [45–54]. The SMEs were all Navy senior officers (25% Lieutenant-Commanders, 58% Commanders and 17% Captains). Most of the respondents were from the Warfare branch (only one was from the Naval Engineers branch). Regarding experience in disaster response operations, all had experience, 75% had obtained it in training exercises, while 25% had been involved in real operations.

2.2 Round 1

In the body of the questionnaire, respondents were asked to rate, on a 6-point Likert scale, the importance (1-Not Important to 6-Extremely Important) of the teams identified (in Table 1) to complete the tasks during a humanitarian assistance operation.

The objective is to identify the degree of priority in assigning each team to perform a given task. For example, with respect to a medical team, a nurse is perfectly capable of performing the task "Prepare and distribute food to wounded", but due to his/her competencies he/she becomes more important that he/she is assigned to the task " Provide first aid". Thus, it is expected for this team that the task "Provide first aid" has a rating close to 6, while for the task " Prepare and distribute food to wounded", a rating close to 1 is expected.

The Likert scale was chosen with an even number of points to avoid neutral responses, forcing respondents to take a positive or negative position for each question. Traditionally in Delphi studies the scales range from 3 to 11 points. The choice of the number of points in the scale considered the comments of Miller [16] and Baddeley [17], according to which a scale of 11 or more points becomes very complex to justify the choice of answer, while, on the other hand, a scale of 5 or fewer points may make it difficult to draw meaningful conclusions. Different proposals for the Likert scale can be found in literature, for example, Mead and Moseley [18] suggest a 7-point scale. However, in order to avoid neutral responses (except for 0 - Do not know/not responding, which is not counted statistically) the 6-point scale was chosen (1 - Not Important, 2 - Little Important, 3 - Somewhat Important, 4 - Quite Important, 5 - Very Important, 6 - Extremely Important).

The questionnaire was organized by grouping the 11 teams into 5 brigades in order to facilitate the response, since the teams of the same brigade had the greatest similarity:

- Reconnaissance Brigade - Reconnaissance team;
- Search and Rescue Brigade - SAR, USAR and Structures teams;
- Logistics Brigade - Supply and Food teams;
- Technical Brigade - Mechanical, Electricity, Water and Sanitation, and Firefighting teams;
- Medical Brigade - Medical teams.

With 50 tasks and 11 teams, 550 questions were provided to each SME. In addition, respondents were asked to indicate their level of experience with each team and the degree of confidence they had in the responses given. Both questions were equally ranked using a 6-point Likert scale.

Figure 2 shows part of an Excel sheet filled out by a SME, containing in the first line the experience of the responder regarding the team, in the last line the level of confidence on the correctness of the answers, and in the intermediate lines the answers regarding the 3 SAR teams that integrate the Search and Rescue Brigade. The methodology used to select whether the answers reached consensus or not, was based on the Interquartile Range (IQR) metric, that is, on the difference between the 1st and 3rd Quartiles, which represents 50% of the closest observations of the median. When this difference is less than 1 it means that more than 50% of the responses obtained are within 1 point of the Likert scale (as discussed by De Vet, Brug, De Nooijer, Dijkstra and De Vries [10]). Notice that different metrics can be proposed. In [15] the results applying distinct dispersion metrics were compared, but the obtained results are similar.

This approach was compared with another one, where a weighting factor was applied based on the experience indicated by each SME (Table 2), so that the opinion of a more experienced SME had a more significant weight in the weighted average (WA) the importance of a certain team to perform a task and vice versa.

Level of expertize regarding this team?	4-Strong	4-Strong	4-Strong
	SAR Team	Urban-SAR Team	SAR_ Structures Team
Identify incidents	5 - Very Important	5 - Very Important	5 - Very Important
Triage of affected and injured people	3 - Somehow Important	3 - Somehow Important	3 - Somehow Important
Provide first aid	6 - Extremely Important	6 - Extremely Important	6 - Extremely Important
Affected people registration	1 - Not Important	1 - Not Important	1 - Not Important
Identify location for [point of interest]	5 - Very Important	5 - Very Important	5 - Very Important
• • •			
Stabilize structures	3 - Somehow Important	6 - Extremely Important	6 - Extremely Important
Clear roads and pathways	6 - Extremely Important	6 - Extremely Important	6 - Extremely Important
Build support structures for rescue	6 - Extremely Important	6 - Extremely Important	6 - Extremely Important
Fire fighting	2 - Little Important	2 - Little Important	2 - Little Important
Flood fighting	2 - Little Important	2 - Little Important	2 - Little Important
Perform anchoring	3 - Somehow Important	5 - Very Important	6 - Extremely Important
Distribute drinking water	6 - Extremely Important	6 - Extremely Important	6 - Extremely Important
• • •			
Evacuate team to [point of interest]	2 - Little Important	2 - Little Important	2 - Little Important
Evacuate affected population to [point of interest]	2 - Little Important	2 - Little Important	2 - Little Important
Evacuate animals	1 - Not Important	1 - Not Important	1 - Not Important
Distribute animal feed	1 - Not Important	1 - Not Important	1 - Not Important
Bury dead animal	1 - Not Important	1 - Not Important	1 - Not Important
Level of certainty of the answers to the questionnaire?	5-High certainty	5-High certainty	5-High certainty

Fig. 2. Example of responses from SMEs

Table 2. Weighting given to each SME's response based on experience

Experience level	Weight
1 - Extremely weak	0.1
2 - Very weak	0.5
3 - Weak	0.9
4 - Strong	1.1
5 - Very strong	1.5
6 - Extremely strong	1.9

Based on this approach, the results presented in Table 3, where the 'task-team' relationships marked in red are those with an IQR greater than 1. In these cases, it is verified that the SMEs did not reach consensus on how important the team to perform the task, as such and according to the Delphi method, these cases will have to go to the next Round of questionnaires.

Table 3. Sample of the results obtained in Round 1

	RECON		SAR		USAR		STRUCT		FIREF.		SANITAT.		ELECTRIC		MECHAN.		LOG.		FOOD		MEDICAL	
Task	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA
Identify incidents	0,0	5,7	1,3	5,0	1,3	5,0	1,3	4,8	3,0	4,1	3,0	4,1	3,0	4,1	3,0	4,1	2,3	2,9	2,3	2,9	2,5	4,2
Triage of affected and injured people	3,0	4,2	1,3	5,4	2,0	5,2	2,0	4,5	2,0	2,2	2,0	2,2	2,0	2,2	2,0	2,1	1,3	1,9	1,3	1,9	0,3	5,5
Provide first aid	3,0	4,6	1,0	5,4	1,0	5,3	1,3	5,0	2,3	3,0	2,3	3,0	2,3	2,9	2,3	2,9	2,3	2,9	2,3	2,9	0,0	5,8
Affected people registration	2,5	3,2	2,3	2,9	1,5	2,8	2,3	2,6	1,0	1,5	1,0	1,5	1,0	1,5	1,0	1,5	1,0	2,2	1,0	1,3	2,5	3,0
Identify location for [point of interest]	2,3	4,0	2,0	3,6	2,0	3,6	2,0	3,6	1,3	2,0	1,3	2,0	1,3	2,0	1,3	2,0	2,0	2,2	2,0	2,3	3,3	3,6
Transport material to install [point of Interest]	2,0	2,4	2,5	3,2	2,3	3,2	2,3	3,4	2,3	3,0	2,3	3,0	2,3	2,9	2,3	2,9	3,0	4,0	4,0	3,5	3,0	2,9
Instalation of [point of interest]	1,5	2,5	3,3	3,6	3,0	3,5	1,5	4,2	2,0	2,7	2,3	2,8	2,3	2,7	2,3	2,7	3,3	3,3	4,0	3,6	3,0	3,0
Transport severe injured person	1,3	3,6	2,0	5,2	2,3	4,8	2,3	4,5	1,5	2,6	1,3	2,2	1,3	2,2	1,3	2,2	2,0	2,1	2,0	2,1	1,0	5,5
Stabilize severe injured person	3,0	4,4	1,3	5,4	1,3	5,0	2,0	4,7	2,0	2,3	2,0	2,3	2,0	2,4	2,0	2,4	2,0	2,4	2,0	2,4	0,0	6,0
Transport slght injured person	0,3	2,8	2,3	4,6	2,0	4,5	2,0	4,3	2,0	2,3	2,0	2,3	2,0	2,3	2,0	2,3	2,0	2,4	2,0	2,4	2,0	4,6
Rescue trapped victim	0,5	2,9	0,8	5,0	0,0	5,3	0,0	5,5	2,0	2,0	2,0	2,0	2,0	2,1	1,3	2,5	1,0	1,7	1,0	1,7	4,0	3,0
Rescue victim from heights	0,5	2,9	0,0	5,4	2,0	4,6	2,0	4,9	1,0	1,5	1,0	1,5	1,0	1,5	1,0	1,5	1,0	1,4	1,0	1,4	2,3	2,5

A more detailed analysis of the results presented in Table 3 shows that the tasks that reached consensus with a lower IQR have a weighted average of importance close to the extreme values of importance (1 or 6). This fact can be seen more clearly in the graphs shown in Fig. 3, which relates the IQR and the weighted average of the answers to each question. In general, the distribution of collected data may approximate a quadratic function. This means that the SMEs group agrees to the response in cases where it is reasonably obvious that a team is 'Extremely Important' (values close to 6) or 'Not Important' (values close to 1) to accomplish a task. However, there is a great dispersion in all other cases, and consensus is not reached in most situations.

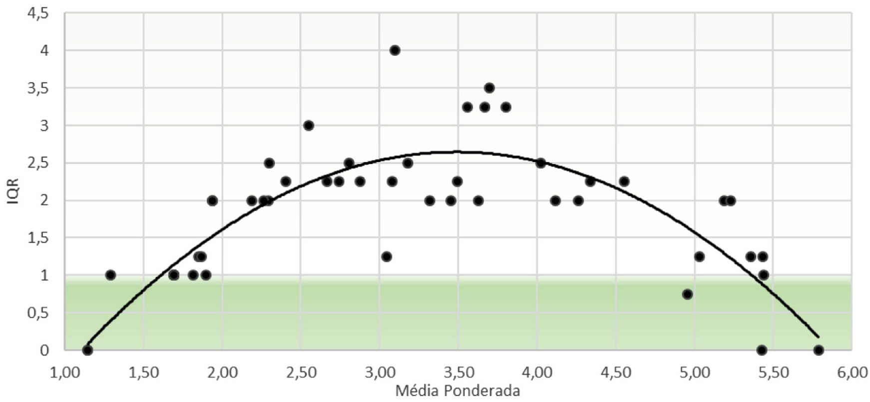


Fig. 3. IQR vs weighted average for the SAR team

The confirmation of this tendency can be obtained by observing the Box Plot diagrams that allow visualizing the dispersion of the responses to the different teams considering the same task. In the example shown in Fig. 4, one can see that the STRUCT team is considered as the most prepared to carry out the task "Stabilizing structures", followed by the USAR team. On the other hand, it is consensual that the Recognition, Logistics and Medical teams are not suitable for the task. The remaining teams present a wide dispersion of responses, with no consensus.

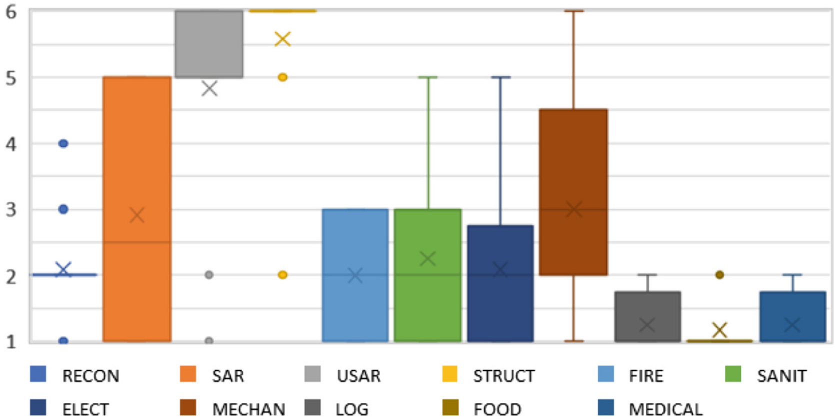
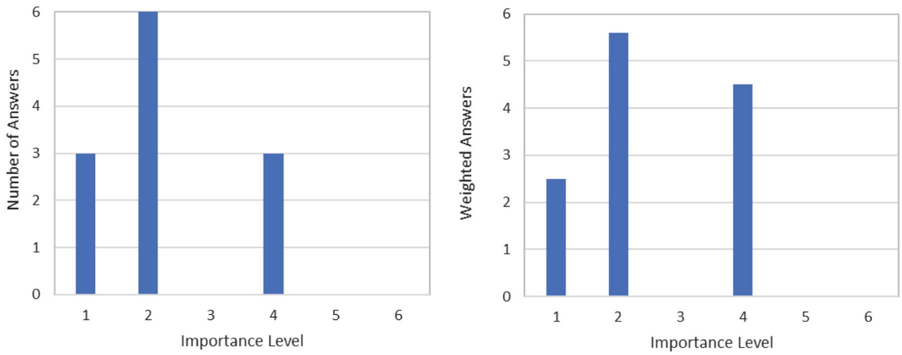


Fig. 4. Boxplot diagram for the task "Stabilizing structures".

In a more detailed analysis of the cases, there are some cases where, although the IQR is less than 1, the observation of the frequency charts does not suggest a consensus. For example, in the case of the "Rescue isolated victim by air" task, shown in Fig. 5(a), for the Reconnaissance team, 9 out of 12 of the SMEs chose a level of importance of 1 or 2, and 3 SMEs chose level 4. The result of the IQR calculation is



(a) Answers frequency (b) Weighted answers

Fig. 5. Answers for the "Rescue isolated victim by air" task.

0.75 which, according to the criterion previously used, would lead to the conclusion that there would be consensus on this task.

In order to deal with these cases (validating or countering the IQR value) we considered the application of the weighting related to the experience of the SMEs (Table 2). Thus, response frequency plots are corrected weighted with the levels of experience, as shown in Fig. 5(b). In this case there is a proximity of the weighted value for the degrees of importance 2 and 4, which suggests that there is no consensus in this task. This is since the 3 SMEs who responded with a level of importance 4 are more experienced with this team, and thus their opinion has a greater weight than the others.

This suggests a new criterion which is that there is no consensus on all questions where there are 2 or more non-adjacent levels of importance that account for at least 20% of the sum of weighted levels of experience. This approach is based on Fuzzy set theory [19] and validates consensus based on the convexity of the aggregated SMEs' opinions.

In the above example, given the weightings, the sum of the experience levels for the Recognition team is 12.6. As such all levels of importance are the ones above the value of α-cut = 2.52 (= 12.6 * 0.2). As levels 2 (5, 6) and 4 (4, 5) are above this limit and are not adjacent, it is concluded that there is no consensus and consequently this question will go to the next round of questionnaires. Applying this methodology, some tasks did not reach consensus.

In summary, in the first round of questionnaires to SMEs, applying the Delphi method and combining as consensus criteria the IQR and the α-cut calculation, it was observed that of the 550 questions asked, there was no consensus in the response to 290 (52.7%); 282 had an IQR greater than 1 and 8 did not meet the convexity requirement when the α-cut calculation was applied.

Thus, the 290 questions in which there was no consensus have passed the second round of application of the Delphi method.

2.3 Round 2

In the second round of Delphi questionnaires a personalized questionnaire was sent to each of the SMEs involved in the previous Round. As mentioned above, the questionnaire sent to the SMEs presented the 290 questions whose answers had not reached the consensus in Round 1. In order to try to converge to consensus, each SME received information on the weighted response frequency of each question, as well as the weighted average and his previous response.

Respondents were asked to review the data provided and to revise their previous response, changing it if they considered it necessary, using the same scale previously used, and to justify their choice, particularly if their answers did not converge to the most common opinion in the group. Eleven responses were obtained, corresponding to 91.6% of the total number of SMEs who answered the previous round. The answers were processed using the same methodology that was used in Round 1. A sample of the results obtained, to the questions asked in this Round are presented in Table 4.

From the analysis of the answers to the 290 questions, it was verified that 265 (91.4%) reached a consensus, having an IQR greater than 1. In this round no question was identified as non-consensual based on the α-cut calculation methodology.

Thus, 25 questions remained out of consensus being considered in the third round of the Delphi method.

Table 4. Sample of Round 2 results

	RECON		SAR		USAR		STRUCT		FIREF.		SANITAT.		ELECTRIC		MECHAN		LOG.		FOOD		MEDICAL	
Task	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA	IQR	WA
Identify incidents	-	-	1,0	5,6	1,0	5,6	0,0	5,0	1,0	4,6	1,0	4,6	1,0	4,2	1,0	4,6	0,0	3,4	0,0	3,4	1,0	4,5
Triage of affected and injured people	1,0	4,6	0,0	5,9	1,0	5,6	1,5	4,7	1,0	2,2	1,0	2,2	0,8	2,0	0,0	2,1	0,8	1,7	0,0	1,9	-	-
Provide first aid	2,0	4,9	-	-	-	-	0,0	4,7	0,0	3,0	0,0	3,0	0,0	2,8	0,8	2,9	0,8	2,8	0,8	2,8	-	-
Affected people registration	1,0	2,8	1,0	2,6	1,0	2,8	1,0	2,5	-	-	-	-	-	-	-	-	-	-	-	-	1,0	3,1
Identify location for [point of interest]	0,5	4,2	1,0	3,6	1,0	3,5	1,0	3,5	0,0	2,0	0,0	1,9	0,0	1,8	0,0	2,0	0,0	2,2	0,8	2,4	0,5	4,0
Transport material to install [point of Interest]	0,0	2,1	0,5	2,9	0,5	3,0	1,0	3,4	0,5	3,2	0,5	2,9	1,0	2,7	1,0	2,9	1,0	4,2	1,0	3,2	1,0	2,5
Instalation of [point of interest]	0,5	2,3	1,0	3,4	1,0	3,5	0,5	4,1	1,0	2,9	1,0	2,8	1,0	2,7	1,0	2,9	1,8	3,8	1,0	3,4	0,0	2,9
Transport severe injured person	1,0	3,0	1,0	5,4	0,5	4,7	1,0	4,3	1,0	2,4	0,5	2,3	0,0	1,9	0,0	2,1	0,0	2,2	0,0	2,2	-	-
Stabilize severe injured person	1,5	4,1	1,0	5,5	0,0	5,1	0,5	4,7	0,5	2,0	0,5	2,2	0,0	1,9	0,0	2,1	0,8	2,0	0,8	2,0	-	-
Transport slght injured person	1,0	3,0	0,5	4,7	1,0	4,5	1,0	4,4	0,5	2,2	0,0	2,2	0,0	1,9	0,0	2,1	0,8	2,3	0,0	2,1	1,0	4,3

2.4 Round 3

The third Round was devoted to assessing the 25 questions on which no consensus had been reached in the 550 questions asked. Round 3 was held in a format different from the previous ones, taking the form of a simultaneous interview with two SMEs. The main objective of the session was to try to arrive at a final assessment of importance for issues where there had been no consensus in previous rounds. A secondary objective was to validate all the answers previously obtained, in order to point out the existence of some possible error.

The SMEs who participated in this Round were two senior officers from the Integrated Training and Assessment Center, the Portuguese Navy body responsible for coordinating and developing training activities (e.g., the DISTEX response to a natural disaster) as well as for establishing the doctrine regarding standards of readiness observed by the operational forces.

During the interview the questions were presented, the results obtained in Round 2, as well as the justifications that were given by the SMEs to base their classification in the previous round. After analyzing these information elements, the SMEs reach a consensus on the final value of importance to be assigned to each task.

With the conclusion of Round 3, the application of the Delphi method was terminated, and consensus was reached on all issues. Table 5 summarizes the results of the 3 rounds, using a color scale to indicate the degree of importance of the teams to carry out the tasks. The green color indicates that a team is adequate to accomplish a task, while the red color indicates that a team is not suitable to accomplish a task.

Table 5. Sample of consensus values on the importance of teams to perform tasks, based on 6-points Likert Scale (1-Not Important to 6-Extremely Important)

Task	RECON	SAR	USAR	STRUCT	FIREF.	SANITAT.	ELECTRIC	MECHAN.	LOG.	FOOD	MEDICAL
Identify incidents	5,7	5,6	5,6	5,0	4,6	4,6	4,2	4,6	3,4	3,4	4,5
Triage of affected and injured people	4,6	5,9	5,6	4,7	2,2	2,2	2,0	2,1	1,7	1,9	5,6
Provide first aid	4,9	5,4	5,3	4,7	3,0	3,0	2,8	2,9	2,8	2,8	5,8
Affected people registration	2,8	2,6	2,8	2,5	1,5	1,5	1,5	1,5	2,1	1,3	3,1
Identify location for [point of interest]	4,2	3,6	3,5	3,5	2,0	1,9	1,8	2,0	2,2	2,4	4,0
Transport material to install [point of Interest]	2,1	2,9	3,0	3,4	3,2	2,9	2,7	2,9	4,2	3,2	2,5
Instalation of [point of interest]	2,3	3,4	3,5	4,1	2,9	2,8	2,7	2,9	3,8	3,4	2,9
Transport severe injured person	3,0	5,4	4,7	4,3	2,4	2,3	1,9	2,1	2,2	2,2	5,6
Stabilize severe injured person	4,1	5,5	5,1	4,7	2,0	2,2	1,9	2,1	2,0	2,0	6,0
Transport slight injured person	2,5	4,7	4,5	4,4	2,2	2,2	1,9	2,1	2,3	2,1	4,3

3 Conclusions

THEMIS' Knowledge Base contains knowledge intrinsic to the response resource types (namely, organization, and skills offered by response teams), intrinsic to the disaster incident types (namely, tasks required to handle the incidents, and skills required to perform the tasks) and relating resources to incident response tasks (response time, skills matching).

Eliciting the knowledge for implementing the THEMIS posed quite demanding challenges, particularly regarding the acquisition of tacit knowledge used by decision-makers in their reasoning processes.

The knowledge acquisition considering the specific problem addressed in this paper involved a group of subject matter experts (SMEs). The Delphi method was identified as a particularly suited approach to deal with the problem at hand, since its goal is to consolidate the views of SMEs of a particular domain, based on an iteratively process designed to converge to a commonly agreed opinion on a specific issue.

The goal of our study was to elicit knowledge on the level of fitness that disaster managers recognize between different types of response units (e.g., search and rescue, technical, medical) and specific disaster response tasks (e.g., first-aid, stabilization of structures, rescue of entrapped victims, firefighting, restoring of basic services, feeding and sheltering), with the purpose of making explicit the knowledge required for the inference process of the THEMIS intelligent system.

Considering that all SMEs had previous experience in disaster management but in different technical areas of disaster response (i.e., different degrees of expertise regarding multiple cross-domain areas), and to help reaching converge, the DELPHI method was applied considering a ponderation of opinions weighted according to the level of experience and confidence on the answers (as stated by the SMEs) and assessing the convexity of the resulting set.

This paper, detailing the preliminary work introduced in [20], describes a variant of the Delphi method used to establish the importance that SMEs assign to the selection of specific response teams to perform typical tasks required in DRO, discussing the approach adopted in each step of the iterative process, the results obtained, as well as the followed validation process. The outcome of the knowledge management process associated with the development of intelligent systems is determinant to ensure the effectiveness, efficiency and satisfaction of a system, which are key factors of its Usability.

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

References

1. Simões-Marques, M., Mendonça, P., Figueiredo, D., Nunes, I.L.: Disaster management support system prototype design evolution based on UX testing. In: Nunes, I. (ed.) Advances in Human Factors and Systems Interaction, AHFE 2019. Advances in Intelligent Systems and Computing. Springer, Cham (to be published)
2. Simões-Marques, M., Figueira, J.R.: How can AI help reduce the burden of disaster management decision-making? In: Nunes, I. (ed.) Advances in Human Factors and Systems Interaction, AHFE 2018. Advances in Intelligent Systems and Computing, vol. 781, pp. 122–133. Springer, Cham (2019)
3. Nunes, I.L., Lucas, R., Simões-Marques, M., Correia, N.: An augmented reality application to support deployed emergency teams. In: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (eds.) Proceedings of the 20th Congress of the International Ergonomics Association, IEA 2018. Advances in Intelligent Systems and Computing, vol. 822, pp 195–204. Springer, Cham (2019)
4. Simões-Marques, M., Correia, A., Teodoro, M.F., Nunes, I.L.: Empirical studies in user experience of an emergency management system. In: Nunes, I. (ed.) Advances in Human Factors and Systems Interaction, AHFE 2017. Advances in Intelligent Systems and Computing, vol. 592, pp. 97–108. Springer, Cham (2018)
5. Correia, A., Severino, I., Nunes, I.L., Simões-Marques, M.: Knowledge management in the development of an intelligent system to support emergency response. In: Nunes, I. (ed.) Advances in Human Factors and System Interactions, AHFE 2017. Advances in Intelligent Systems and Computing, vol. 592, pp. 109–120. Springer, Cham (2018)
6. Sackman, H.: Delphi Assessment: Expert Opinion, Forecasting, and Group Process. RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/reports/R1283.html

7. Azevedo, S.G., Govindan, K., Carvalho, H., Cruz-Machado, V.: Ecosilient Index to assess the greenness and resilience of the upstream automotive supply chain. J. Clean. Prod. **56**, 131–146 (2013). Elsevier (1974)
8. Duvarci, Y., Selvi, Ö., Gunaydin, H.M., Gür, G.: Impacts of transportation projects on urban trends in İzmir. Teknik Dergi **19**(1), 4293–4318 (2008)
9. Fraga, M.: A economia circular na indústria portuguesa de pasta, papel e cartão. M.Sc. dissertation. FCT, Universidade Nova de Lisboa (2017)
10. De Vet, E., Brug, J., De Nooijer, J., Dijkstra, A., De Vries, N.K.: Determinants of forward stage transitions: a Delphi study. Health Educ. Res. **20**(2), 195–205 (2005)
11. Powell, C.: The Delphi technique: myths and realities. Methodol. Issues Nurs. Res. **41**(4), 376–382 (2003)
12. Gunaydin, H.M.: Impact of information technologies on project management functions. Ph.D. dissertation. Chicago University, USA (1999)
13. Adler, M., Ziglio, E.: Gazing into the Oracle. Jessica Kingsley Publishers, Bristol (1996)
14. Wissema, J.G.: Trends in technology forecasting. R & D Manag. **12**(1), 27–36 (1982)
15. Nunes, I.L., Calhamonas, G., Simões-Marques, M., Teodoro, M.F.: Contributing to a decision support system that manages teams in emergency case (submitted)
16. Miller, G.A.: The magical number seven, plus or minus two. Some limits on our capacity for processing information. Psychol. Rev. **101**(2), 343–352 (1956)
17. Baddeley, A.: The magical number seven: still magic after all these years? Psychol. Rev. **101** (2), 353–356 (1994)
18. Mead, D., Moseley, L.: The use of Delphi as a research approach. Nurse Res. **8**(4), 4–23 (2001)
19. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
20. Nunes, I.L., Calhamonas, G., Marques, M., Teodoro, M.F.: Building a decision support system to handle teams in disaster situations - a preliminary approach. In: Madureira, A., Abraham, A., Gandhi, N., Varela, M. (eds.) Hybrid Intelligent Systems. HIS 2018. Advances in Intelligent Systems and Computing, vol. 923, pp. 551–559. Springer, Cham (2020)

A Study of An Airworthiness Requirement Influence of Human Factors on Aviation Incidents

Yinbo Zhang(✉), Zhefeng Jin, Ruijie Fan, Hua Meng, and Shasha Lu

COMAC Shanghai Aircraft Design and Research Institute,
No. 5188 Jinke Road, Pudong New Area, Shanghai, China
zhangyinbo@comac.cc

Abstract. The intent of this paper is two-fold: (1) find a method to reduce human errors in the pre-flight phase, and (2) reduce the accident rate during take-off. This paper recognises task analysis as one of the most effective methods based on the character of this paper and the methods in AMC25.1302. This paper will adopt the task analysis method to analyze the specific work task. Through the analysis of the FMC (Flight Management Computer) operation task based on the CS25.1302, it can be known that FMC is a kind of common equipment which can achieve the intended function. However, in the process of the FMC operation, there are still some defects, such as the information tips, information check and information modification. In comparing the old to the new airworthiness requirements of FMC, the CS25.1302 is identified as easily discovering such problems. Before the CS25.1302, such problems were not effectively dealt with.

Keywords: CS25.1302 · Tail strike · FMC · Human errors · Incident

1 Introduction

With the development of airplane structure strength and system reliability, the air accident rate has decreased significantly, especially for some serious accidents. Despite serious accidents decrease, however, aviation accidents/incidents are still a fairly continuous occurrence [1]. Studies began into why accidents/incidents cannot be totally eliminated. In fact, the cooperation of both the equipment and the pilots is required to accomplish the flight; to eliminate an accident, therefore, it is not enough to design perfectly functional equipment, but also to consider the human factor present. The whole process of airplane design proceeds according to the requirements of airworthiness. Equipment function can meet the airworthiness requirement. Equipment reliability is high enough, and reliability of some equipment is up to 10^{-9} [2]. However, the observers find that flight crew usually make some kind of inconceivable mistake during equipment operation, and that maybe the main reason for the accident/incident. Flight crew may make more mistakes especially in reaction to complicated conditions.

Aviation accidents are associated not only with the equipment function, but with the human. Some aviation incidents associated with human factor (such as tail strike)

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 399–409, 2020.
https://doi.org/10.1007/978-3-030-20040-4_36

happen every year. However, the airworthiness requirements which have been published seem unable to effectively solve the problems of the relationship between equipment and human. EASA added a new requirement CS25.1302 in 2009 [3]. This requirement mainly indicates the relationship between the pilot's task and the function of the system. It seems that these incidents can be reduced by this new requirement. Hence the paper primarily researches whether this airworthiness requirement when applied to aircraft design can help reduce the tail-strike incidents which regularly happen.

2 Incidents

A number of aviation incidents happen in the taking off and landing phase [1]. Many tail-strike incidents are caused by human errors every year [4], and it seems there is still not an effective way to solve them. One incident, on Nov 11, 1998, happened during the landing phase. The tail of an airplane struck the runway on landing, due to a miscalculation of the landing speed by the pilots [5]. After this incident, Boeing published a flight operation technical bulletin for operators in 2000. Entitled "Inadvertent Entry of ZFW (Zero Fuel Weight) into the GW (Gross Weight) Line of the FMC flight management computer" [6], Boeing expressed the importance of correct takeoff data. However, this bulletin has not been able to prevent the incident occurring. A tail-strike incident happened on an Airbus 330-343 in Frankfurt, Germany on 14th Jun. 2002 [7]. Another similar take-off incident happened on March 12, 2003 [8]. The tail of a Singapore airplane struck the runway and caused widespread damage during the take-off phase.

The same problems occur:

1. The flight crew made mistakes in the calculation of take-off speed. Human factor plays an important role in the incident of aircraft. Many incident/accident happened because of human error [4].
2. This incorrect data lead to tail strike incidents.

Through the analysis of many incidents, it is found that the tail-strike incident is common. One reason for a tail-strike incident is the wrong operation of FMC. In order to reduce the incidents in the take-off phase, the operation of FMC must be clear and understood by pilots in the pre-flight phase.

3 New Airworthiness of Human Factors (CS25.1302)

Traditional evaluation systems of human factors in flight deck mainly focus on system functions rather than function of equipment. It simply considers whether the function of equipment can meet the design requirements. It is believed that the design is good if all the functions can meet the requirements as expected. The process detail of a flight crew's operation usually cannot be depicted and evaluated; moreover, this evaluation system often analyses and evaluates with some single design points. It lacks comprehensive consideration of flight crew behaviour during the whole design process. To avoid the majority of flight incidents, the "human centre" design ideology should be used throughout flight deck design. This design is to further reduce human errors. EASA added a new section CS25.1302 in CS25 in 2009.

CS25.1302 provides requirements about equipment installed in flight deck to ensure flight safety. Based on a flight crew's tasks, it studies equipment controls and display systems in flight deck to meet the requirements of flight crew's tasks in each flight phase. It shows the "human centre", from the view of flight crew. Designers design flight decks with a definite goal based on flight crew safety operation, not only for achieving equipment function. Considering human factors in initial phase of design can reduce design defects in future. For example, more space room can be designed for pilots and equipment arrangement for pilots can be more convenient, no matter whether the operation manner is fit for flight crew, especially in emergency conditions.

CS25.1302 focuses on a flight crew's tasks. Equipment function is analyzed based on a flight crew's tasks. Flight crews complete different tasks and use certain equipment during different flight phases. From this, the designer can identify the equipment used in flight and they also can decide the main equipment which should be considered during the design phase.

4 Method

4.1 Background of the Method

CS25.1302 flow chart contains four steps as follows (Fig. 1):

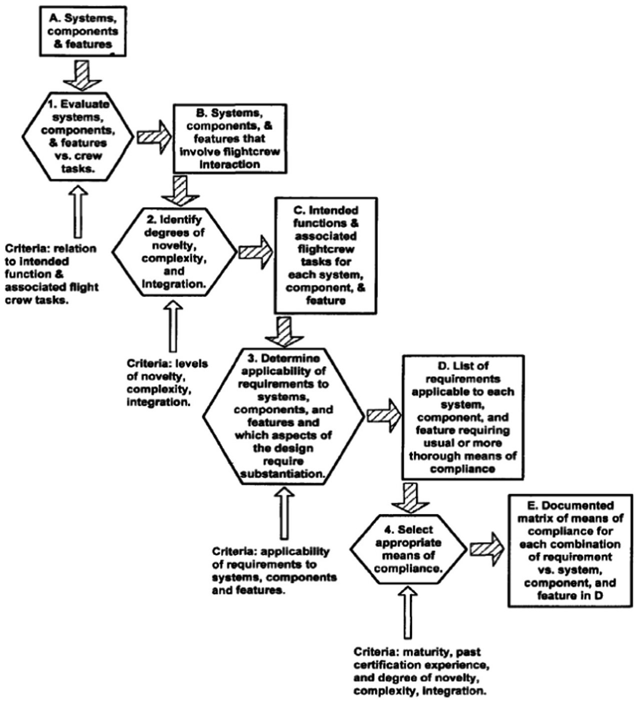


Fig. 1. Flow chart of method (AMC25.1302)

Step 1. Discuss all the systems, equipment and their characters in flight deck. Select the systems and equipment associated with flight crew's tasks.

Step 2. Confirm which systems, equipment and their characters are associated with flight crew's tasks. Estimate the level of innovation, integration and complexity through certain criteria associated.

Step 3. Study the functions of each system, equipment and their characters. Evaluate whether the function of these systems, equipment and characters can meet the airworthiness requirements, using the relative criterions.

Step 4. Select compliance validation manners of each system, equipment and character. Select an applicable manner to validate based on relative criterions.

4.2 The Scope of Research

The CS 25.1302 establishes some requirements to ensure that the design supports flight-crew member's ability to perform tasks associated with a system's intended function [3]. Two things should be indicated:

1. Flight crew's task
2. Equipment function

Those are the expected function of the equipment and the pilot's expected task.

For the first point, the expected aim of the pilot is also the pilot's operation task. These operation tasks can be found in an operation manual or a pilot handbook. These materials contain the basic information such as information description, the operation procedure of the pilot and the way of control.

The paper does not research the whole flight phase, but the pre-flight phase. Pilot's preparation in the cockpit should be considered. Pilot's work during the flight phase is not of concern.

For the second point, the design function of the equipment contains the data input and data output. The data input includes weight, zero fuel weight and centre of gravity. The data output includes the take-off speed, such as V1, VR and V2. Regarding the requirements of the CS25.1302, the designer should be concerned with the flight crew indications and controls for that equipment, and individual features or functions of that equipment.

Select the equipment which will be used in this research. After the definition of the operation task, it should be clear that this paper only studies the operation of FMC before take-off but not all the preparation of all the equipment in the cockpit.

When the initial definition of the work content and the selection of the equipment is complete, the conclusion will be that the working content of this paper is the operation task of FMC in the cockpit during the pre-flight phase.

4.3 FMC Operation Analysis with CS25.1302

Use the CS25.1302 to analyze the operation task. The CS25.1302 has many items, but not all the items can be used for analyzing FMC or tail-strike incident.

1. Analyze the FMC operation task, according to each item in the CS25.1302.
2. Make the study emphasize on the task which does not comply with the CS25.1302, after the analysis of FMC operation task.

5 Analysis and Discussion

Based on the methodology mentioned in the last Sect. 4.3, this chapter studies each item one by one. The basic analysis content obtained will provide the discussion. The methodology of this paper is task analysis method. The main focus is the analysis of the FMC operation task by the CS25.1302.

Two things should be declared before analysis:

Firstly, this paper studies the task analysis of FMC operation process during the pre-flight phase on Boeing 737NG.

Secondly, this paper mainly emphasizes the methodology research; the airplane type is a case study which can be changed.

According to Sect. 4.2, this paper mainly studies the FMC operation task in the flight deck during the pre-flight phase. The novelty, complexity and integration of the FMC equipment will be analysed first.

5.1 The FMC Equipment Analysis

Novelty Analysis: FMC is common equipment nowadays. It is installed in almost all commercial airplanes. Pilots will check and reset the content of FMC during each pre-flight phase, thus most pilots are quite familiar with the function of FMC. The location of FMC is usually the same; installed in both sides of the front part of central pedestal, which is one of the most convenient places for pilot operation.

Integration: FMC is integrated physically and functionally into the flight deck and may interact with other flight deck systems [3]. As the integration of FMC is very high, many types of equipment in the flight deck are associated with FMC. For example, the take-off speed provided by FMC will be displayed in the PFD. The navigation information will be displayed in the navigation display.

Complexity: in general, the complexity is closely related to integration. The more functions of integration, the more complex the operation. For pilots, the complexity of the FMC is high. FMC is integrated with many functions and requires a lot of data to be inputted, one page at a time, through the page control management. Some of the data requires the pilot to pre-calculate prior to input. The amount of pages and data requiring input makes the pilot vulnerable to mistakes.

In general, FMC is a type of equipment which limited novelty but significant complexity and integration.

5.2 The Flight Crew's Task and the Function of FMC

The Flight Crew's Task. At first, the tasks of normal pre-flight take-off procedures of Boeing 737 NG operation manual [9] should be analyzed. As this paper is concerned primarily about the tasks when the airplane is on the ground, specifically the take-off data in FMC in the pre-flight preparation phase, the analysis does not include the non-normal condition. If human error or problems with equipment are encountered during the operation process, the flight crew have enough time to deal with the problems. This will not result in an aviation incident.

Operation of FMC contains three steps: Data preparation for FMC, FMC operation and the crosscheck procedures.

The Function of FMC. FMC is a type of equipment with high integration and complexity. FMC integrates many flight control functions. Pilot should input the certain information in each page in the pre-flight phase.

One function of FMC relates to take-off references. Pilots need to enter and edit take-off plans, such as gross weight, zero fuel weight and slap setting, either manually or automatically. There are three pages in the FMC control part regarding take-off. They are IDENT Page, PERF INIT Page and TAKEOFF REF Page.

Pilot should finish the data and information input step by step following the information of the display on the FMC. Pilots should enter the data input page, and then choose the line which the data stands, press the button and input the data. The data will be displayed on the line. Each piece of data has its own pattern. Pilots should input the data into the right location as required.

When the pilots finish the data input, the FMC will calculate the take-off performance data based on its own calculation system and display the data of take-off speed on the screen. After the data is confirmed by the flight crew, the take-off basic data will be displayed on the PFD (Primary Flight Display).

5.3 The FMC Operation Task with CS25.1302 Analysis

The FMC operation task will be analyzed with the CS25.1302 in the subsequent paragraphs. The key of the task is to operate the FMC and input some data, according to the suggestive information of FMC. This aim has been emphasized in the first paragraph of the CS25.1302 – to ensure the design can support the flight crew's operation associated with the system's expected function.

"This paragraph applies to installed equipment intended for flight-crew members' use in the operation of the aeroplane from their normally seated positions on the flight deck" [3]. The first sentence: "installed equipment intended for flight-crew members' use in the operation of the aeroplane", requires that the function of FMC must meet the requirements of the pilot's operation. It is obvious that the function of FMC can meet the requirements. The FMC can obtain the take-off data and transfer them to the PFD after the pilot inputs the basic data. The second sentence: "from their normally seated positions on the flight deck" requires that the pilot can operate the FMC when he is sitting normally (usually with belt fastened). FMC can meet these requirements.

"This installed equipment must be shown, individually and in combination with other such equipment, to be designed so that qualified flight-crew members trained in its use can safely perform their tasks associated with its intended function by meeting the following requirements:" [3]. It should ensure the function of FMC is intact and the FMC can finish the take-off speed display function with PFD, so the pilot's work can proceed successfully. FMC can meet this requirement based on its function.

"(a) Flight deck controls must be installed to allow accomplishment of these tasks and information necessary to accomplish these tasks must be provided" [3]. This item regulates the minimum level of FMC function. FMC should provide enough function and information to assist the pilot to finish the operation of FMC. The FMC will provide general information to assist the pilot to finish the operation of FMC.

"(b) (1) Flight deck controls and information intended for flight crew use must: Be presented in a clear and unambiguous form, at resolution and precision appropriate to the task" [3]. FMC does not provide the relative helpful information or some suggestive information to instruct the pilot to finish the operation when the pilot operates each item of FMC. For example, if the pilot inputs the zero fuel weight data into the location of gross weight by mistake, FMC will not provide the information to tell the pilot otherwise and that the pilot should check again. If the pilot does not discover the mistake, the take-off speed calculated will be less than usual. This will cause a tail-strike incident. Through the analysis of the CS25.1302 (b) (1), it is discovered that the weight unit of FMC is not clear enough with kg and lb.

"(b) (2) Flight deck controls and information intended for flight crew use must: Be accessible and usable by the flight crew in a manner consistent with the urgency, frequency, and duration of their tasks, and" [3]. As for FMC, the location is good as the pilot can find and operate it quickly and easily.

"(b) (3) Flight deck controls and information intended for flight crew use must: Enable flight crew awareness, if awareness is required for safe operation, of the effects on the airplane or systems resulting from flight crew actions" [3]. This item requires the FMC to provide more information to the pilot, so the pilot can understand the consequence of each data item input. when the pilot operates the FMC, it should provide the relative information to make the pilot understand his behaviour and the task which the equipment will implement. It should give the alarm when it is necessary. If the take-off data calculated by the first officer is less than normal, it will influence the airplane on taking off after it is inputted into the FMC. If there is no suggestive information to be provided to the pilot, the pilot may easily ignore this problem and it will influence the airplane on taking off. Through the analysis of the CS25.1302 (b) (3), it is discovered that FMC does not have the suggestive information; the pilot may input ZFW data into GW location. it is also discovered that FMC does not have the prediction function. The pilot does not know the influence of his operation on the airplane take-off after he finishes the FMC operation task.

"(c) (1) Operationally-relevant behavior of the installed equipment must be predictable and unambiguous" [3]. For the operation of FMC, the meaning of predictable has two aspects: (1) to make the pilot know what he is doing and why; and (2) to make the pilot know the consequence of his action and the operation to the airplane. At present, FMC may not provide such kinds of information to the pilot, to let the pilot know the consequence after he inputs the data. Through the analysis of the CS25.1302 (c) (1), it is discovered that FMC does not provide the suggestive information to instruct the pilot to modify the data when the pilot modifies the data in FMC.

"(c) (2) Operationally-relevant behaviour of the installed equipment must be designed to enable the flight crew to intervene in a manner appropriate to the task" [3]. This item ensures that the pilots have the right to make final decisions. The pilots can modify the data which they have previously inputted, so that they can manage mistakes better. When the pilots operate the FMC, they may input wrong data. They should have the ability to modify and correct the wrong data to meet the working requirements. After the first officer calculates the take-off speed, the captain should change the take-off speed on the TAKEOFF REF page. The aim of the change is to make the pilot manage the airplane better and reduce potential incidences.

"(d) To the extent practicable, installed equipment must enable the flight crew to manage errors resulting from the kinds of flight crew interactions with the equipment that can be reasonably expected in service, assuming the flight crew is acting in good faith. This sub-paragraph (d) does not apply to skill-related errors associated with manual control of the airplane" [3]. The item (d) is different from the other items above, which reduce the errors in the FMC operation task through the requirement of design. However, the item (d) is to give the requirement after the pilot has made mistakes. The take-off data which the pilot will input into the FMC may be wrong. The captain and the first officer will prepare a lot of work before the operation of FMC, such as the information transfer, data calculation. If the first officer writes the zero fuel weight on the textbox of the gross weight on the bug card by mistake, he will use the wrong data to calculate the take-off speed. During the check process of the pilot, the captain does not check all the information which the first officer uses. The captain checks the amount of the fuel on the bug card and in the FMC; he also checks the gross weight on the dispatch paper and in the FMC. Both of these actions will not discover the mistake. However, the gross weight on the bug card and the dispatch paper are not checked. This random check will easily result in mistakes [8]. Some errors may be easily ignored. The take-off speed may be wrong. Thus, the error management should be used to ensure the reliability and accuracy of the data.

5.4 Discussion

The Problem of Weight Unit—kg or Ib. The requirement in the CS25.1302 (b) (1) is to provide clear and unambiguous information to the pilot. Pilots input the gross weight or the zero fuel weight into FMC according to the dispatch list when they operate the FMC. However, the dispatch paper is handed to them by the ground crew, who may use a different weight unit compared with the weight unit in FMC. Moreover there is no weight unit behind the weight textbox in FMC, and this allows for the pilot to judge the weight unit by his experience, lb or kg - the two most widely used weight units worldwide.

According to the requirement in the CS25.1302 (b) (1), this situation should be avoided, so the pilot can clearly understand each unit of every textbox in FMC. When he inputs the data into the FMC, he can see the unit behind the data. The unit can remind the pilot to pay attention to the problem of unit, and then he will notice the unit on the dispatch paper. This can ensure the two kinds of data are in accordance. The weight unit can help the pilot make fewer errors, so that the take-off incidents can be reduced.

Through analysis of FMC by CS25.1302, it can be found that the weight unit loss may lead to errors in the FMC, which may lead to take-off problems. However, the other airworthiness standards cannot discover this problem efficiently.

According to the requirement of CS25.1302 (b) (1), if the weight unit can be displayed on the FMC page, the tail-strike incident can be avoided. The airplane can be filled with enough fuel and fly as normal. The standard of CS25.1302 (b) (1) is useful for anticipating human factors problems with FMC.

The Position of Weight Data Input—Put ZFW Data into GW Position. The CS25.1302 starts from the pilot's task and considers the pilot's FMC operation task. As the CS25.1302 (b) (1) requires, "At resolution and precision appropriate to the task". When the designer designs the FMC data input function, he should consider the pilot's possible mistakes. The FMC should provide some suggestive information to the pilot to tell him what to input at a certain place before the pilot inputs the weight information. The FMC instructs the pilot to finish the data input according to the suggestive information. No matter how rigorous the pilot's training is, human error cannot be eliminated thoroughly. The suggestive information of FMC can remind the pilot and prevent some of the mistakes. The best way to make the pilot understand whether he has inputted the corresponding data at the right place is to provide the suggestive information about operation. The FMC reminds the pilot of information input by providing suggestive information when the pilot operates the FMC to input the GW or ZWF. The FMC should ensure there is some suggestive information to be provided to the pilot, such as "Input the GW here, not ZFW". This may reduce incidents of the pilot wrongly placing information. After the data is inputted into the FMC, the mistake can be reduced further by the pilots' crosscheck.

As the CS25.1302 (b) (1) requires, "At resolution and precision appropriate to the task". If these aspects can be taken into consideration during the design of FMC, that the FMC can inform the pilot when he inputs the weight, the incidents in 1998, 2003 and 2008 may not happen. The standard of CS25.1302 (b) (1) is useful for anticipating human factors problems with FMC.

Predict the Influence of Data Input on Takeoff. The CS25.1302 (b) (3) requires that after the pilot operates the FMC, there must be some relative methods to assure the pilot can know how his operation will influence the airplane on take-off. However for the situation now, the pilot only knows that the data which he inputs into the FMC such as gross weight or zero fuel weight is used to calculate the take-off speed, but does not have clear understanding about the take-off speed V1, VR, V2 and how the speed will affect the take-off. This obviously cannot meet the requirement in the CS25.1302 (b) (3) and may lead to an aviation incident.

The CS25.1302 (b) (3) requires that the pilot must know how his operation influences the airplane in future, not only predicts the consequence depending on the pilot's experience. The FMC should provide such information to the pilot that how this take-off data will influence the airplane during the take-off phase, so the pilot will understand whether his operation will influence the airplane's safety. The airworthiness requirements published before the CS25.1302 cannot discover this kind of problem unlike the CS25.1302 (b) (2) which can easily discover problems such as those identified.

The Influence of Take-Off Data Correction on Take-Off. FMC now has the function of calculating the take-off speed automatically. FMC will calculate the V1, VR, and V2 in the TAKEOFF page automatically, after the gross weight, flaps setting and the thrust parameters are set. This take-off data is calculated depending on the data inputted into the FMC without considering the actual situation during take-off phase, so the data calculated is just as reference. In general, the first officer will manually calculate the take-off speed depending on the situation during take-off phase; the factors considered include the length of the runway, the friction between the airplane and ground, the humidity of ground, the wind direction and so on. The actual take-off speed which will be modified into the TAKEOFF page of FMC is the speed the first officer calculates.

According to the requirement of the CS25.1302 (c) (1), “Predictable and unambiguous”, modification of flight data cannot meet the airworthiness requirement. The take-off speed is made clear because the specific take-off data is provided. However, in terms of being predictable, the situation at present cannot meet the airworthiness requirement. When the pilot modifies the data, he cannot understand the data well enough. The pilot cannot determine which take-off speed is right and why. Especially, when the data calculated by FMC is not far from the data calculated by the first officer, it is harder for the pilot to determine the accuracy of the data and predict the flight status. So this cannot meet the airworthiness requirement.

6 Conclusion

The conclusion can be made that the new airworthiness requirement is helpful for reducing aviation accidents through the research methods of FMC in the third chapter, the analysis of FMC in the fourth chapter and the discussion of FMC in the fifth chapter. These aviation accidents can be reduced by lessening the occurrence of a pilot’s wrong operation due to the new airworthiness requirement on FMC. The new airworthiness can identify a pilot that cannot execute the FMC operation task well. Through the analysis of the pilot’s task with the CS25.1302, it can be known that four problems exist with the pilot’s FMC operation at present.

1. The problem of weight unit—kg or lb
2. The position of data input—put ZFW data into GW position
3. The problem of predicting the influence of data input on takeoff
4. The influence of take-off data correction on take-off

Through the discussion of the problem existing at present, the FMC operation still has some problems and the old airworthiness requirements before cannot discovery these problems. The conclusion can be got that the new airworthiness requirement can be helpful in reducing the airplane take-off accident. Discovering the problems is just the first step to solving them. There is no analysis regarding solutions to the problems. Much work exists to improve the design of FMC and ensure the function serves the pilots, and this will hopefully be addressed by further study.

References

1. Boeing: Statistical summary of commercial jet airplane accidents: worldwide operations 1959–2009. Boeing published, Seattle (2009)
2. EASA: Certification specifications for large aeroplanes CS25 AMC, AMC25.1309, EASA, 2009/017/R, Cologne, Germany (2009)
3. EASA: Certification specifications for large aeroplanes CS25.1302, EASA, 2009/R, Cologne, Germany (2009)
4. Civil Aviation Authority (CAA): Global fatal accident review 1980–96, (CAP 681). Civil Aviation Authority, London (1998)
5. NTSB: Safety recommendation, refer to: A-05-03 through -07, NTSB report, Washington, DC, US (2005)
6. Boeing: Boeing 737NG operations manual normal procedures amplified procedures, flight deck preparation. Boeing Published, Seattle (2000)
7. TSBC: Tail strike on take-off and aircraft pitch-up on final approach. Aviation investigation report, June 2002, Gatineau, Quebec (2002)
8. TAIC: Boeing 747-412 9V-SMT, flight SQ286, tail strike during take-off. Aviation occurrence report, report no.: 03-003, Wellington, New Zealand (2003)
9. Bulfer: Boeing 737NG flight management computer user's guide, advanced guide to the 737 flight management computer. U10 software (2000)

Innovative Methods and Approaches for Systems Interaction Design

System-of-Systems Approach for Sustainable Production Planning and Controlling in Manufacturing Companies

Maximilian Zarte[1(✉)], Agnes Pechmann[2], and Isabel L. Nunes[3]

[1] Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, Campus Caparica, 2829-516 Caparica, Portugal
m.zarte@campus.fct.unl.pt

[2] Department of Mechanical Engineering, University of Applied Sciences Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany
agnes.pechmann@hs-emden-leer.de

[3] Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, UNIDEMI, Campus de Caparica, 2829-516 Caparica, Portugal
imn@fct.unl.pt

Abstract. The need for sustainable development has been widely recognized as important by governments, as well as by enterprises, and has become a popular topic in various disciplines, including ergonomics and industrial engineering. Sustainable development can only be achieved if impacts on current and future generations of humans are focused according to the three dimensions of sustainability (environment, social and economic). The paper presents a system-of-systems approach to sustainable production planning and controlling in manufacturing enterprises. The results of the paper can be used as a basis to support and improve the outcome of such process discussions with company managers, experts, and employees' representatives.

Keywords: Sustainability · Human factor · Production management

1 Introduction

Our world is more interconnected than ever before. This provides an opportunity for researchers to consider complex realities and support a better life for all considering questions of sustainability. Sustainable development, in particular, requires understanding and remediating the conditions under which our resources are gathered; the way our information is collected and stored; the working conditions in factories; the sophistication of the transport and energy networks that bring products and services to homes and workplaces; and our methods of disposal or re-use of products [1]. However, the need for sustainable development has been widely recognized as an important by governments, as well as by enterprises [2]. No standard definitions exist for the terms "sustainability," "sustainable development" and "sustainable manufacturing." The nature of sustainability can vary according to the system considered (e.g. household, industry, country, world), the affected stakeholder groups (e.g. employees, customers, suppliers, etc.), and/or the perspective and scientific background of the

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 413–421, 2020.
https://doi.org/10.1007/978-3-030-20040-4_37

researchers. Therefore, researchers always need to define what is meant by sustainability in relation to their research approaches. This leads to multiple theories about the understanding of sustainable products and processes, which are all true for a specific case.

There are several widely accepted definitions for sustainability. For example, the Brundtland Commission defined sustainability as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” [3]. Moreover, sustainability is bonded by the triple bottom line, which balances economic goals, environmental cleanness, and social responsibility [4]. Conventional manufacturing of products can be defined as the transformation of raw materials into products, usually to achieve a designated output (in terms of product quantity and quality) at minimum cost and in the desired time frame [5]. Figure 1 presents the main input and output flows for a production system.

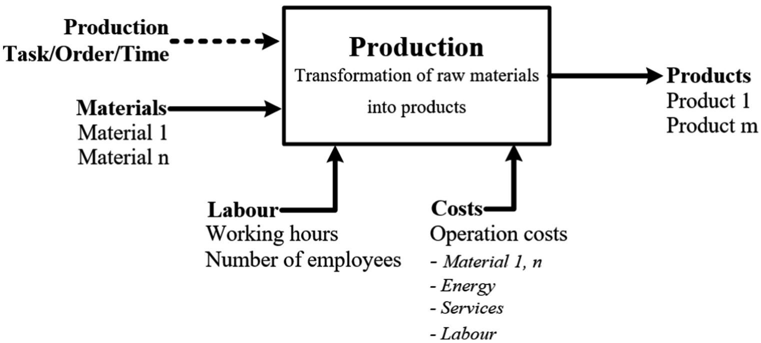


Fig. 1. Input/output model of production for conventional production planning and controlling. The model considers the input materials, labor, and costs, as well as the output products. The input and output can be normalized to production task, order and time (dash arrow).

Conventional production management according to conventional goals (such as minimum cost, production output) is already a complex task, because many aspects of a production system contribute to its planning processes. These elements not only include large production systems and changing product portfolios, but also unpredictable events (e.g., the breakdown of machines). Addressing sustainability objectives, such as limits for emissions and the health status of employees, in manufacturing operations scheduling in addition to classical production objectives makes production scheduling much more complex [6]. The change from conventional to more sustainable production planning and controlling (PPC) expands the boundaries of conventional production systems, and additional product life-cycle (PLC) phases (such as material acquisition, product use, product recycling, and product disposal) should be considered for sustainable PPC [7].

This extension of system boundaries for sustainable PPC was already investigated in a previous study [8]. The study presented an overview of stakeholders beyond the production system (costumers, suppliers, local and global communities) that affect the sustainability performance of the enterprise. Sustainability goals and indicators for these stakeholders, which can be used for sustainable PPC, were presented. Such an approach assumes that the stakeholders have an effect on the sustainable performance of enterprises. In reality, the sustainable behavior of enterprises and the needs and interests of the stakeholders are interdependent [9]. To describe these interdependencies, this paper presents a system-of-systems approach for sustainable PPC, describing the nested hierarchy of systems and related stakeholders and systems that interact with the enterprise.

The definition and general concept for a sustainable system-of-systems approach have been developed and presented in Thatcher and Yeow [9]. The authors introduced an approach that integrates the current hierarchal conceptualization of possible interventions with important concepts from the sustainability literature, including the triple bottom line approach and the notion of time frames [10]. Here, this approach has been adapted to develop a system-of-systems approach for sustainable PPC, one that takes the needs and interests of stakeholders into account in determining goals and indicators, PLC and time frames for sustainable manufacturing.

Section 2, which follows this introduction, provides a more detailed description of the sustainable system-of-systems approach. Section 3 outlines that approach for sustainable PPC. Section 4 demonstrates the application of the system-of-systems approach to sustainability goals for PPC. Section 5 presents conclusions and directions for future works, followed by references.

2 Theoretical Background of the Sustainable System-of-Systems Approach

According to Maier, “a system is generally understood to be an assemblage of components that produces behavior or function not available from any component individually. A system-of-systems is an emergent class of systems that are built from components which are large-scale systems in their own right” [11]. As noted, Thatcher and Yoew developed a sustainable version of this class of systems, which has three main components [10]:

1. A nested hierarchy of systems.
2. A time frame for a system is sustainable.
3. Goals and indicators for sustainability.

First, a nested hierarchy is used to describe a system-of-systems. Thatcher and Yeow prefer to use the nomenclature from Wilson to describe of the relationship between a target and related systems. A target system can interact with numerous sibling systems (e.g. the organizational systems of the organizations with which it cooperates or competes); parent systems (e.g. the economic climate, customer systems, and the organizational ecosystem reflecting the mix of organizations around it); and various child systems (e.g. the organization’s culture and communication systems) [12].

Second, Costanza and Patten noted that no system is infinitely sustainable in its current form, not even the universe. In fact, it is the very nature of systems that they are dynamic [13]. All systems have a natural lifetime (i.e. termination point), after which they will become brittle and unable to cope with changes. For example, at an individual level, many of the classic jobs have vanished with the change to office automation systems at the organizational level. A sustainable system is thus a system that reaches its full, expected lifetime. Any threat to the system that causes it not to reach its potential lifetime (e.g. disease if the system is a human; social unrest if the system is a community) is not sustainable [13].

Third, a system-of-systems approach cannot be considered sustainable unless it considers multiple goals [1]. These goals have been more extensively investigated in previous studies of sustainable PPC [8, 14, 15].

3 System-of-Systems Approach for Sustainable Production Planning and Controlling

Figure 2 presents the system-of-systems approach for sustainable PPC.

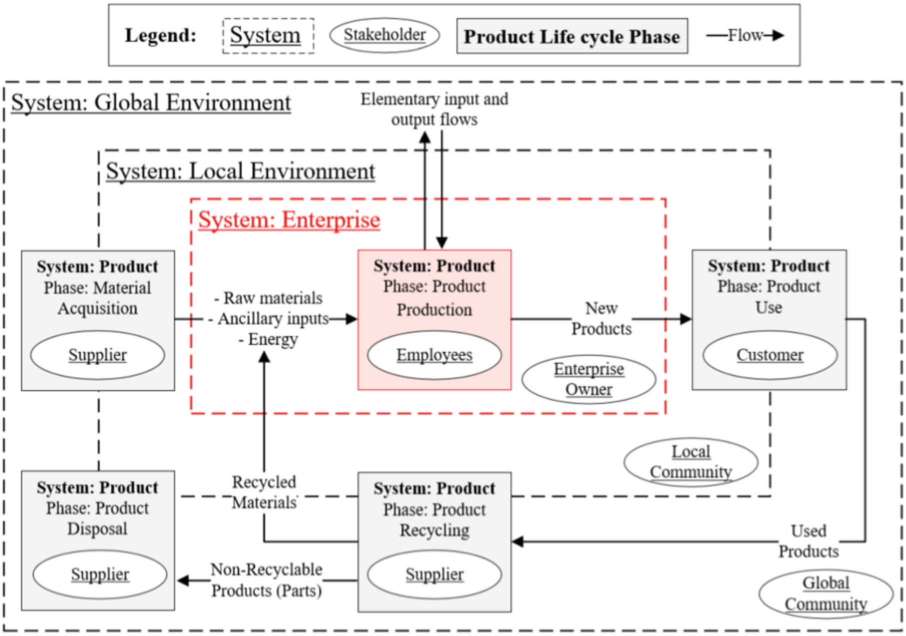


Fig. 2. Nested hierarchy of the systems for sustainable PPC, including stakeholders for sustainable manufacturing (owner, employee, customer, supplier, local community and global community) and product life-cycle phases (material acquisition, production, use, recycling, and disposal).

3.1 Component 1: Nested Hierarchy

The nested hierarchy consists of four systems:

– Global environment
– Local environment
– Enterprise
– Product

The target system is the enterprise system (see red mark in Fig. 2). The local and global environment are considered parent systems. The physical border between enterprise and local environment system can be clearly defined in terms of the land owned by the enterprise. The physical border between the local environment and the global environment is fluid, and must always be defined for each enterprise. Factors such as the size of the enterprise, its type, and the marketplaces for its products play essential roles in the range of social and effects it has local and global environments. The local and global communities can include human populations, cooperating organizations and competing enterprises.

The product system cannot clearly be defined as a sibling, child or parent system. A product passes through several life-cycle phases, which effect different systems. In the "material acquisition" PLC phase, external suppliers collect and generate materials, energy, and ancillary inputs, which are provided to the enterprise system. The collection and generation processes affect the sustainability of local and/or global environment systems; the extent of such effects depends on the geographical relationship of these systems to the enterprise systems. The "product production" PLC phase involves the enterprise system, and transfers materials into products. The production of products has direct effects on the sustainability of the enterprise system. Additional elementary flows that affect the stakeholders' local and global communities are also considered. The elementary flows include the use of elementary resources and releases to air, water and land associated with production. The elementary input flows considered include climate information (such as radiation, wind speed, moisture, and air pressure); water; air; and fossil fuels (natural gas, oil). The elementary output flows considered include emissions to water (e.g. sewage), ground (e.g. solid waste), and air (e.g. greenhouse gases [GHG]). In the "product use" PLC phase, costumers use the products provided by the enterprise system. The use of the products has effects on the sustainability performance of the local and global environmental systems. The "product recycling" PLC phase provides recycled materials to the production, and is performed by external supplier or employees. Non-recyclable products are disposed of in the "product disposal" PLC phase, which has effects on local and global environmental systems.

3.2 Component 2: Time Frame

The sustainable system-of-systems model considers the natural span of time over which a system should be sustainable, which is dependent on its relative position in the parent-sibling-child nested hierarchy. For example, an organization is expected to have a longer natural lifespan than its component systems [1].

The temporal dimension of the systems considered for sustainable PPC is applied throughout the PLC phases. PLC phases related to the enterprise system have a longer time span in which to be sustainable than PLC phases related to the local and global environment systems, because it is assumed that the enterprise system can react faster to changing requirements for sustainability than the other systems. However, to determine the time dimension to be sustainable of products is a challenging task, and depends on several factors, such as the lifetime of a product, political restrictions, the needs and interests of costumers, and environmental and social impacts on other systems.

3.3 Component 3: Goals for Sustainability

The goals for sustainable manufacturing from a human perspective are defined according to the interests of the stakeholders. Table 1 gives a brief overview of such interests.

Table 1. Overview of stakeholders in sustainable manufacturing and their main interests.

Stakeholder	Description of the stakeholder	Description of the interest
Enterprise Owner	Individual with a financial investment in the business who is responsible for the strategic orientation of the company	Interested in the economic value creation of the company
Employee	Individual who provides their skills to a firm, usually in exchange for a monetary wage	Interested in safe and healthy workplaces, opportunities for training and qualification activities, and fair salaries
Customer	Can be viewed as an end-user of a product, service, or process	Interested in demanding products to satisfy their needs
Supplier	Provides goods or services to the company	Interested in economic value creation
Local community	A spatially related group of individuals using a shared resource base within which a company enterprise exists	Interested in a healthy and safe living environment
Global community	A community outside the boundaries of the local community (e.g. a state, national, or international entity)	Interested in a healthy and safe world

The stakeholders are related to specific PLC phases and systems (see Fig. 2). Goals must be in accordance with current restrictions for sustainable manufacturing, as well as the interests of the stakeholders. Interests and resulting goals for sustainable manufacturing have been addressed more extensively in previous studies (see [8, 14, 15]). These studies used the 15 sustainable development goals (SDG) defined by the United Nations [16]. The studies identified four national SDGs, which can be affected by considering sustainable aspects of PPC processes:

- Responsible consumption and production.
- Good health and well-being.
- No poverty.
- Climate action.

Depending on the needs and interests of the stakeholders, the identified goals have different meanings, which require different indicators to control the goals. For example, the sustainable goal "responsible consumption and production" is allocated to multiple stakeholder groups: owners, customers, suppliers, and the local community. Local communities are interested in protecting local resources and the sustainable indicators are focused on the local resources available for manufacturing companies. However, costumers are interested in products, which are produced using renewable and recycled materials, green energy, and environmentally compatible ancillary inputs. The sustainable indicator "use of resources" is selected for both stakeholder groups, but with a different focus on the type of resources controlled.

4 Application Example of the System of Systems Approach for Sustainable Production Planning and Controlling

In her recent book, Klein argues that climate change is the most pressing problem of our age [17]. This thesis is acknowledged by several researchers (e.g. [1]). Therefore, this application example is focused on the sustainability goal "responsible consumption and production," and considers the effects of the emissions of GHG and the use of renewable energy on the considered systems. Figure 3 presents the system boundaries as the nested hierarchy of the case study, which is structured according to the identified goal. Some positive and negative effects among the systems and related stakeholders are marked with colored arrows to clarify how the system-of-systems approach can be used to describe sustainability goal for PPC.

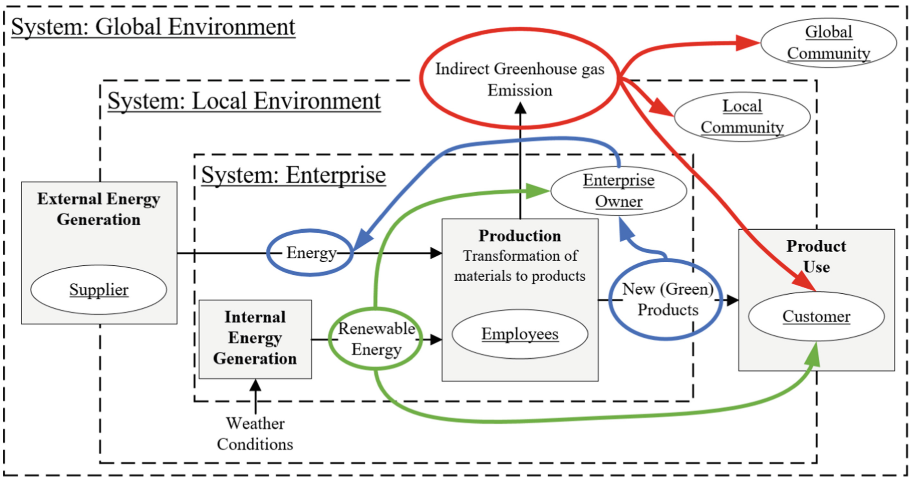


Fig. 3. System boundaries as nested hierarchy of the systems for sustainable PPC, defined in relation to the identified goal. Effects among the systems and life-cycle phases are marked red (negative effects) and green (positive effects).

In this example, it is assumed that an external energy supplier and internal energy generation processes supply the production facility with energy. The external supplier uses the fossil fuel gas for the energy generation and causes indirect GHG emissions for the enterprise system. The internal energy generation processes produce renewable energy with the aid of solar and wind energy, which depends on local weather conditions. The renewable energy does not generate any GHG emissions.

The GHG emissions have several negative effects on the stakeholders (red arrows in Fig. 3). The GHG emission affects the climate, which influences the local and global communities. Moreover, the emission of a high amount of GHG affects the buying decisions of customers. The internal energy generation processes, which cause no GHG emissions and involve fewer production costs, and have positive effects on the stakeholders (green arrows Fig. 3). However, the use of self-generated renewable energy, the production costs are decreased and the profit for the enterprise owner increases. Moreover, the use of renewable energy positively affects the buying decisions of customers. If more consumers buy green products, the enterprise owner will try to decrease the ratio of energy produced by external or internal processes, which will increase the green image of the enterprise system (blue arrows in Fig. 3).

5 Conclusion and Future Work

The paper presents a system-of-systems approach for sustainable PPC, one based on the methodology developed by Thatcher and Yoew [10], which considers human factors and ergonomics. The approach for sustainable PPC detailed above expands the system boundaries of conventional PPC for enterprise systems and considers additional PLC phases (material acquisition, product production, product use, product recycling and product disposal); stakeholders for sustainable manufacturing (enterprise owner, employees, costumers, suppliers, local communities and global communities); parent systems (local environment, global environment); and child systems (product systems). One of the key findings is that the product systems can be allocated to different systems depending on their PLC phase. The lifetime over which a product system should be sustainable is dependent on its relative position within the parent-sibling-child nested hierarchy and the sustainable impacts to the related systems. However, the results can be used to identify and describe the sustainability effects of enterprise systems among parent systems (local and global community systems), child systems (product systems) and related stakeholders. Knowledge of these effects can be used to introduce new sustainability goals and indicators in PPC processes, as well as to facilitate strategic and tactical enterprise management, and particularly the needs and interests of affected stakeholders in sustainable manufacturing. Moreover, the results of the analysis can be used as a basis to support and improve the outcome of the PPC process discussions with company managers, experts and employees' representatives.

The system-of-systems approach offers several opportunities for future work. The approach can be applied to a real use case to describe the sustainable performance of enterprise system in a nested hierarchy. The results of such a case study can then be used to identify applicable goals and indicators for sustainable decision-making in manufacturing enterprises.

References

1. Thatcher, A., et al.: This changes everything: macroergonomics and the future of sustaianbility. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **60**(1), 871–875 (2016)
2. Radjiyev, A., Qiu, H., Xiong, S., Nam, K.: Ergonomics and sustainable development in the past two decades (1992–2011): research trends and how ergonomics can contribute to sustainable development, (eng). Appl. Ergon. **46**(Pt A), 67–75 (2015)
3. WCED: Report of the World Commission on Environment and Development: Our Common Future (1987)
4. Elkington, J.: Cannibals with Forks: The Triple Bottom Line of 21st Century Business. Capstone, Oxford (1997)
5. Wieneke, F., Schmidt, J.: Produktionsmanagement: Produktionsplanung und Auftragsabwicklung am Beispiel einer virtuellen Firma; mit den Übungsversionen der Anwendungsprogramme ERP-Software "PMS-ERM", Simulationssoftware "DOSIMIS-3", 4th edn. Verl. Europa-Lehrmittel, Haan-Gruiten (2012)
6. Giret, A., Trentesaux, D., Prabhu, V.: Sustainability in manufacturing operations scheduling: a state of the art review. J. Manuf. Syst. **37**, 126–140 (2015)
7. Xin, Y., Ojanen, V., Huiskonen, J.: Knowledge management in product-service systems – a product lifecycle perspective. Proc. CIRP **73**, 203–209 (2018)
8. Zarte, M., Pechmann, A., Nunes, I.L.: Indicators and goals for sustainable production planning and controlling from an ergonomic perspective. In: Nunes, I.L. (ed.) Advances in Intelligent Systems and Computing, volume 781, Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2018 International Conference on Human Factors and Systems Interaction, 21–25 July 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA, pp. 363–373. Springer, Cham (2019)
9. Thatcher, A.: Longevity in a sustainable human factors and ergonomics system-of-systems. In: 22ª Semana De La Salud Ocupacional (2016)
10. Thatcher, A., Yeow, P.H.P.: A sustainable system of systems approach: a new HFE paradigm (eng). Ergonomics **59**(2), 167–178 (2016)
11. Maier, M.W.: Architecting principles for systems-of-systems. Syst. Eng. **1**(4), 267–284 (1998)
12. Wilson, J.R.: Fundamentals of systems ergonomics/human factors, (eng). Appl. Ergon. **45**(1), 5–13 (2014)
13. Costanza, R., Patten, B.C.: Defining and predicting sustainability. Ecol. Econ. **15**(3), 193–196 (1995)
14. Zarte, M., Pechmann, A., Nunes, I.L.: Indicator framework for sustainable production planning and controlling. Int. J. Sustain. Eng. **45**, 1–10 (2019)
15. Zarte, M., Pechmann, A., Nunes, I.L.: Influencing factors in sustainable production planning and controlling from an ergonomic perspective. In: Arezes, P.M., et al. (eds.) Occupational Safety and Hygiene VI: Proceedings of the 6th International Symposium on Occupation Safety and Hygiene, SHO 2018, 26–27 March 2018, Guimarães, Portugal. Chapman and Hall/CRC, Milton (2018)
16. United Nations, Sustainable Development Goals: 17 Goals to Transform our World. http://www.un.org/sustainabledevelopment/sustainable-development-goals/. Accessed 8 Feb 2018
17. Klein, N.: This Changes Everything: Capitalism vs. the Climate. Penguin Books, London (1997)

Set-Top Box Automated Lip-Reading Controller Based on Convolutional Neural Network

Yuanyao Lu(✉) and Haiyang Jiang

School of Information Science and Technology,
North China University of Technology, Beijing, China
luyy@ncut.edu.cn

Abstract. Automated lip reading (ALR) is a new mode of human-computer interaction and can directly recognize the speech content from the lip motion image sequences of the speaker. Using ALR based on convolutional neural network to realize the control of the set-top box to power on/off, change the channel and adjust the volume is a challenging and useful work. The paper uses a coupled three-dimensional convolutional neural network (3D-CNN) architecture that calculates the similarity between the captured lip instruction characteristic image and the standard image after training. We identify the control commands, change the channel and adjust the volume to achieve new control mode for set-top boxes. The experimental accuracy of lip-reading recognition is obviously improved compared to other similar methods. We are making a meaningful exploration for the practical use of ALR and our results and application experience can be easily extended to other machines and smart home systems.

Keywords: Automated lip-reading · Set-top box · Convolutional neural network · Human-computer interaction

1 Introduction

Automated lip reading (ALR) is a technique that analyses the content of lips by capturing the movements of the person's mouth while speaking. ALR is a comprehensive technology that integrates machine vision, artificial intelligence and natural language processing. Early researches on speech recognition technology introduced visual features of pronunciation [1–8] and improved recognition accuracy by extracting comprehensive features of audio and visual information.

In recent years, with the rapid development of artificial intelligence and big data, deep learning has attracted increasing attention of researchers and has been widely used in ALR technology. Compared with traditional lip-reading methods including lip segmentation and feature extraction algorithms [9–12], deep learning often uses depth network model to extract and understand the features of lip motion image sequences. Maulana [13] achieved the extraction of lip features by using convolutional neural network (CNN) and recurrent neural network (RNN). Rahmani [14] used the traditional GMM-HMM and the existing DNN-HMM to realize the recognition task of continuous

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 422–431, 2020.
https://doi.org/10.1007/978-3-030-20040-4_38

lip pictures. These previous studies provide technical support and guarantee the better application of lip-reading recognition to other machines and smart home systems.

The traditional set-top box controller generally uses infrared remote-control system and set-top box changing system which uses Android device via Wi-Fi network. In this paper, we propose lip-reading system to achieve a new control of the set-top box. Considering lip-reading aims at understanding the contents of speaker from lip movement sequences, the correlated information of inter-frame in video also plays a significant role in the final recognition accuracy. The paper captures the visual information of the lip movement through the high-definition camera, and builds a limited instruction set automated lip-reading control system for set-top box on a GPU platform. We use a coupled three-dimensional convolutional neural network (3D-CNN) architecture that makes special adjustments to the frequently occurring dynamic lip visemes. It can capture features from space dimension and time dimension simultaneously. 3D-CNN can join motion information in multiple adjacent frames. Therefore, the automated lip-reading system based on 3D-CNN is used to control the set-top box instruction set in this paper. The advantage of the automated lip-reading recognition system constructed by 3D-CNN is that it does not require manual extraction of features. Feature extraction and classification recognition have been constituted to an end-to-end complete closed-loop system.

The main contributions of this paper are as follows:

1. We designed a set-top box automated lip-reading controller based on deep learning method, which can realize the basic functions for visual control of TV set-top box and further explore the application of lip-reading recognition technology in other machines and smart home systems.
2. This paper reduces the initialization value of the parametric rectified linear unit (PRelu) activation function. The experiment demonstrates that a better network model can be trained even when the CNN architecture is small.

2 System Scheme

The physical connection of the set-top box automated lip-reading controller is shown in Fig. 1. The camera acquires instruction videos which are uttered by users. Then, the set-top box automated lip-reading controller uses lip-reading recognition technology to recognize commands. Finally, the commands are sent to set-top box via USB cable and achieve control of the set-top box automated lip-reading controller.

2.1 The Data Preparation of 3D-CNN

There are very few public databases available for lip-reading recognition studies. Considering the practical needs of study, we have recorded a database with ten instructions, including up, down, left, right, enter, plus, minus, mute, channel and volume. The database collects video clips from thirty people and each people repeats thirty times for each instruction. Therefore, the amount of dataset is composed of 300 videos.

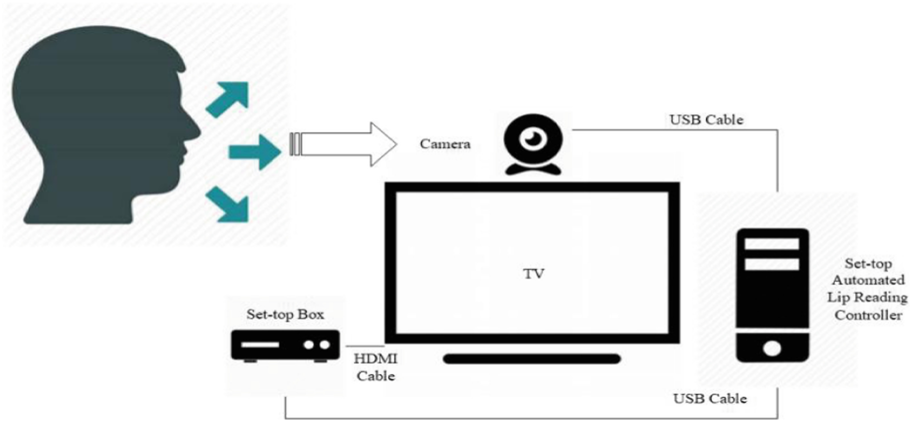


Fig. 1. The physical connection of the set-top box automated lip-reading controller.

Since there are only thirty samples for each category, the amount of database is not enough for deep learning. We use data enhancement technology to make the number of videos in each category reach to one hundred. Thus, after the video processing, the number of images for each category turns into 3200.

We analyze and contrast the previous researches which involve in face detection and lip feature extraction methods. We use Dlib library to achieve face tracking and mouth area extraction procedures to our own recorded videos. Then, the normalization of mouth area picture is used to make sure that the input data size of neural network is same. Therefore, the input data of 3D-CNN is a visual stream which formed by standardized consecutive image frames.

The visual stream of single sample is shown in Fig. 2. The size of visual stream is 32 × 32 × 32. The first 32 represents the number of frames extracted from sample video and 32 × 32 represents the size of mouth area image.

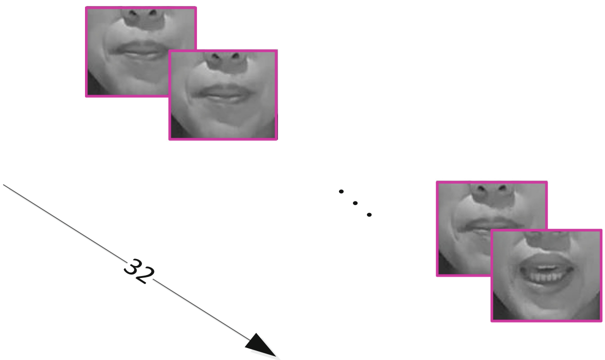


Fig. 2. The visual stream of 3D-CNN.

2.2 The Construction of 3D-CNN

The architecture and parameters of the proposed 3D-CNN are described in Table 1. It can be seen from the table that the format of the convolution kernel of the 3D-CNN is D × H × W × C, where D is the time dimension, H and W represent convolution kernel size of the height and width dimensions of the image respectively, C is the number of channels for input pictures. Like the 2D-CNN, the number of convolution kernel is equal to the number of input channel. The mouth area images used in this paper are all color images, so the number of input channel C is three.

Table 1. The architecture and parameters of 3D-CNN.

Layer	Input-size	Output-size	Kernel	Stride
Conv1	32 × 32 × 32 × 1	30 × 30 × 30 × 64	3 × 3 × 3 × 1 × 64	1 × 1 × 1 × 1 × 1
Pool1	30 × 30 × 30 × 64	14 × 14 × 10 × 64	1 × 3 × 3 × 3 × 1	1 × 2 × 2 × 3 × 1
Conv2	14 × 14 × 10 × 64	10 × 10 × 6 × 64	5 × 5 × 5 × 64 × 64	1 × 1 × 1 × 1 × 1
Pool2	10 × 10x6 × 64	4 × 4 × 2 × 64	1 × 3 × 3 × 3 × 1	1 × 2 × 2 × 2 × 1
Fc3	64 × 32	64 × 328	32 × 384	–
Fc4	64 × 384	64 × 192	384 × 192	–
Softmax	64 × 192	64 × 10	192 × 10	–

The specific operations of each layer for the proposed 3D-CNN network are introduced in the following. Firstly, the input data is convoluted with 64 convolution kernels of the first convolution layer and obtains output data which contains 64 feature maps. The output data passes through the first pool layer to extract the main features. Secondly, the output of the first layer is calculated to the same convolution and pooling operations exactly the same as the first layer. Then, we can get highly abstracted features after two fully connected layers. Finally, we use softmax function to get the probability value for each category and select the biggest probability value as the final recognition result.

For the proposed 3D-CNN architecture, convolution operations are performed on successive time frames towards input data stream. Except for the final full connected layer, all convolutional layers are followed by a modified form of the Relu activation function, named as parametric rectification linear unit (PRelu) [15]. It differs from the Relu activation function in that it multiplies a constant when the independent variable is less than zero and the equation is shown as follows:

$$f(y) = \begin{cases} y, & y \geq 0 \\ ay, & y < 0 \end{cases} \tag{1}$$

Compared with the Relu activation function, the PRelu activation function does not have the phenomenon of gradient disappearance and it can greatly reduce the over-fitting risk of the model. Thus, our network more stable. The comparison of the Relu activation function and PRelu activation function is shown in Fig. 3.

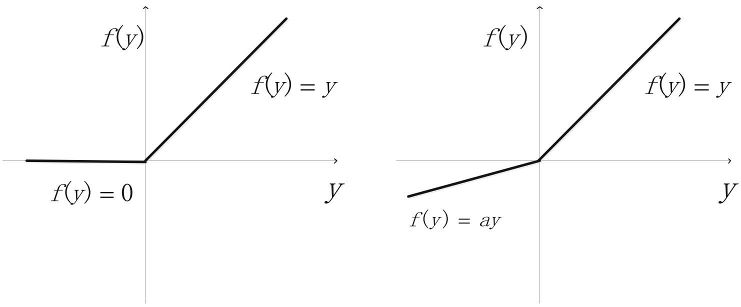


Fig. 3. Comparison of Relu and PRelu.

2.3 The Classifier of 3D-CNN

We use the cross-entropy function as loss function. The specific implementation of cross-entropy function is divided into following two steps:

Step 1: Softmax

For the classification tasks, the number of neurons of the output layer in neural network is the number of categories. The role of the softmax function is to normalize the output components so that the sum of the components is one. The softmax function converts the output component values of the output vector into the probability values which corresponding to categories. Figure 4 shows the specific operation flow of softmax function with three classifications.

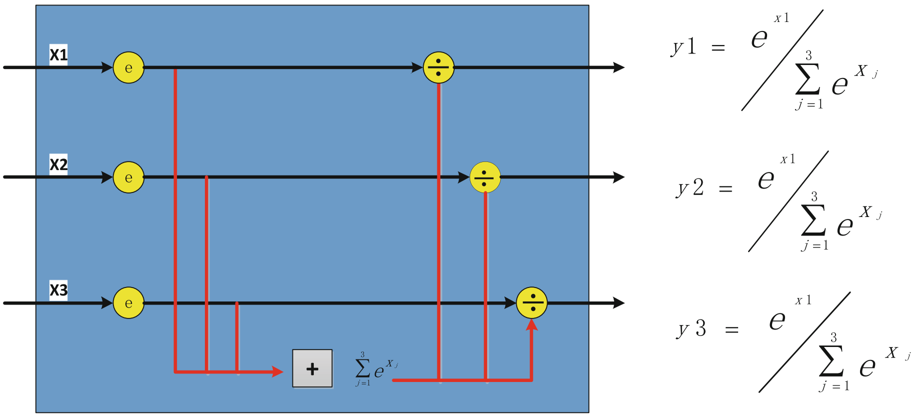


Fig. 4. The operation flow chart of softmax function.

The formula of the softmax function is as follows:

$$S_i = \frac{e^{V_i}}{\sum_j e^{V_j}} \tag{2}$$

Step 2: Calculation of cross-entropy

After calculating by the softmax function, the output layer of 3D-CNN starts to calculate the loss. In this paper, we use cross-entropy as loss function. Since the input of label format of tf.nn.sparse_softmax_cross_entropy_with_logits () is a one-dimensional vector, we need to convert it into a one-hot format. For example, if the component is 3, it means that the sample belongs to the third category and its corresponding one-hot format label is [0, 0, 0, 1, ... 0]. And if your label is already in one-hot format, you can use the tf.nn.softmax_cross_entropy_with_logits() function to perform softmax and calculate loss. After converting the label of the data to one-hot format, we calculate cross-entropy. The formula is as follows:

$$H_{y_i'}(y) = -\sum_i y_i' \log(y_i) \tag{3}$$

where y_i' is the label value, y_i is the corresponding component of the vector normalized by softmax. When the classification is more accurate, the corresponding component will be closer to one, and the loss value will be smaller. The convolution processes of the 3D-CNN are shown in Fig. 5.

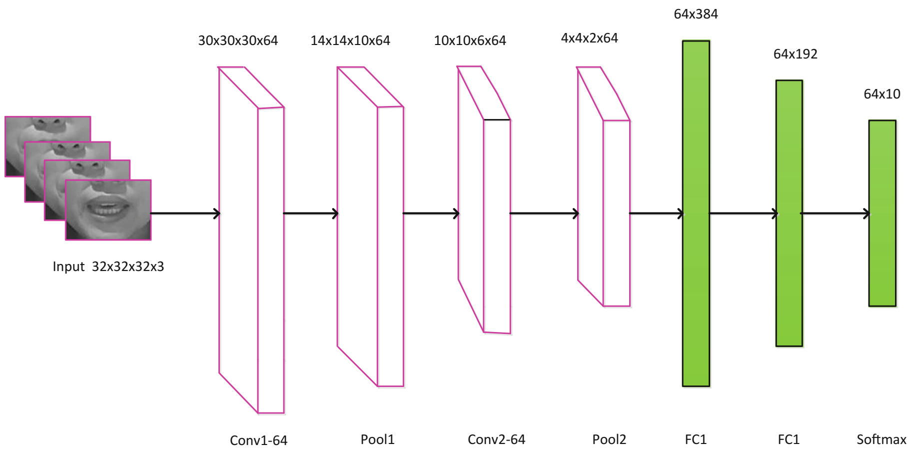


Fig. 5. The construction of 3D-CNN.

3 Experiment and Data Analysis

In the experiment, we use the high-definition camera linked to the set-top box to get the video data of the human lip movement sequences. We record 64 samples of each category which stand at different positions with 3–5 meters from the camera as our test data. We use the hybrid contour model as segmentation method to segment the lips, as shown in Fig. 6. Then we select 32 pictures with obvious changes in lip motion as input data of the network.

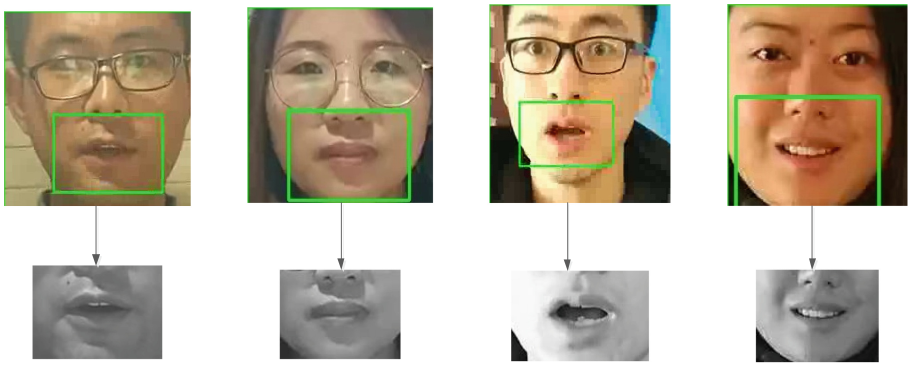


Fig. 6. Lip segmentation examples.

Kappa coefficient is a common method to evaluate the performance of multi-classification tasks. The paper uses kappa coefficient evaluation method to measure the accuracy of the model classification. The range of kappa coefficient is [−1,1]. In practical application, we usually use [0,1]. The kappa coefficient is calculated as follows:

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{4}$$

where p_0 represents total classification accuracy, p_e is expressed as:

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \ldots + a_n \times b_n}{n \times n} \tag{5}$$

where a_i represents the number of real samples and b_i represents the number of predicted samples. The larger value of k, the higher accuracy of our classification model. We use ten categories of data samples to test RNN which has performed well in lip-reading fields in recent years and the 3D-CNN in this paper. The classification results obtained by the test samples are shown in Tables 2 and 3 respectively.

Table 2. The classification results of 3D-CNN.

Prediction	Truth									
	Up	Down	Left	Right	Enter	Plus	Minus	Mute	Channel	Volume
Up	58	0	0	2	0	0	0	2	1	0
Down	2	57	1	1	1	2	1	0	1	1
Left	1	1	57	0	1	0	1	1	0	2
Right	1	1	1	60	0	0	1	0	0	0
Enter	0	1	1	0	60	0	0	0	2	0
Plus	0	2	0	0	1	59	0	0	1	0
Minus	1	0	0	1	0	0	57	0	1	0
Mute	0	1	2	0	1	1	2	58	0	0
Channel	0	0	2	0	0	1	0	3	58	0
Volume	1	1	0	0	0	1	2	0	0	61

Table 3. The classification results of RNN.

Prediction	Truth									
	Up	Down	Left	Right	Enter	Plus	Minus	Mute	Channel	Volume
Up	57	0	0	2	0	0	0	2	1	1
Down	3	56	2	2	1	2	1	0	0	1
Left	1	2	55	0	1	0	1	1	1	2
Right	0	2	1	58	0	0	1	1	0	0
Enter	1	1	2	0	60	0	0	0	2	0
Plus	0	1	1	1	1	60	0	0	1	0
Minus	2	0	0	0	0	0	57	0	1	0
Mute	0	0	2	1	1	1	2	57	1	0
Channel	0	1	2	0	0	1	0	3	57	0
Volume	1	1	0	0	0	0	1	0	0	60

Table 4 compares the recognition results of RNN and 3D-CNN with kappa coefficient. By calculating the kappa coefficient, it can be known that the performance of using the 3D-CNN model in lip-reading recognition is better than the RNN model.

Table 4. The comparison of kappa coefficient between RNN and 3D-CNN.

	RNN	3D-CNN
Kappa	84.60%	86.10%

By increasing the number of test samples, we calculate the accuracy of each instruction that controls the basic operation of the set-top box through the set-top box lip-reading controller. The comparison of recognition accuracy between RNN and 3D-CNN is shown in Fig. 7.

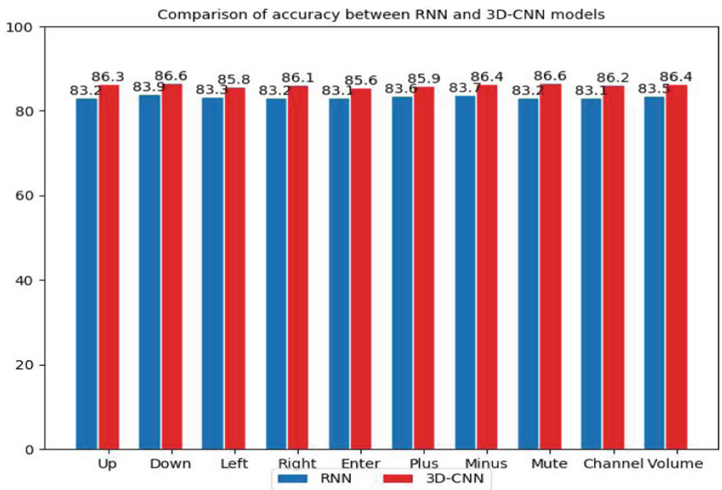


Fig. 7. Comparison of recognition accuracy between RNN and 3D-CNN.

From Fig. 7, it can be found that the accuracy of each instruction of the set-top box lip-reading controller using the 3D-CNN model is slightly higher than the RNN model. The average accuracy of RNN model is 83.38%, while the average accuracy of 3D-CNN model is up to 86.19%. It verifies the feasibility and effectiveness of the 3D-CNN architecture used in lip-reading recognition.

In 3D-CNN architecture, we use PRelu activation function to replace Relu activation function. The PRelu activation function does not have neuron death phenomenon. It is found that when the initialization value of the learnable parameter a is small, a small CNN architecture can also train network model with better performance.

4 Conclusions

This paper proposes 3D-CNN architecture to implement the lip-reading recognition system and uses the lip-reading recognition system to replace the infrared remote-control switching channel system of the set-top box. The experiments have shown that the lip-reading technology can effectively control channels of the set-top box. Our study lays the theoretical foundation for the research of lip-reading technology applied to other machines and smart home systems.

Acknowledgements. The research was supported by the National Natural Science Foundation of China (61571013), by the Beijing Natural Science Foundation of China (4143061), by the Science and Technology Development Program of Beijing Municipal Education Commission (KM201710009003) and by the Great Wall Scholar Reserved Talent Program of North China University of Technology (NCUT2017XN018013).

References

1. Yadav, I.C., Shahnawazuddin, S., Pradhan, G.: Addressing noise and pitch sensitivity of speech recognition system through variational mode decomposition based spectral smoothing. Digit. Sig. Process. **86**, 55–64 (2019)
2. Huang, J., Kingsbury, B.: Audio-visual deep learning for noise robust speech recognition. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7596–7599 (2013)
3. Shahnawazuddin, S., Maity, K., Pradhan, G.: Improving the performance of keyword spotting system for children's speech through prosody modification. Digit. Sig. Proc. **86**, 11–18 (2019)
4. Yi, J.Y., Tao, J.H., Wen, Z.Q., Bai, Y.: Language-adversarial transfer learning for low-resource speech recognition. IEEE-ACM Trans. Audio Speech Lang. Process. **27**, 621–630 (2019)
5. Alsharhan, E., Ramsay, A.: Improved Arabic speech recognition system through the automatic generation of fine-grained phonetic transcriptions. Inf. Process. Manag. **56**, 343–363 (2019)
6. Zeiler, S., Nicheli, R., Ma, N., Brown, G.J., Kolossa, D.: Robust audiovisual speech recognition using noise-adaptive linear discriminant analysis. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2797–2801 (2016)

7. Wang, J., Zhang, J., Honda, K., Wei, J., Dang, J.: Audio-visual speech recognition integrating 3D lip information obtained from the Kinect. Multimed. Syst. **22**(3), 315–323 (2016)
8. Petridis, S., Pantic, M.: Deep complementary bottleneck features for visual speech recognition. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2304–2308 (2016)
9. Lee, K.D., Lee, M.J., Lee, S. Y.: Extraction of frame-difference features based on PCA and ICA for lip-reading. In: IEEE International Joint Conference on Neural Networks (IJCNN 2005), Montreal, Canada, pp. 232–237 (2005)
10. Nath, R., Rahman, F.S., Nath, S.: Lip contour extraction scheme using morphological reconstruction based segmentation. In: 1st International Conference on Electrical Engineering and Information and Communication Technology (ICEEICT), Dhaka, Bangladesh (2014)
11. Li, Y., Hang, Y., Wang, Y. K., Chang, Y.C.: A lip localization method based on HSV transformation in smart phone environment. In: 12th IEEE International Conference on Signal Processing (ICSP), HangZhou, Peoples Republic of China, pp. 1285–1290 (2014)
12. Gritzman, A.D., Rubin, D.M., Pantanowitz, A.: Comparison of colour transforms used in lip segmentation algorithms. Sig. Image Video Process. **9**, 947–957 (2015)
13. Maulana, M.R.A.R., Fanany, M.I.: Sentence-level Indonesian lip reading with spatiotemporal CNN and gated RNN. In: International Conference on Advanced Computer Science and Information Systems. 9th International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 375–380. Inst Pertanian, Dept Ilmu Komputer, Jakarta, Indonesia (2017)
14. Rahmani, M.H., Almasganj, F.: Lip-reading via a DNN-HMM hybrid system using combination of the image-based and model-based features. In: 3rd International Conference on Pattern Analysis and Image Analysis (IPRIA), pp. 195–199. Faculty of Technology and Engineering, Shahrekord University, Shahrekord, Iran (2017)
15. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1026–1034. IEEE Press, Santiago, Chile (2015)

Learning from Employee Perceptions of Human-Work and Work-Organization in Digitized Production-Drilling Activity in Mines

Mohammed-Aminu Sanda[1,2](✉)

[1] University of Ghana Business School, P. O. Box LG 78, Legon, Accra, Ghana
masanda@ug.edu.gh
[2] Luleå University of Technology, 97187 Luleå, Sweden
mohami@ltu.se

Abstract. This paper discusses workers' perception of human-work and work-organization in a digitized production-drilling activity in mines and its associated learning that could lead to the creation of harmony between the technical and the social system in the design of the mining-work environment. Underlined by the systemic structural activity theory, data was collected in a digitized mine through interviews and observations of workers engaged in production drilling activities. Using the systemic analytic approach, the workers' perceptions of their digitized-world of work were analyzed functional. It was found that the workers perceive the mutuality of the exchanges, interactions, and understanding with the human-work and work organization design as lacking quality inputs from them, but which knowledge remains tacit and are not shared. It is concluded that by understanding workers' perception of the human-work and work-organization designs of their autonomized work environment, an optimized work-system, entailing sociotechnical systemic characteristics can be formulated.

Keywords: Employee perception · Autonomized work-system · Human work · Work organization · Digitized production drilling · Autonomized work environment · Deep mine

1 Introduction

Mining firms are known to face several operational constraints in their supply chains which has enormous consequences on their productive performances. Among these constraints are issues associated with the design and management of actor-technology interactions (both technically and socially) in the automated-work environment issues. Several mining firms have had to endure the consequences of such challenges for the past two decades, in terms of operational inefficiency, low productivity, and decreasing profitability. In bids to overcome such operational constraints, firms in the mining industry are shifting their focus toward the introduction of innovative and efficient practices underlined by digitizing the mining operations along their supply chain. According to [1], mining firms' abilities to enhance the digitization process will require that they develop new mental images of their firms introducing new technologies, to be complemented by efficient work-systems and

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 432–441, 2020.
https://doi.org/10.1007/978-3-030-20040-4_39

conditions that support high productivity [2]. In this vein, most mining firms are today, focusing the design of operational activities in their supply chains on highly automated systems [2]. Such systems, according to [2], are built around lean organizations with multi-skilled employees who have the capability to handle and manage a variety of tasks [2]. As such, technology and work-systems are expected to be integrated efficiently, meaning that the expected interactive cooperation between machines and workers will have to be built around a new system of autonomation [2].

Arguing from the perspectives of [3], a mining work environment entailing such system of autonomation will be a future representation of an intelligent work environment that is required to fully automate the operational systems mines [3]. It has been noted by [2, 3] and [4], that the modeling of such intelligent work environment, which incorporates a fully functional system of autonomation in the mines, will enable learning in work-system design, and collaboration between technology and human-work across organizational borders. Thus, going by the argument of [4], the supply chain of a modern mine with an intelligent work environment will be so technologically advanced that the number of its workers who are unskilled will decrease significantly and then, disappear altogether with time [4]. Hence, with the progressive movement from mechanization to automation in the mining industry, the development of new technologies and work methods need to be enhanced [5] and [6]. By developing work-systems that challenge the prevailing macho-culture in the mines, and which will create a safer work environment, the unequal gender balance that existed in most mines will be broken [4]. It is therefore important to identify innovative ways of enhancing work-organization in the intelligent mine work environment where gender equality, cooperation, skills development is key.

Sanda [6] has examined challenges that are associated with the design of a work-organization system in mining work environments that are highly autonomized. The underlined purpose of the study was to identify constraints that could be viewed as not only challenges but also as starting points for designing innovation systems for enhancing collaboration and social harmony between workers and their autotomized work environments. Specifically, [6] examined the effect of work environment autonomation on the performance of workers engaged in mining activities that are object-oriented and subject-oriented. According to [7], object-oriented activities are comprised of physical and mental operations for which the use of advanced technological tools by workers are required. The subject-oriented activities, according to [7], involved the dynamics of the workers' social interaction and collaboration with their autonomized work environments [6]. Thus, [6] emphasized the importance of capturing the unshared knowledge of workers in an autonomized work environment as an additive resource for developing an effective and efficient autonomized work-system model. This, according to [6], requires that workers with knowledge on the autonomized work environment be incentivized to share such knowledge. This sharing, according to [6], will help work-system designers gain an additive understanding of constraints in the social interaction and collaboration between workers and their autonomized work-systems. In this regard, [6] argued for the need to understand the emerging mental images of workers in the design of an autonomized work environment that are friendly, efficient and effective.

The purpose of this paper, therefore, is to understand workers' perception of human-work and work-organization in a digitized production-drilling activity in mines and its associated learning that could lead to the creation of harmony between the technical and the social system in the design of the mining-work environment. The objective is to challenge the old ways of doing things in the mining industry (especially in the deep-mines) by stimulating creativity and innovativeness in their organizational processes that will help them create attractive and stimulating work environment for their workers (i.e. to create new knowledge). In other words, it is to come out with knowledge that could enhance the creation of an attractive workplace in the mines which can engage and motivate even youngsters who are not particularly interested to work within the industry, as well as facilitate the development of an efficient work culture which will not endanger the health and welfare of mine workers operating in an autonomized work environments.

2 Literature Review

The mining industry has, for the past two decades, been undergoing vigorous technological revolutions with several organizations restructuring the functionality of their work environments towards autonomation. The restructuring is part of bids by firms in the industry to accelerate their production activities while being efficient and effective. In this respect, many organizations have autonomize their entire value chains [6]. The autonomation process, as noted by [6], appears to focus more on developing organizational structures that are autonomized. The autonomized system, according to [6], uses very efficient and effective technologies with high-performance capabilities and less reliant on humans [6]. It is argued by [6] that the autonomation of the work environment concerns the design of automated systems that are adapted to humans. As such, the process of autonomation must regard humans as resources in designing better-automated systems. As it is explained by [6], autonomation of an organization's work environment involves the harmonious integration of the systems' human and technological components [6]. In this vein, [2] described the automated system as work practices that enable learning for enhanced human-technology collaborative work. This systemic description is informed by the argumentations of [4, 7] and [8] to the effect that social relations in work environments are human-dominated and cannot be completely replaced by technologies. Though there are indications that the autonomation process in most firms, such as the recent conception of Industry 4.0 model, is progressing rapidly, [5] argues that such rapid development needs considerations and reflections that could result in the emergence of new learning from designers. Arguing from the perspectives of [2], such considerations and reflections are a manifestation of the challenges that need to be overcome by designers when developing specialized knowledge in autonomizing the work environment, as it relates to the work-organization and human systems. In this regard, workers unshared knowledge on the design characteristics of the autonomized work environment, when captured can be used to design efficient and effective human-technology interfaces [6]. In this vein, it is important for designers to understand workers' narratives about the functionalities of their autonomized work-system and its corresponding effect on their performances.

Arguing from the perspectives of [8] and [9], it is imperative for designers of autonomized work environments to incorporate workers shared mental images when designing work-systems with modern work-organization that supports harmonious human-technology collaboration. Thus, in the industrial drive toward autonomization, as emphasized by [8], development of autonomated production systems that are highly-integrated and smarter is a prerequisite. Such systems could serve as enablers for learning and collaboration across organizational borders [8], as well as facilitate organizations' capability to become lean. By implication, organizations with autonomated production systems can engage a few workers who are multi-skilled but have the capabilities to perform in multiple areas of the organizations' autonomized activities [4].

3 Methodology

The methodological approach was underlined by the notion professed by [2] that, for practice in a highly automated work environment to be understood, it is important to move beyond the environment's institutional similarities. By implication, [2] deemed it important to penetrate the situated and localized nature of practice in the environment's context. Underlined by this notion, the activities of workers engaged in production mining activity in a highly automated mining work environment was studied. The workers' practices are categorized as skill-based and problem-solving, and each practice has a level of automaticity with which it is performed [10]. The workers' skill-based tasks are performed in rapid automatic ways with minimum concentrations [10]. These tasks include production operations, which are guided by standardized methods of performance with logical decisions required on the sequence of possible actions.

3.1 Data Collection

Data were collected in a high-tech underground mine. The subjects of the study were experienced mine workers who were engaged in production mining activities in a highly autonomized work environment. Figure 1 below shows the external context of the work environment. The internal context is also shown in Fig. 2 below.

Fig. 1. External context of the production mining activity in the autonomized work environment with worker positioning highly digitized machine in preparation for work commencement.

Fig. 2. Internal context of the production mining activity in the autonomized work environment with worker operating highly digitized machine during work.

Firstly, the work of the workers was directly observed and video-recorded by the researcher. The direct and video observations were then used by the researcher to engage the workers in conversations (interviews) soliciting their subjective perceptions and unshared knowledge about their work-organization and work environments.

3.2 Data Analysis

The systemic structural analytical approach [11] was used as the basic paradigm for analyzing the positional actions of the miners who were studied while engaged in the production mining activity. In this analysis, the elements in the work-system that are deemed relevant and the dynamic interactions between actors in the work context are extracted. Such extraction makes it feasible to differentiate elements in the work environment, in terms of their functionalities, so that the systematic inter-relationship among the elements and their work-organization can be described [10]. Therefore, in the conduct of the systemic-structural analysis, a hierarchical description of activities engaged in by the miners are outlined. This is because, the organization of the miners' activities is hierarchical, and as such, need to be described and analyzed at different operational levels. This, therefore, necessitated the need to express the units of analysis in ways that will permit identification in the real work processes under study [10]. It also necessitated the need to combine the workers' activity descriptions with quantities of measures as a way of enhancing the prognosis of the efficiency of their performances [10]. In this regard, the description of the workers' activities is performed in a way that allowed inferences to be made on how the efficiency of work performances can be increased [10].

In analyzing the collected data, the functional analysis method of the systemic analytical approach was applied. Firstly, the cognitive aspect of complexity that depended on the specificity of information processing by the miners in when engaged in the production mining activity was evaluated. Secondly, the emotional-motivational aspects of complexity that reflected the energetic aspects of the production mining activity engaged in by the miners were also evaluated. In this respect, the workers' perceptions of their work environments design, informed by their mental capture of the images, concepts and non-verbal signs associated with their work, which were deemed to constitute their non-shared knowledge were analyzed. In other words, the potential strategies of activity performance associated with the workers' actions and operations, were appraised qualitatively. These included an appraisal of the workers' goals and past experiences. The workers' subjective perceptions on standards of a successful result and admissible deviation, decision-making on explorative performance actions, and assessment of activity's sense and difficulty were also appraised.

4 Results and Discussion

The following is a conversation between the observer (researcher) and an experienced worker (miner) engaged with the production drilling activity who was subject of observation and whose work was video-recorded;

Observer: What do make you feel relaxed when working, and that you can come to work the following day very fit and fully recovered?

Worker: The most important thing when you come to work every day is that the machine should have repairments. we should have a team to fill diesel, change oil, everything, and when you come to the machine to work, it is newly serviced. Also, we have no vehicle to drive around. In my opinion, if we have one, and if we are on launch break, the team can go around to fill oil and diesel, and if you have something you want to be fixed, they fix it for you. That is the most important thing.

Observer: If you want to visit the washroom, what do you do?

Worker: You must take the car. we have the washroom only at level 980 m where we eat. We must drive from here, level 1330 m up to level 980 m and back.

Observer: So, in your mind, do you think it will be better for management to provide more washrooms so that you do not have to be going up and down?

Worker: Yeah! More toilets, and more places to eat in the mines so that you do not need to transport yourself to and fro.

Observer: So, do you think it is possible in the next fifteen years to replace you with robots? That is to bring robots here.

Worker: No, not in this environment

Observer: Why?

Worker: I know it is impossible. You can see that I spent about six minutes trying to get the robotic arm complete the drilling of this marked spot. So, I do not think so. We need people here to work on the machine

The exchanges highlighted above between the worker and researcher (observer) were analyzed functionally. It is found that the miner's individual object-oriented activity entails operations that were characterized by physical and mental actions both of which had an influencing effect on the miners' perceptions of their work entrainment. This is in line with [7]'s argument to the effect that in mine work, the object of the worker's self-regulation of orienting activity is influenced by his/her perception of the work environment design, informed by his/her physical and mental activities. This is because, self-regulation is not a psychological notion, but a cybernetic one, and is an influence on a work-system that derives from the system itself in order to correct its behavior or activity [11]. It can therefore be derived from the exchanges above that the object of the workers in the autonomized work environment, in terms of their physical activity, is their engagement in work activity in which they use informed decisions and programs to explore and innovate additive practices that can enhance their work performances as well as their sensemaking capability when performing their work. Furthermore, the object of the workers in the autonomized work environment, in terms of their mental activity, is the concurrence of their work performance with task-difficulty assessment, and communication alertness, for which dynamic mental is a pre-requisite.

The following is also a conversation between the observer (researcher) and another experienced worker (miner) engagement with the production drilling activity who was also subject of observation and whose work was also video-recorded.

Observer: I know you want to help your organization make more money by work faster.

Worker: Yes, you are right! At the time, I try to work faster, but my supervisors will ask me to slow down. But I hope you will agree with me that the more the machine works, the more money the organization will make.

Observer: I observed that when the machine develops some faults while you are using it to work, you mostly stop to repair it yourself, instead of notifying and waiting for the maintenance crew. Why?

Worker: Your observation is right. There are very few mechanics and it is a one-minute journey between them (i.e. at levels 920 m and 816 m). Sometimes you must wait for two hours. Therefore, instead of waiting for two hours, you go ahead and fix the machines yourself. You have to do it. The managers say they do not have the money to employ many mechanics. We have only two guys here.

Observer: You think that the management does not have the money? Maybe, the management is not aware that you always do this kind of machine repairs as an addition to your normal work, rather than informing and waiting for the maintenance crew.

Worker: No! They know, but they do close their eyes.

Observer: Those who manufactured this high-tech machine you are using, as well as your work environment that is, the engineers and the designers. Do they come to talk to you the workers and ask you to tell them how you feel about interaction and usage with the machines and the work environment?

Worker: It is only when the company wants to buy a new machine from them that we see them here. Even with this, they do not show any interest in conversing with us about the performances of their machines that we are using. Maybe, because they are the designers, they feel that they know more than we do and are not be interested in our views about their machines

The exchanges highlighted above between the worker and researcher (observer) were functionally analyzed. It is found that the miners' have developed subjective perceptions of various objective activities in the organization's supply chain, which remained unshared. These include issues related to the efficiency of their work-systems and the operational functionality of their work environments. The workers use such tacit knowledge to negotiate constraints associated with work in their autonomized work environment. The findings indicate that in the course of their work performance in the autonomized work environment, the workers develop their own perceptions which translate into their non-shared procedural knowledge of the work environment. When such knowledge is shared with designers of the work-system, enormous insight be gained which will provide space for learning and innovation in the future design of a highly autonomized and work-friendly work environment. The workers perceive the mutuality of exchanges, interactions and understanding regarding the human-work and work organization systemic designs as lacking quality inputs from them. This quality input, if understood by their management, could have been incorporated to enhance the quality of the production-drilling task and associated work design. The workers also perceive the autonomized work-system as negatively affecting human social relations in the work environment in its structuring of transactions between human-technology roles as the building blocks of the digitized production-drilling activity. The base for this postulation is underscored by [7]'s observation that when workers gain experience from their work performances, the tendency for them to be more abstract in the cognitive conceptualization of their work environment grows. Such cognition, with its inherent complexity, influences workers perceptions and interpretations of their work environment design, leading to an openness towards their tacit knowledge development influenced by their experiences in the work environment. Arguing from the perspectives of [12], this cognitive complexity could be viewed as the most useful integrating model of psychosocial influences in work-organization system design. This is because, such cognitive complexity, as it is explained by [13], has a higher order structural personality dimension which persons to develop different conceptual systems for perceiving reality [13]. Thus, the design of an effective autonomized work environment can be enhanced by understanding and learning from workers perceptions on the environment. Arguing from the perspectives of [14], organizations seeking to autonomize their work-systems and making their work environment intelligent must encourage the practice of procedural knowledge sharing, especially by workers who interact with the work context. Such practice of knowledge sharing, when cultivated will enhance the practice of knowledge transfer, competence development, and learning at work for production workers in mining organizations. All these practices provide spaces for learning and are relevant human-work and work organization elements that can enhance organizations' digitization of their work environment. By implication, learning in work scenarios should be viewed not only in terms of training individuals and work-teams [14] but also as part of organizational change processes and the development of an autonomized workplace culture that can handle rapid changes such as, the introduction of highly digitized production technologies and techniques.

5 Conclusion

This study has shown that the perception of workers who are engaged in production mining activity in an autonomized work environment of their work environment is an important learning medium for designers of such a work environment. It is, therefore, concluded that organizations seeking to develop an autonomized work environment should put in place strategies and mechanisms to encourage workers to share their tacit knowledge on the functionalities of the human work-system as well as the work-organization systems. By sharing their knowledge, workers will bring their subjective and perceptual insights to bear, and which insight is key to identifying both the innovative and constraining characteristics of the human-work and work-organization systems. Thus, it is important for designers of autonomized work environments to capture and understand the relived experiences and perceptions of workers who interact daily with the work context. This understanding will enable designers to create intelligent work environments and design autonomized work-systems that are friendly, efficient and effective for the digitized production drilling activity. As emphasized by [14], those responsible for designing both human-work and work organization systems need to appreciate that their organizations are operating in a new industrial context that creates new prerequisites and new opportunities for development. Involving the individual's total competence, as a resource, in the organization's internal rationalization work [14] is a learning focus that must be considered as key in the changes required in the autonomized mining industrial environment. The implication is that the creation of space in organizations, especially in an autonomized work environment, for capturing workers perceptions and specialized perspectives in the human-work and organization systems design processes must be deemed important. This is could result in the development of an innovative autonomized work environment design that is efficient, and which can enhance a collaborative and harmonious sociotechnical co-existence between workers and autonomized systems for enhanced productivity.

References

1. Abrahamsson, L., Johansson, J.: From grounded skills to sky qualifications: a study of workers creating and recreating qualifications, identity and gender at an underground iron ore mine in Sweden. J. Ind. Relat. **48**(5), 657–676 (2006)
2. Sanda, M.A.: Automation of the work environment and the human-technology collaboration challenge: a critical reflection. Int. Rob. Auto. J. **3**(6), 00074 (2017)
3. Bassan, J., Srinivasan, V., Knights, P., Farrelly, C.T.: A day in the life of a mine worker in 2025. In: Proceedings of First International Future Mining Conference, Sydney, NSW, pp. 71–78 (2008)
4. Abrahamsson, L., Johansson, B., Johansson, J.: Future of metal mining: sixteen predictions. Int. J. Min. Miner. Eng. **1**(3), 304–312 (2009)
5. Johansson, J., Abrahamson, L., Kårebon, B.B., Fältholm, Y., Grane, C., Wykowska, A.: Work and organization in a digital industrial context. Mrev **28**(3), 281–297 (2017)

6. Sanda, M.A.: Dichotomy of historicity and subjective perception of complexity in individuals' activity goal formation and decision outcomes. In: Ayaz, H., Mazur, L. (eds.) Advances in Neuroergonomics and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol. 477, pp. 265–277. Springer (2018)
7. Sanda, M.A., Johansson, J., Johansson, B., Abrahamsson, L.: Using systemic structural activity approach in identifying strategies enhancing human performance in mining production drilling activity. Theor. Issues Ergon. Sci. **15**(3), 262–282 (2014)
8. Sanda, M.A.: Cognitive and emotional-motivational implications in the job design of digitized production drilling in deep mines. In: Hale, K.S., Stanney, K.M. (eds.) Advances in Neuroergonomics and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol. 488, pp. 211–222. Springer (2016)
9. Jarzabkowski, P.: Strategy as social practice: an activity theory perspective on continuity and change. J. Manage. Stud. **40**(1), 23–55 (2003)
10. Bedny, G.Z., Karwowski, W.: A Systemic-Structural Theory of Activity: Applications to Human Performance and Work Design. Taylor and Francis, Boca Raton (2007)
11. Bedny, G.Z., Karwowski, W.: Analysis of strategies employed during upper extremity positioning actions. Theor. Issues Ergon. Sci. **14**, 175–194 (2011)
12. Hendrick, H.W., Kleiner, B.M.: Macroergonomics: An Introduction to Work System Design. HFES, Santa Monica (2001)
13. Harvey, O.J., Hunt, D.E., Schroder, H.N.: Conceptual Systems and Personality Organization. Wiley, New York (1961)
14. Johansson, J., Abrahamsson, L.: The good work: a Swedish trade union vision in the shadow of lean production. Appl. Ergon. **40**, 775–780 (2009)

Vehicle Multi-sensory User Interaction Design Research Based on MINDS Methods

Qianqian Wu and Zhang Zhang(✉)

East China University of Science and Technology,
Shanghai 200237, People's Republic of China
873616093@qq.com, zhangzhang@ecust.edu.cn

Abstract. With the Development of Artificial Intelligence, there are different ways to interact with vehicles besides graphical user interface (GUI). Technology such as speech recognition, gesture recognition, facial expression recognition, and gaze tracking has gradually gained application in vehicle interaction. The advances in autonomous driving technology will also lead to the transformation of vehicle functions from transportation to multi-sensory space. Users can get a comprehensive experience based on sight, hearing, touch, taste and smell in future vehicles. Therefore, an exploration of the future vehicle interaction design combined with vehicle technology is needed. The objective of this paper is to deliver a vehicle multi-sensory user interaction design model based on the MINDS method [1]. This paper starts with the function analysis based on user requirement. On this basis, a connection between user and innovative services can be illustrated which enhances the user experience. Then an interaction sketch offered a detailed view of how users interact with vehicles and how technology impacts interaction. Finally, a new vehicle interaction model can be created by integrating the scenarios, technology, and users.

Keywords: Vehicle interactive design · Artificial intelligence · MINDS method

1 Background

In the past 30 years, the human-vehicle interactive mode has changed significantly. The in-car interactions have gradually transformed from physical buttons to graphical user interface, voice user interface, gesture interaction, and gaze interaction. A previous research revealed that humans got 83.0% information from sight, 11.0% from hearing, 3.5% from smell, 1.5% from touch, and 1.0% from taste [2]. This indicated that there are more ways for users to obtain information, process information, and express their in-car needs clearly. And it is found that in communication, verbal information only accounts for 7%, while the remaining are non-verbal information. The non- verbal information is composed of gestures (55%) and intonations (38%). It means that the vehicles can "understand" more about their users via facial identification, gesture identification, intonation identification, and individual style identification. This is closely related to artificial intelligence technology. AI impacts human-computer interaction, AI promotes the interactive mode from technology-centered to user- centered. And finally, a

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 442–447, 2020.
https://doi.org/10.1007/978-3-030-20040-4_40

sense-centered interactive mode developed using AI to enhance context awareness, conscious awareness, and emotion awareness [3]. In summary, users will get multi-sensory interactive experience through a variety of interactive ways.

The development of the Internet of Things also leads to the promotion of human-vehicle interactive mode. Internet of Things is now evolving rapidly, and the CAGR will reach 14.4% from 2018 to 2021 [4]. We can see a huge potential in the growth of Internet of Things, and one of the key development fields is the Internet of Vehicles. The Internet of Vehicles involved in vehicles, transportation, electronics, and other fields. The Internet of Vehicles connects users with various users-centered services applications. These service applications, mainly contain intelligent in-car information system, users can apply applications such as navigation, music, communication, information research. As technology develops, users will get more services. The key element of the Internet of Things is that it connects with all transportation resources. Thus, In- car interaction will cover multiple factors and multiple usage scenarios. Now the autonomous driving technology is updated quickly. And the autonomous driving level was classified into six levels (from level 0 to level 5) in SAE (J3016) [5] based on the level of automation. Once the technology and the law system get perfected, the amount of fully autonomous vehicles will increase quickly. Under this circumstance, the function of drivers will become weakened, and the in-car environment will finally turn into mobile space. Therefore, in-car user interactive mode and user experience will meet a great difference.

Based on the above consideration, this article is to take a deep exploration into future vehicle user interaction connecting with technology and usage scenarios. It is essential to consider the relationship of users, product (software and hardware), and the scenarios in a more systematic way, for the future vehicle interaction depends heavily on these elements. Thus, in that context, this article makes research on future vehicle interaction model based on the MINDS method.

2 The MINDS Method

The MINDS method was proposed by Grenha Teixeira et al. in 2017 [1, 2], which is a service design method that combing with interaction design method.

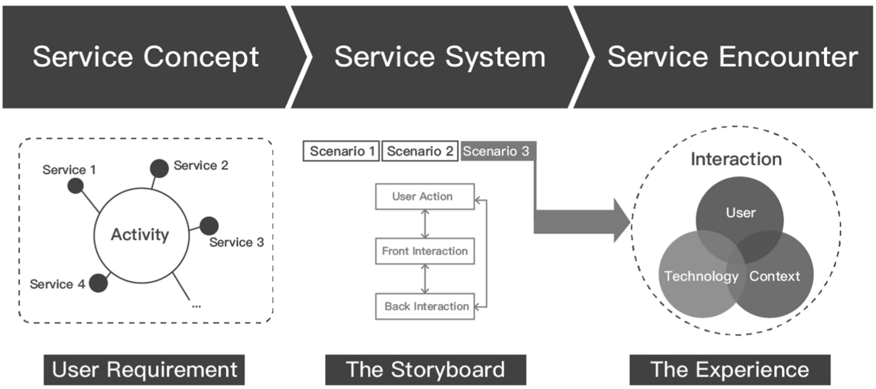


Fig. 1. Three key processes of the MINDS method

The MINDS method paid much attention to how technology affects service experience and user interactions. This method is to analyze the relation between users, technology and the context. There are three key processes in this method, as shown in Fig. 1. In step 1, designers need to collect the demand of users via questionnaire or user interview. And then typical scenarios should be drawn, while the relations between users, front interactions, and back interactions can be described in step 2. In the final step, designers can make interaction sketches for further development. In this paper, we apply step 2 to study the relation of users and technology.

3 Research on Future Vehicle Multi-sensory Interaction

Combining with related technology and key users researches, the trend of the automotive industry was summarized as "eascy" by PwC – electrified, autonomous, shared, connected and yearly updated [6]. The renewal of the technology and the industrial structure will lead to a revolution of vehicles. Then, the application scenarios and the usages of users are corresponding to be changed in the future.

This article is to discuss the future vehicle multi-sensory interaction based on the above trends. The objective of this document is to explore a future vehicle interaction model connected with users, technology, and the context.

3.1 User Requirements Research

The main purpose of this paper is to find and explore the user requirements while the vehicles are highly automated. We made brief interviews with six drivers to find out their perspective toward self-driving vehicles, and what action they might do facing this situation. After an overview about their attitudes, we also gathered more information from a report about public attitude toward autonomous vehicles in Berkeley [7] and a similar survey in the U.S., the U.K, and Australia [8].

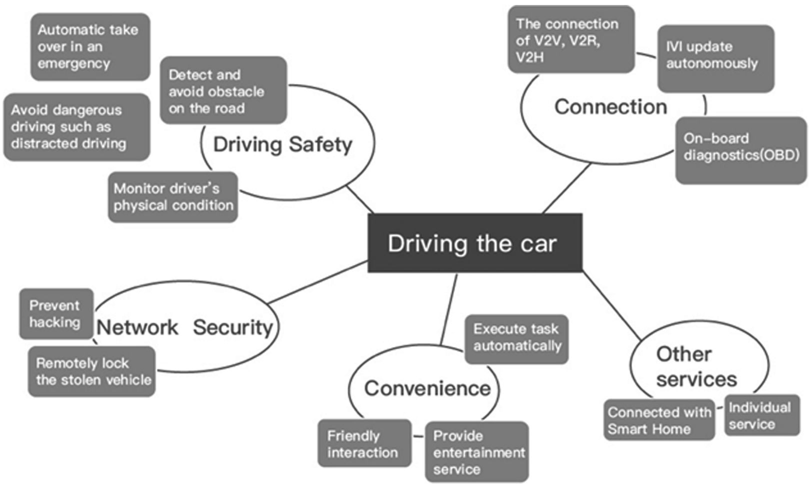


Fig. 2. Five main concerns in the future self-driving car

Based on these analyses and our brainstorming, the user preference was created. As shown in Fig. 2, the primary concern of users is driving safety. Almost all of them show a sense of distrust toward self- driving car. Is the car really safe? Will it get operation mistake? Can I take over my car at any time? They also take network security seriously, fearing about the privacy leaking. The rest elements they care about are the connection, convenience, and other services.

Then this article proposed a new interaction system. This new proposal pays much attention to build the feeling of trust with the users and provided optional interactive modes. The main characteristics are as follows: (1) the in-car system will give the notification initiative to the users about the vehicle operation, help to establish reliance between users and the system. (2) Various interactive modes can be identified to help the users convey their requirements easier, such as speech, eye movements, and gesture. (3) The proposal provides cross-domain scenarios through the Internet of Vehicles.

3.2 Vehicle Multi-sensory Interaction System

After the definition of the future vehicle interaction proposal, this article explored the relationship between users, vehicle, and technology based on the MINDS method (Fig. 3). We used Storyboard to draw below five scenarios, trying to integrate users, service (the vehicle and other associated locations), and the backend system (the technology and big data).

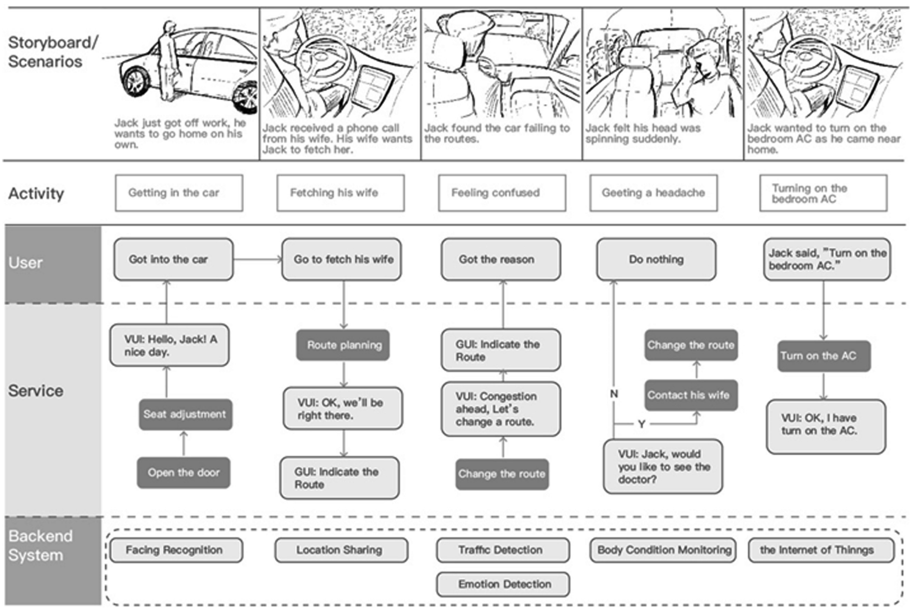


Fig. 3. Vehicle multi-sensory interaction system

3.3 Future Vehicle Interaction Model

Based on the above analysis and exploration, a future vehicle interaction model was put forward. As shown in Fig. 4, users are the core elements. The environment can impact the in-car interaction, so the users will get multi-sensory experience. When the users proposed their commands intuitively or unconsciously, the environment will be affected and generate different interactive modes.

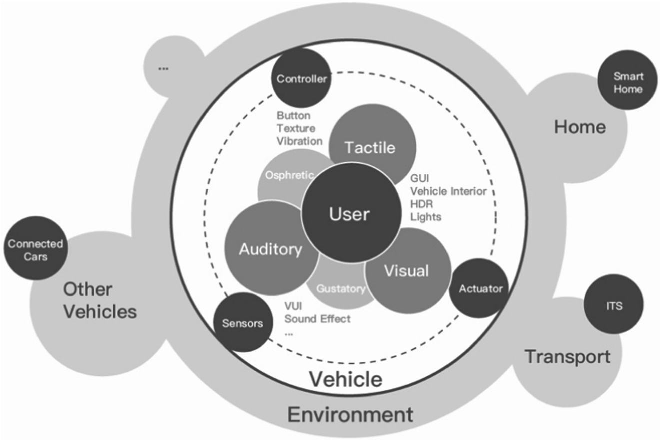


Fig. 4. Future vehicle interaction model

4 Conclusion and Prospect

This article made an in-depth discussion on the future vehicle multi-sensory interaction combining with users, technology, and the environment. There is no doubt the future vehicle interactive pattern will face great changes as technology develops, and different scenarios will become connected. The vehicle in the future may play a bigger role, not just be considered as transportation. While in vehicle interaction design, it is essential to building credibility between users and in-car intelligent systems. Only then can the users get enhanced experiences.

Acknowledgments. We thank everyone who provided valuable suggestions and feedback during the writing of this paper, especially Dr. Zhang Zhang. Thanks to the supported by Shanghai Summit Discipline in Design - DA18304.

References

1. Grenha Teixeira, J., Patrício, L., Huang, K.H., et al.: The MINDS method: integrating management and interaction design perspectives for service design. J. Serv. Res. **20**(3), 195–197 (2017)
2. Stolovitch, H.D., Keep, E.J.: Telling ain't training. American Society for Training and Development (2011)

3. Jinyan, T.: The interactive design in the era of big data. J. Packag. Eng. **36**(8), 1–5 (2015)
4. Xiaoshan, Y., Liang, Y.: The development of the Internet of Things industry in 2018 and the outlook. J. China Autom. Ident. Technol. **6**(8), 51–55 (2018)
5. SAE Standard J3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems, vol. 4, pp. 593–598 (2014)
6. Kuhnert, F., Stürmer, C., Koste, A.: Five Trends Transforming the Automotive Industry. PricewaterhouseCoopers, Frankfurt (2018)
7. Howard, D., Dai, D.: Public perceptions of self-driving cars: the case of Berkeley, California. In: Transportation Research Board 93rd Annual Meeting, vol. 14, no. 4502 (2014)
8. Schoettle, B., Sivak, M.: A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia (2014)

Human Computer Interaction in e-Health and Environmental Applications

Evaluation of the Usability of a Mobile Application for Public Air Quality Information

Jorge-Luis Pérez-Medina[1(✉)], Rasa Zalakeviciute[1,2], Yves Rybarczyk[1], and Mario González[1]

[1] Intelligent and Interactive Systems Lab (SI2 Lab), Universidad de Las Américas (UDLA), Quito, Ecuador
{jorge.perez.medina, rasa.zalakeviciute, yves.rybarczyk, mario.gonzalez.rodriguez}@udla.edu.ec
[2] Grupo de Investigacion en Biodiversidad Medio Ambiente y Salud (BIOMAS), Universidad de Las Américas (UDLA), Quito, Ecuador

Abstract. This contribution summarizes the results achieved from a summative usability study considering the efficiency, effectiveness, and subjective satisfaction of final users when using Air Quality, a new mobile software for Public Air Quality information. Fifty-one participants tested the user interfaces of a prototype of the application. We use the IBM Computer System Usability Questionnaire (CSUQ) in order to assess the user experience (UX). The Microsoft Desirability Toolkit is also applied as complementary tool. We present experimental results based on subjective appreciations and self-reported feed-back. It suggests a favorable trend towards the usability of the application. Nevertheless, we have found that the quality of the information is a deficiency in the mobile application. We analyze some recommendations to improve the mobile application.

Keywords: User interfaces · User experience · User study · Air pollution mobile application

1 Introduction

Rapid global human population growth, technological development, industrialization and urbanization are responsible for environmental quality deterioration, overuse of natural resources, and environmental pollution, among other anthropogenic induced issues. Constant emission of gaseous and particulate pollution from industrial, residential and traffic related activities has been increasingly changing the composition of the atmosphere, ranking air pollution as one of the most critical concerns to human and environmental health [1, 2].

A vast array of atmospheric contaminants are triggering cardiovascular and respiratory diseases, and causing incidents of cancer, which kills seven million people prematurely every year [3] While in the developed countries the air pollution is slowly decreasing due to regulations, it is not quite the same in the developing countries [4], potentially increasing the mortality rates due to this cause in the next few decades [5]. While it is quite challenging to reduce the unfortunate progression of harmful air

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 451–462, 2020.
https://doi.org/10.1007/978-3-030-20040-4_41

pollution, escaping the exposure could be a reasonable solution for sensitive population. Raising the air quality awareness is limited to the newspaper and television reports common only during extreme pollution episodes. While the increased air pollution episodes are extremely harmful to every exposed citizen, the long-term day-to-day exposure can also cause serious health problems [6]. Therefore, we believe, that a more efficient way to inform the population of the risks of exposure to poor urban air conditions is applications used on easily accessible mobile devices [7].

More than 54% of the world population resides in the cities. This number is estimated to reach 66% by 2050. Studies assert their predictions in the fast urbanization in developing countries [8]. These effects are represented in the shortage of natural resources, motorization, climate change and, especially, pollution in our environment. Actually, 92% of the global human population living in cities with more than 100,000 inhabitants are intensively exposed to the air pollution not meeting the guidelines of the World Health Organization (WHO), while this value increases to 98% in low-income countries [1, 9]. At this point, air pollution is one of the main causes of premature mortality, in addition to being the main environmental risk, causing 7 million premature deaths per year, and it is expected to double by 2050 [3].

Urban centers contain millions of people of all ages and health states. The population most suffering the health effects due to the poor air quality are children, elderly and risk groups with cardiovascular and above all respiratory diseases. However, certain daily activities could be avoided by sensitive population, if the awareness of air quality conditions would be more user friendly/accessible.

In recent years, mobile applications for Public Air Quality Information [7] have reached increasing levels of presence in the field of environmental quality. Assuring the acceptance and ease of use of these kinds of applications requires the implementation of usability studies during the development cycle of such a tool. The usability allows to measure if a specific product can be easily used by users who wish to reach certain objectives. The measurements consider aspects such as efficiency, effectiveness and satisfaction in a specific context of use. The evaluation measures may be both formative or summative [10]. Formative or summative usability evaluation process are incorporated in UCD practices [11]. Through formative usability evolution, designers detect early usability concerns. This formative evaluation is complemented with the summative evaluation or user experience (UX). Formative usability facilitates the refinement of a product during its design, just as it focuses on studying the quality of the user's experience. This method that consists of describing the user experience is widely adopted in early stages of a User-Centered Development process (UCD) [12]. In [7], a formative usability is realize considering heuristics and metrics to measure the potential problems of an Air Quality mobile software for Public Air Quality information.

The ISO 9241-11 [13] regulates the way in which usability should be applied within the framework of the use of interactive systems, products and/or services. This standard defines usability as a measure variable of efficiency, effectiveness and acceptance of a product by a user who needs to complete several tasks in specific use contexts [13].

Tullis et al. [14] recommends the System Usability Scale (SUS) [15] and the IBM Computer System Usability Questionnaire (CSUQ) [16] to evaluate the usability, as they are considered to be the most effective. The SUS offers a "quick and dirty" assessment of the usability of a system, while the CSUQ allows to study aspects such as, the utility, the information quality and the interface quality of an interactive application. The utility implies the ease of use, the learning, the speed in accomplishing the tasks and the friendliness of the user with respect to the interactive system. The information quality study the errors concerns, the intelligibility and the simplicity of the information. Finally, the interface quality study the affective attitude of a user when use the interactive system.

The SUS only provides one score to identify how usability is perceived by finals users. Regrettably, it has no other measure to measure the appreciations of the users. On the contrary, the IBM CSUQ facilitates obtaining the satisfaction of the user when using any interactive system. The IBM CSUQ have nineteen validated 7-point Likert scale questions organized in 4 categories. All questions have 3 levels for positive appraisals as well as negative ones. The reliability of this kind of scale is guaranteed by the alpha coefficient described in [17].

The Microsoft Desirability Toolkit is a method composed of 118 emotional words [18]. The words allow to obtain the reactions of the participants. These are oriented in visual aspects of design, functionalities, user experiences that users perceive when using a product or design. Designers use this method to obtain the emotional feedback and desirability from the participants after a user experience.

The Microsoft Desirability Toolkit is a method composed of 118 emotional words. In this toolkit, words are diverse and cover positive or negative reactions in aspects such as: design, functionality, performance and user experience. Designers use this method to obtain the emotional feedback and desirability from the participants after a user experience.

The presented document is organized as follows. Next section describes the main tasks of the application. In the third section, details of the usability study are described. Section four summarizes the main findings of the study and discussions. Finally, conclusions and upcoming efforts are presented.

2 App Description

The value of the AQI of sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and suspended particles (PM2.5 and PM10 particulate matter with aero- dynamic diameters under 2.5 and 10 μm, respectively) in the air are obtained from the website of the Environmental Secretariat of Quito Metropolitan District. This city office manages nine monitoring stations of the atmospheric conditions (meteorology and air pollutants) across the city. Then, the concentration of the pollutants for a specific position is obtained by interpolation. The method used to perform this interpolation is called Inverse Distance Weighted (IDW) [19]. Since the distribution of the monitoring stations is relatively sparse, the IDW is used to estimate the in-situ AQI index of a determined pollutant by only considering the actual values provided by the stations inside the footprint area surrounding the user. The purpose of such an App is to provide the population with a ubiquitous tool to inform and alert on possible pollution peaks and consequent health threats. “Figure 1” displays a screen-shot of the main user interfaces of the mobile air quality prototype and the navigation from one interface to another.

These functionalities are the following:

1. Since it is necessary to load several data from a server, the program starts with a progression bar that provides the user with a feedback on the delay before the application is available.
2. The geolocation of the user and the stations are displayed on a map of the city. The users can also deactivate their location.
3. It is possible to move across and zoom-in/zoom-out in the map by using the established joint gesture of thumb and forefinger.
4. It is possible to reinitialize the map according to the user’s position.
5. The user may change the map display settings, as shown in “Fig. 1” (panel 4). It means that the user can select the type of map according to the views: normal, terrain, satellite and hybrid. The color of the text is yellow when the map is shown in satellite or hybrid mode.
6. The pollution value is associated with a color scale. It is also possible to consult the documentation of the scale, by means of a help option, as shown in “Fig. 1” (panel 2). According to the pollution levels, the citizens are provided with specific recommendations. This option is particularly relevant for people with health condition, children and those who wish to practice outdoor activities.
7. The user may activate or deactivate the visual description of each station, as shown in “Fig. 1” (panel 3). If the user turns OFF this option, the stations will only show the pollution value.
8. It is possible to filter the contaminant parameters (status ON or OFF), in order to display only relevant pollutants for the user (“Fig. 1”, panel 5).
9. The user may change the radius of the measurement area, so that a different area is considered to estimate the concentration of the pollution at the user’s location.
10. A menu must be placed in the superior left corner of the user interface enables the users to select the contaminant they are interested in. Once the pollutant is chosen, its name is highlighted to give a visual confirmation that the request is proceeded.
11. The user can receive notifications or alerts when the system detects dangerous values of pollution.
12. The user may share the pollution values and their recommendations through the use of social networks. This allows the user to alert their relatives, friends and other citizens.

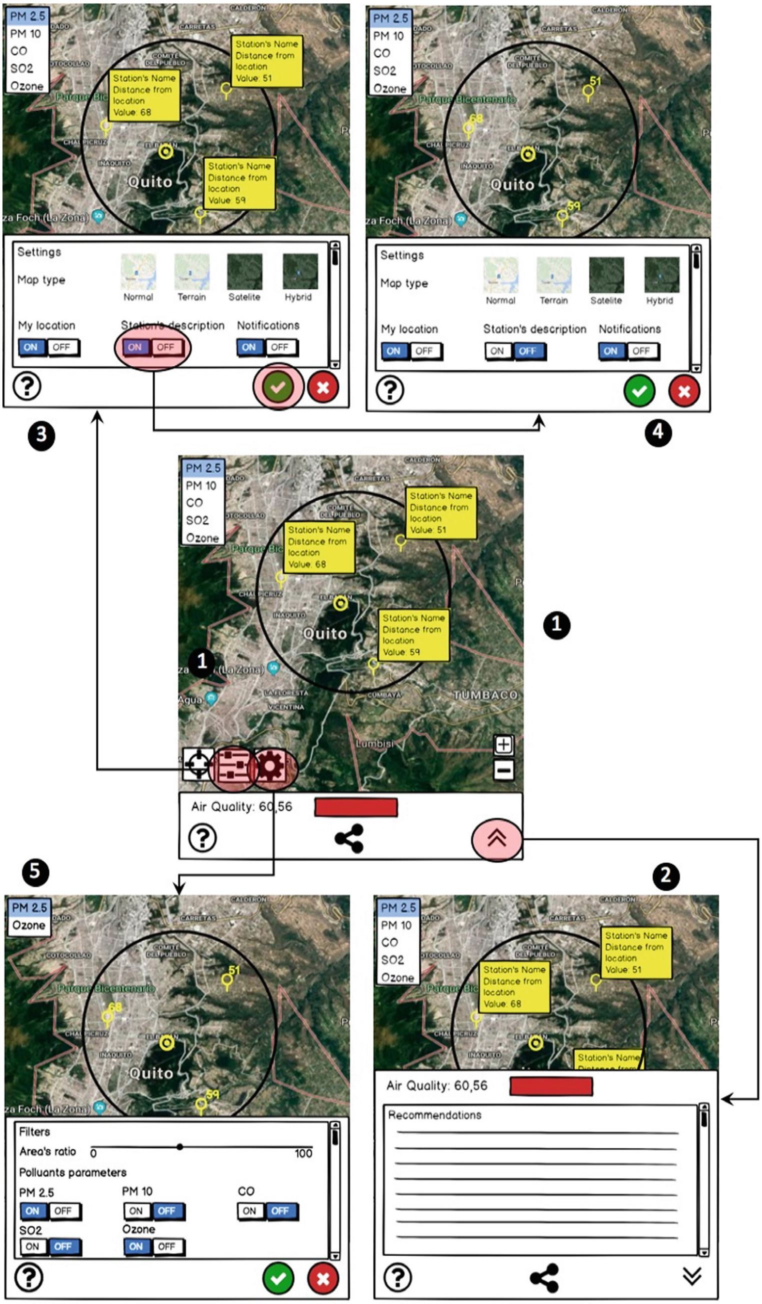


Fig. 1. The user interfaces of the air quality mobile application

3 The Experiment

We conducted an experiment involving final users to comprehend the usability for the mobile air quality interactive prototype.

3.1 Participants

A total of 51 people (distributed in 42 males or 82.3% and 9 females or 17.6%) participated in our usability experiment with the mobile air quality interactive prototype. The participants were selected by convenience sampling. For the participation in the experiment no specific skill was required. From the total of participants, 42 participants (or 82.3%) were between ages of 18 and 21 years old, while 7 participants (or 13.7%) were between 22 and 25 years old. Only 2 participants (or 3.9%) were between ages of 26 and 30 years old. From the 51 participants, 40 participants (or 78.43%) reported that they had never used an air quality app before. On the contrary, 11 (or 21.57%) participants who claimed to use an air quality system. Likewise, from the 11 participants, 5 (or 9.8%) have used these systems for less than 1 year. 2 participants (or 3.95%) have used these systems between 1 and 2 years. 3 participants (or 5.9%) have used these systems between 3 and 4 years. Finally, only 1 participant has used these systems between 5 and 6 years. The daily use of these systems was the following: 8 participants (or 15.7%) use these systems for less than 1 h. 1 participant (or 2.0%) use these systems between 1 and 2 h. 2 participants (or 3.9%) use these systems between 3 and 4 h.

3.2 Tasks

Table 1 presents the four tasks to be performed by participants during the experiment.

Table 1. Experimental tasks.

Nro.	Task
1	Change the map display settings to a terrain view
2	Consult the recommendations of the air quality index
3	Deactivate the visual description of each station
4	Filter the contaminant parameters

3.3 The Experimentation Procedure

We started by having each participant sign an informed consent. Next, we ask each participant to answer a demographic questionnaire. Then, each participant received training on how to use the mobile application. The participants were able to ask their questions after using the mobile air quality prototype. Immediately, when the scenarios of the experimentation were clarified, and after all the participants were self-confident with the use of the mobile application, they were invited to perform the four tasks. The experimentation lasted about 10 min. Once all the tasks accomplished, the participants were requested to fill in the CSUQ [16], plus the optional open question: Are there any comments you'd like to make about the air quality mobile application or your experience in using it during the experiment? Finally, participants were asked to select as many words as they wished from a Microsoft Desirability Toolkit [18] in order to describe their experience with the user interfaces.

4 The Results

This section presents the results of applying the IBM CSUQ instrument. Table 2 shows the measure scores grouped by the four categories of the CSUQ. Figure 2 (left panel) shows also, the results of the IBM CSUQ questionnaire.

In general, a favorable trend is observed, given that most participants valued positive responses. In almost the majority of cases, the averages in the answers to the questions were superior than 4.02 and the standard deviation obtained was less than 1.28. The results of the question Q2, Q3 and Q4 indicate that participants consider the system as simple to use and promoting an effective completion of the tasks.

The results of questions Q9, Q10, Q11, and Q15 concerning to the information quality (Mean (μ) = 5.15, Median (M) = 5.00, Std. Dev. (σ) = 1.47) suggest that the participants involved on the study were not in full agreement with the online help, error messages and other documentations offered with the mobile air quality prototype. On the contrary, the results obtained in question Q14, allows us to affirm that the participants found effective the information and helping them to perform the tasks asked with the system.

Table 2. Structure of the IBM CSUQ. Mean = (μ). Median (M). Std. deviation (σ)

Category	#	Question	μ	M	σ
SYSUSE	01	Overall, I am satisfied with how easy it is to use this system	5.45	6.00	1.00
	02	It was simple to use this system	5.66	6.00	1.21
	03	I could effectively complete my work using this system	5.70	6.00	1.00
	04	I was able to complete my work quickly using this system	5.54	6.00	0.96
	05	I was able to efficiently complete my work using this system	5.54	6.00	1.10
	06	I felt comfortable using this system	5.49	6.00	1.17
	07	It was easy to learn to use this system	5.80	6.00	1.31
	08	I believe I could become productive quickly using this system	5.47	6.00	1.34
INFOQUAL	09	The system gave error messages that clearly tell me how to fix problems	4.01	4.00	1.78
	10	Whenever I made a mistake using the system, I could recover easily and quickly	5.03	5.00	1.48
	11	The information (such as online help, on-screen messages, and other documentation) provided with this system was clear	5.11	6.00	1.45
	12	It was easy to find the information I needed	5.47	6.00	1.28
	13	The information provided for the system was easy to under-stand	5.60	6.00	1.20
	14	The information was effective in helping me complete the tasks and scenarios	5.41	5.00	1.20
	15	The organization of information on the system screens was clear	5.41,	6.00	1.28
INTERQUAL	16	The interface of this system was pleasant	5.37	5.00	1.21
	17	I liked using the interface of this system	5.49	6.00	1.23
	18	This system has all the functions and capabilities I expect it	5.54	6.00	1.08
OVERALL	19	Overall, I am satisfied with this system	5.60	6.00	1.04

Regarding the interaction quality (Mean (μ) = 5.47, Median (M) = 6.00, Std. Dev. (σ) = 1.18), the question Q17 suggests that only 31.37% of the participants did not like the user interfaces of the system.

Figure 2 (middle panel) shows the score values obtained for each question of the IBM CSUQ. The results to the system usefulness (Mean (μ) = 5.59, Median (M) = 6.00, Std. Dev. (σ) = 1.14). It allows us to indicate that the participants appreciated the platform. However, the results of question Q7 assert that 45.1% of participants do not believe that it was easy to learn the system. However, we corroborate that it is necessary to carry out complementary studies that measure the performance of using system tasks.

In relation to the information quality (Mean (μ) = 5.15, Median (M) = 5.00, Std. Dev. (σ) = 1.47), the results confirm a positive appreciation of the quality of the information for the participants involved on the study. Nevertheless, we consider that the improvements of the mobile air quality application should point towards: (i) the icons must be more representatives, and (ii) a more direct and short recommendation regarding the pollution value should be provided. Also, the results obtained in relation to the quality of the interface quality (Mean (μ) = 5.47, Median (M) = 6.00, Std. Dev. (σ) = 1.18) suggest that the UIs of the air quality mobile prototype are acceptable. However, we believe that we must make an effort to reduce the interactions between the interfaces. Overall satisfaction (Mean (μ) = 5.61, Median (M) = 6.00, Std. Dev. (σ) = 1.04) indicates that the mobile air quality prototype is positively appreciated. The responses of the participants suggest a positive attitude to resolve the errors when they are presented. In general, the usability of the system was also positively perceived by the users. We consider that the participants like the system and its functionalities.

4.1 The Complementary Analysis of Variance and Correlation

This section presents the complementary study to analyze the one-way variance and the correlations of the results. Figure 2 (right panel) presents a strip plot containing the CSUQ average response for all the categories. The "Jittering" representation was selected in order to provide a better visualization and understanding. We highlight three measures: (1) the minimum score in each category given by the participants; (2) the mean value obtained for each category; and (3) the confidence interval for each category. Based on the results, we can affirm that there are no important differences between the categories. The value of (p- value = 0.1) indicate an approximating positive appreciation for all statements of the IBM CSUQ.

Regarding the correlation between the answers between the categories. Figure 2 (right panel) shows that the OVERALL and SYSUSE measurements were the highest (r = 0.77). This allows us to state that the positive general perception of the participants is associated with each positive response for the questions of the SYSUSE category. This phenomenon can also be observed in Fig. 2 (left panel) where the results of the aforementioned questions were perceived as highly positive. The higher correlation was acquired with the SYSUSE category (r = 0.76) for INTERQUAL. It means that the more appreciate and more pleasant is using the system for the user, the more friendly and productive is.

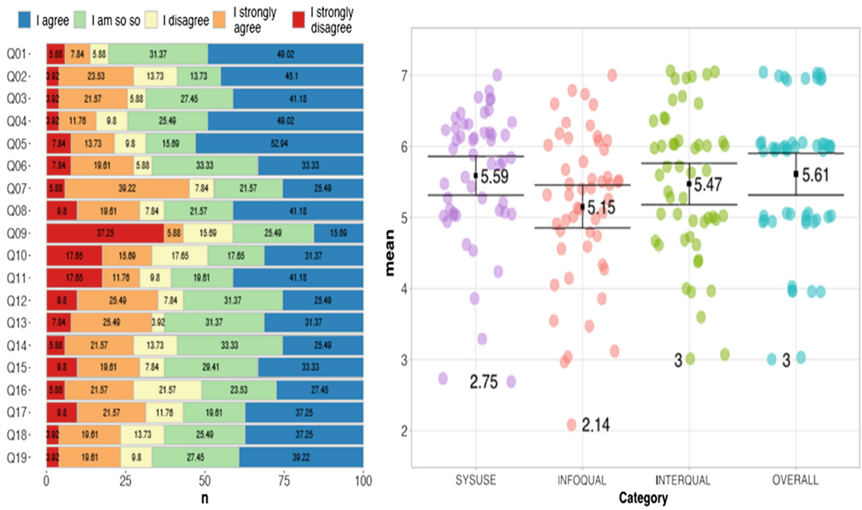


Fig. 2. The IBM CSUQ results. Left panel: the IBM CSUQ questions summarized. Right panel: mean average responses grouped by categories.

4.2 The Analysis of Microsoft Desirability Survey

This section describes the score values achieved when applying the Microsoft Desirability Toolkit [18] to describes the experience of participants with the application. The participants selected an average of 12.5 words, with half of them selecting more than 9 words.

Figure 3 (left panel) shows a word cloud using all the 91 words selected by the participants. We found that the first 10 words with the highest frequency where positive (see right panel of the Fig. 3). Table 3 shows the relationships between the emotional words: Clear, Usable and Easy-to-use, which describes how often the selected emotional words appear together relative to how often they appear separately. The words Clear, Usable and Easy-to-use are the most frequent words picked by the users and are presented in an individual column in Table 3 Then follows a column with the top correlated words and the correlation value. A threshold value of $\theta^{corr} > 0.3$ is used for selecting the correlated words.

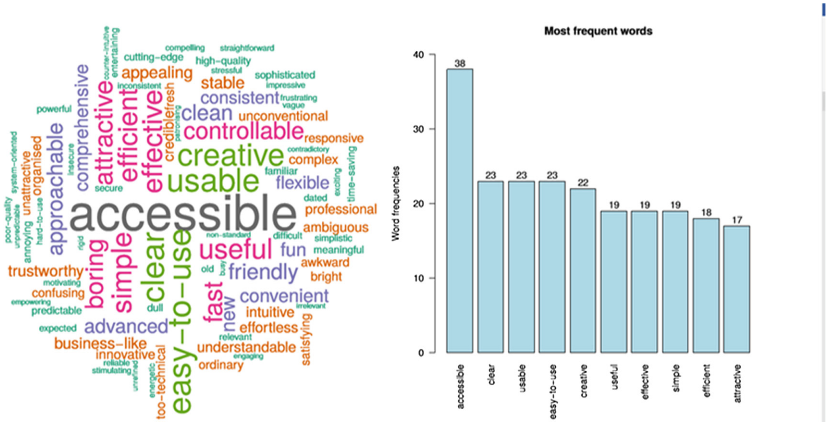


Fig. 3. Desirability results from the Microsoft reaction cards.

Table 3. Correlations between words

Word	Correlated with	Corr	Word	Correlated with	Corr	Word	Correlated with	Corr
Clear	Intuitive	0.44	Usable	Credible	0.44	Easy to use	Controllable	0.54
	Comprehensive	0.39		Business-like	0.37		Consistent	0.45
	Flexible	0.35		Trustworthy	0.37		Approachable	0.41
	Effortless	0.33		Useful	0.36		Friendly	0.41
	Efficient	0.32		Fresh	0.36		Comprehensive	0.39
	Time-saving	0.32		Flexible	0.35		Simple	0.36
	Predictable	0.32		Approachable	0.33		Satisfying	0.36
	Appealing	0.30		Sophisticated	0.32			

5 Conclusions

Throughout this contribution, we detailed a usability study that considers the efficiency, effectiveness and satisfaction of 51 participants who made use of a mobile application for public air quality information. The work intends to provide some guidelines to improve the development of air quality mobile applications. In the context of this research, an appropriate usability is fundamental for a correct understanding and interpretation of the data offered by the system.

In general, the results obtained induce a good usability of the application. For us the most important results obtained during the study, and its subsequent analysis, was the possibility of identifying the early usability problems and interaction concerns. Among the problems encountered, we have identified that a better handling of errors and system documentation is necessary. Likewise, the performed test by the participants shows that a significant part of the population does not believe that the mobile application was intuitive. Likewise, the presence of poor user feedback was considered as a barrier in the acceptance of the application.

We must do what is necessary to improve these aspects and make users feel comfortable using the mobile application in development. Future efforts should have the following directions: (i) considering the opinion of these individuals to redesign certain aspects of the interface; (ii) carrying out tests to evaluate the user's performance of users according to the completion time, number of interactive movements, error rates, among others; (iii) carrying out cognitive tests considering aspects like: mental workload and frustration; and (iv) conducting accessibility evaluations.

As a final reflection, we conclude that User-Centered Development practices in conjunction with air quality analysis research and the usability of monitoring tools are necessary when implementing intuitive techniques of data visualization and information interaction, such as multi-modal user interfaces.

Acknowledgements. The work is supported by Universidad de Las Américas (UDLA) under Grant No.: ERI.WH.18.01 and the Consorcio Ecuatoriano para el Desarrollo de Internet Avanzado CEDIA-CEPRA under Grant No.: CEPRA XII-2018-13.

References

1. Osseiran, N., Chriscaden, K.: Air pollution levels rising in many of the world's poorest cities (2016)
2. UNECE: Air pollution and health. https://www.unece.org/environmentalpolicy/conventions/envlrtapwelcome/cross-sectoral-linkages/air-pollution-and-health.html. Accessed 17 Jan 2019
3. World Health Organization: 7 million premature deaths annually linked to air pollution. World Health Organization, Geneva (2014)
4. Zalakeviciute, R., Rybarczyk, Y., López-Villada, J., Suarez, M.V.D.: Quantifying decade-long effects of fuel and traffic regulations on urban ambient PM2.5 pollution in a mid-size south american city. Atmos. Pollut. Res. **9**(1), 66–75 (2018)
5. Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature **525**(7569), 367 (2015)
6. Hoek, G., Krishnan, R.M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., Kaufman, J.D.: Long-term air pollution exposure and cardio-respiratory mortality: a review. Environ. Health **12**(1), 43 (2013)
7. Pérez-Medina, J.L., Zalakeviciute, R., Rybarczyk, Y.: Assessing the usability of air quality mobile application. In: 2019 International Conference on eDemocracy & eGovernment (ICEDEG). IEEE (2019)
8. United Nations: World population prospects: the 2015 revision. United Nations Economic and Social Affairs, vol. 33, no. 2, pp. 1–66 (2015)
9. Limb, M.: Half of wealthy and 98% of poorer cities breach air quality guidelines. BMJ **353** (2016). https://doi.org/10.1136/bmj.i2730
10. Theofanos, M., Quesenbery, W.: Towards the design of effective formative test reports. J. Usability Stud. **1**(1), 27–45 (2005)
11. Kieffer, S., Ghouti, A., Macq, B.: The agile UX development lifecycle: combining formative usability and agile methods (2017)
12. Pérez-Medina, J.L., Vanderdonckt, J.: Sketching by cross-surface collaboration. In: Rocha, Á. et al. (eds.) 2019 International Conference on Information Technology Systems, ICITS 2019. AISC, vol. 918, pp. 1–12. Springer (2019). https://doi.org/10.1007/978-3-030-11890-7_38

13. Stewart, T.: Ergonomic requirements for office work with visual display terminals (VDTs): part 11: guidance on usability. International Organization for Standardization ISO 9241 (1998)
14. Tullis, T.S., Stetson, J.N.: A comparison of questionnaires for assessing website usability. In: Usability Professional Association Conference, vol. 1 (2004)
15. Brooke, J., et al.: SUS-A quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
16. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Hum.-Comput. Interact. **7**(1), 57–78 (1995). https://doi.org/10.1080/10447319509526110
17. Nunnally, J.C., Bernstein, I.H., ten Berge, J.M.: Psychometric Theory, vol. 226. McGraw-hill, New York (1967)
18. Benedek, J., Miner, T.: Measuring desirability: new methods for evaluating desirability in a usability lab setting. Proc. Usability Prof. Assoc. **2003**(8–12), 57 (2002)
19. De Mesnard, L.: Pollution models and inverse distance weighting: some critical remarks. Comput. Geosci. **52**, 459–469 (2013)

User Experience Assessment of a Tele-Rehabilitation Platform: The Physiotherapist Perspective

Jorge Luis Pérez-Medina(✉), Karina Beatríz Jimenes-Vargas, Patricia Acosta-Vargas, Yves Rybarczyk, and Mario González

Intelligent and Interactive Systems Lab (SI2 Lab), Universidad de Las Américas (UDLA), Quito, Ecuador
{jorge.perez.medina,karina.jimenes,patricia.acosta, yves.rybarczyk,mario.gonzalez.rodriguez}@udla.edu.ec

Abstract. This paper summarizes the heuristic evaluation carried out to the *ePHoRt*, a web-based platform that aims at improving the rehabilitation of patients after an arthroplasty. The development of the interfaces is carried out by an agile, collaborative and user-centered methodology. The heuristic evaluation was performed by four experts, considering four ergonomics principles as follows: *user guidance*, *user support*, *user control and freedom*, and *user support*. The results have allowed us to obtain a series of recommendations that could be applied to improve the functionalities and the user interface of our tele-rehabilitation platform.

Keywords: Usability · Heuristic evaluation · E-health · Physical rehabilitation · Physiotherapist activities

1 Introduction

The *ePHoRt* project is a web-based platform that aims at improving the rehabilitation of patients after an arthroplasty [1, 2]. Usability studies are performed from the patient's perspective [3, 4]. However, new challenges are primarily related to develop a system to support the physiotherapist for planning, monitoring and evaluating rehabilitation exercises. These aspects are important in the therapeutic process, because it allow for selecting the most appropriate exercises (planning), assessing in real time the execution of the movements (monitoring), and quantifying the progress of the patient (recovering) [5–7]. The current clinical practice lacks computerized methods to support physical rehabilitation programs, although the stakeholders (health professionals, caregivers, and patients) agree on the fact that IT can enhance a certain autonomy of the patients by promoting a therapeutic education [8]. In order to make sure that the developed platform supports effectively and efficiently the physiotherapist's activities, an agile, collaborative and user-centered methodology is used. The use of the system requires studying the usability of the interfaces in early stages of a collaborative development process focused on the user.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 463–473, 2020.
https://doi.org/10.1007/978-3-030-20040-4_42

A sets of usability heuristics have been grouped under four ergonomics principles as follows: user guidance, user effort, user control and freedom, and user support. This paper summarizes a preliminary heuristic evaluation carried out on the *ePHoRt* application and from the physiotherapist's perspective. The heuristic evaluation is performed by three usability experts and is focused on the physiotherapist's interface. The results are intended to provide a series of recommendations that will be applied to improve the functionalities of the user interface. The remaining of the paper is organized into six section. The next section presents some related work in the domains of tele-rehabilitation and usability evaluation. Section 3 describes the case study. Section 4 presents a task-oriented usability evaluation method. In Sect. 5, the usability evaluation results are presented. Section 6 suggests some improvements according to the usability evaluation results. Finally, the last section presents conclusions and future work.

2 Related Work

The *ISO* standard 9241-11 [9] provides a framework for understanding the concept of usability and applying it to situations where people use interactive systems, and other types of systems (including built environments), and products (including industrial and consumer products) and services (including technical and personal services). The official *ISO* 9241-11 definition of usability is "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" [9]. Usability evaluation can be formative or summative [10]. Formative usability evaluation aims to evaluate a product or a design. It allows to identify shortcomings in order to produce a set of recommendations with the goal of improving the design before it has been finalized. Currently, there are many methods of inspection and testing that make it possible to obtain the usability of an application. These inspection methods are based on usability problems known as heuristic metrics, among which Nielsen's 10 usability heuristics are the most widely used [11].

A systematic literature review to develop usability heuristics was presented in [12]. This work performs an exhaustive review of 73 studies related to usability heuristics conducted between 2006 and 2016. Recently, an extended set of usability heuristics integrating the usability heuristics of Nielsen [13] with the ergonomics criteria presented in [14] has appeared. However, we focus on using a set of usability heuristics for the evaluation of the *ePHoRt* platform because they have been grouped under four ergonomics principles as follows: user guidance, user effort, user control and freedom, and user support [13].

3 Case Study

3.1 The Home Motor Rehabilitation Platform

ePHoRt [2] is a web-based tele-rehabilitation application to support people recovering from a hip replacement surgery. We adopt the *User-Centered agile.*

Development (UCD) approach [15] to specify and design the main functionalities of the therapeutic programs (TPs), through a flexible and efficient manner. This process is described by using the task tree model. It was inspired by the *Concur Task Trees (CTT)* notation described in [16]. In our case, icons may vary due to the toolkit used. We have adopted this representation because it supports the design of interactive applications for user interface model-based design. It generates a tree graphical representation of the hierarchical decomposition of the tasks provided by the platform. A TP is a plan that describes a set of exercises that should be accomplished by the patient over a certain period of time. Figure 1 shows the tasks that must be performed by the physiotherapist to create a therapeutic program.

3.2 Application's Tasks

According to the case study, Table 1 shows the main tasks to create a therapeutic program with *ePHoRt*. These tasks are used by the experts to make the heuristic evaluations of the physiotherapist interfaces of *ePHoRt*.

Table 1. The evaluation tasks

Reference	Task
T1	Add general information
T2	Add medical history
T3	Write a pre-surgical evaluation
T4	Define the stages of a therapeutic program
T5	Assign exercises to the therapeutic program

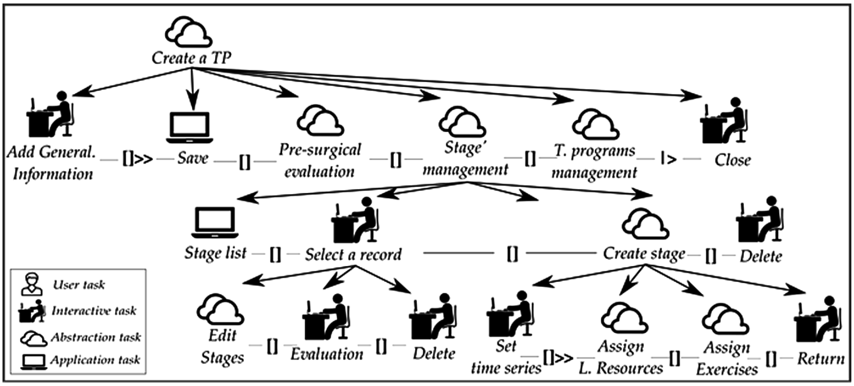


Fig. 1. Task model to create a therapeutic program

3.3 Add General Information

The physiotherapist can add general information regarding the therapeutic program. This includes a title, a brief description, a period of beginning and ending, the patient who requires rehabilitation and his/her referring physician. Figure 2 shows the user interface to create a therapeutic program.

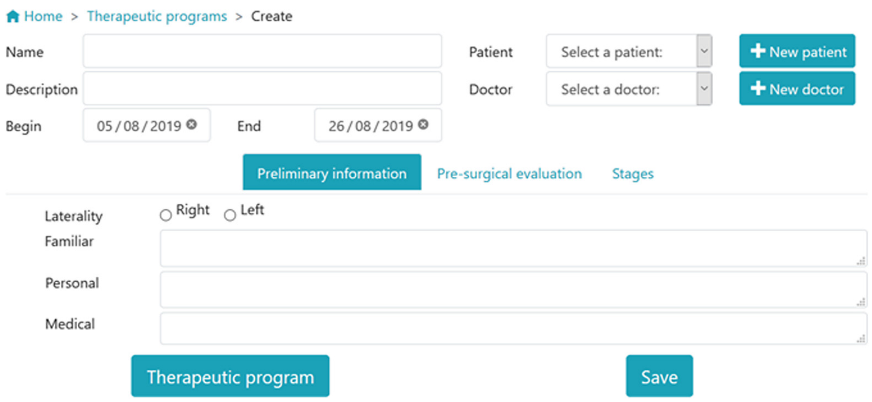


Fig. 2. UI to create a therapeutic program

3.4 Add Medical History

The physiotherapist has the possibility of documenting the preliminary information of the patient. This information refers to any relative, and/or personal medical record (see Fig. 2).

3.5 Write a Pre-surgical Evaluation

With the creation of the therapeutic program, the physiotherapist has the possibility of registering a pre-surgical evaluation. Two types of evaluation are available: (1) functional, or (2) by body segments. The functional evaluation requires an assessment of the patient's physical-motor conditions. The evaluation by body segments considers an assessment of the limbs of the patient, in order to evaluate the levels of pain, strength or movement.

Upper panel of Fig. 3 shows the user interface to write a pre-surgical evaluation.

3.6 Define the Stages of a Therapeutic Program

The physiotherapist must indicate the number of stages in which a therapeutic program will be divided (see middle panel of Fig. 3). At each stage the physiotherapist can record the following elements:

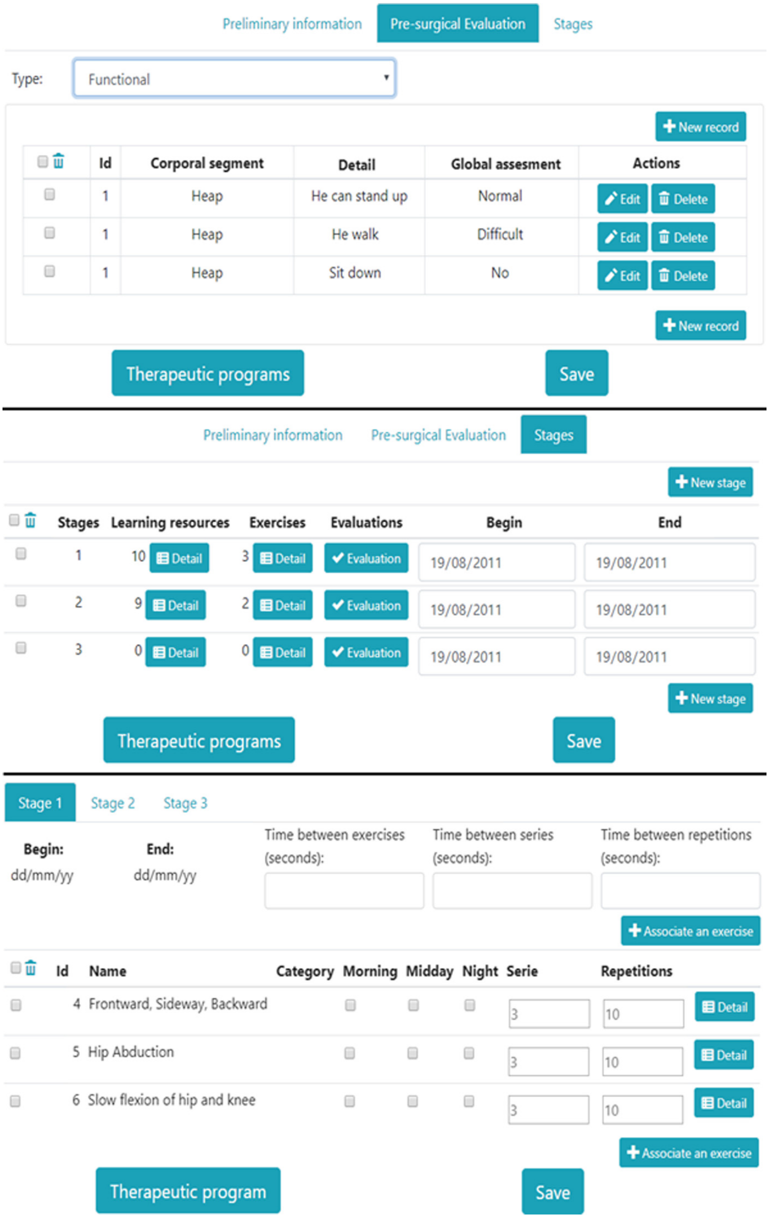


Fig. 3. UIs to create a new therapeutic program. Upper panel the UI to write a pre- surgical evaluation. Middle panel the UI to define the stages of a therapeutic program. Bottom panel the UI to assign exercises to a therapeutic program.

- a consecutive period of validity.
- a set of learning resources.
- a set of exercises the patient must perform during a rehabilitation session.
- a post-surgical evaluation the physiotherapist must perform at the end of a stage.

The physiotherapist can define the rest time the patient must respect between the exercises, the series and the repetitions. The hierarchical links between these functionalities are illustrated in Fig. 1.

3.7 Assign Exercises to the Therapeutic Program

The functionality of assigning exercises to a therapeutic program allows the physiotherapist to establish the link between the therapeutic program and the exercises that should be performed. For each stage, the physiotherapist associates a set of exercises, as shown in the bottom panel of Fig. 3. Each associated exercise must contain the number of sessions for a particular day, the number of series in each session and the number of repetitions per series.

4 Usability Evaluation

This section describes the details of the task-oriented usability inspection method [13] used in this study.

4.1 Materials and Methods

Four experts tested the mobile application in order to find the difficulties end- users might have when using the application. Each expert received a user manual for the application, including the evaluation tasks, the set of usability heuristics (according to [13]), a series of steps to perform the evaluation, and an evaluation template. The usability experts used the set of 14 heuristics presented in Table 3. Usability problems are evaluated based on the severity listed in Table 2. The levels of severity range from 1 to 5. The value of 1 represents the lowest level of severity while a value of 5 represents the most serious problem of usability.

Table 2. The list of severities

Reference	Description
1	A usability problem is not considered, in its entirety
2	Barely aesthetic problem: it does not need to be modified, unless there is time available
3	Less usability problem: the solution to this problem should have low priority
4	Major usability problem: it is important to solve it, for this high priority must be given
5	Usability catastrophe: it is mandatory to resolve it, before the product is disclosed

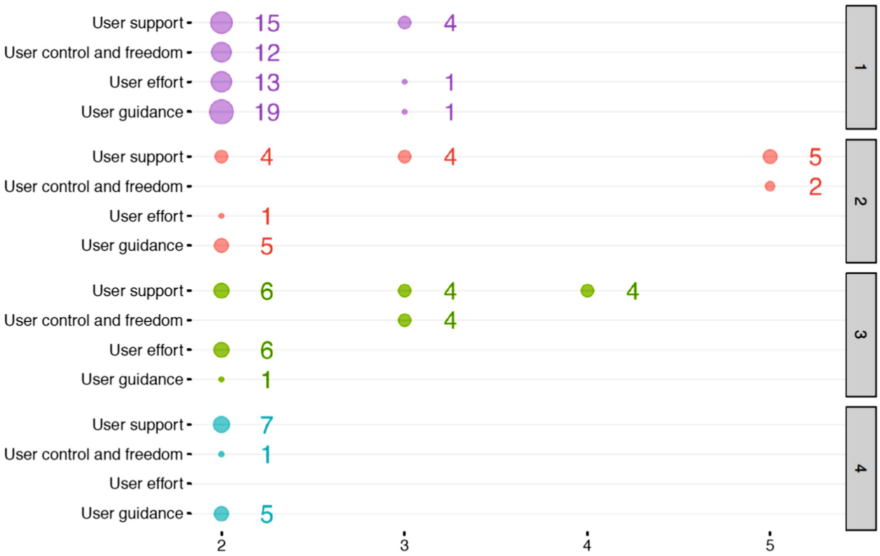


Fig. 4. Usability problems and severity for all experts

4.2 Procedure

The usability problems are evaluated by following a task-based approach. Table 1 presents the evaluation tasks that were used during the usability evaluation. Each heuristic presented in Table 3 was used to explain and comment the usability problems found by experts when analyzing the tasks presented in Table 1. The evaluation procedure was as follows:

1. Each expert became familiar with the application.
2. Each expert made an individual evaluation considering the set of heuristics evaluation and the severity's levels.
3. The results of each expert were merged for further analysis.

5 Results and Discussion

After the individual evaluations, the number of problems detected by each expert varied between 13 and 65 (see Fig. 4). We analyzed each individual evaluation in order to eliminate the duplicates and the false problems. Table 3 shows the collaborative consolidation of usability problems, grouped by categories and severities. Overall, a total of 124 usability concerns where founded. Most of the usability problems detected are related to:

1. There is no help option and documentation for the user.
2. It is not clear when the user should use the *"Save"* button.
3. The buttons must be more representative of the action that the user must perform.

4. Some of the buttons do not suggest the action that the user should do. For instance: *"New record"* in the upper panel of Fig. 3.
5. The registration groups in the tables present in the interfaces are not identified.
6. Tables using in the UIs are not the same presentation style.
7. The user can not delete a particular stage.
8. The user can not dissociate a particular exercise.
9. As the records of the tables increase, it will increase the level of scrolling that the user should perform.
10. There is no separation between the evaluations by functional type and by body segments.

Table 3 shows the usability problems per usability heuristics. User support and are mainly related to all heuristics. The user interfaces does not have help and documentation. Error messages are non-existent and the messages of the buttons do not guide the therapist in the task to be performed. User guide problems are mainly related to the fact that it is difficult to have the feedback of the user tasks and the does not help the user to understand the result of their task. These problems are aligned with user support. This is mainly because the system does not offer good compatibility with the user, nor does it provide help and documentation. The problems obtained in the category of user effort are due to the that the user does not have a good understanding of the table elements presented in the user interfaces (see Fig. 3). Finally, user control and freedom are related to the fact that the user does not have enough flexibility to delete the stages and the exercises.

Table 3. Usability problems and severity for all tasks

Usability heuristics			Severity					Total problems
Group	ID	Description	1	2	3	4	5	
User guidance	H1	Prompting	13	7	0	0	0	7
	H2	Feedback	9	11	0	0	0	11
	H3	Information architecture	14	5	1	0	0	6
	H4	Grouping/distinction	13	7	0	0	0	7
Total - user guidance			0	30	1	0	0	31
User effort	H5	Consistency	14	6	0	0	0	6
	H6	Cognitive workload	12	7	1	0	0	8
	H7	Minimal actions	13	7	0	0	0	7
Total - user effort			0	20	1	0	0	21
User control and freedom	H8	Explicit user actions	14	4	2	0	0	6
	H9	User control	14	4	2	0	0	6
	H10	Flexibility	13	5	0	0	2	7

(*continued*)

Table 3. (*continued*)

Usability heuristics			Severity					Total problems
Group	ID	Description	1	2	3	4	5	
Total - user control and freedom			0	13	4	0	2	19
User support	H11	Compatibility with the user	15	5	0	0	0	5
	H12	Task guidance and support	5	13	2	0	0	15
	H13	Error management	3	10	5	2	0	17
	H14	Help and documentation	4	4	5	2	5	16
Total - user support			0	32	12	4	5	53
Totals			0	95	18	4	7	124

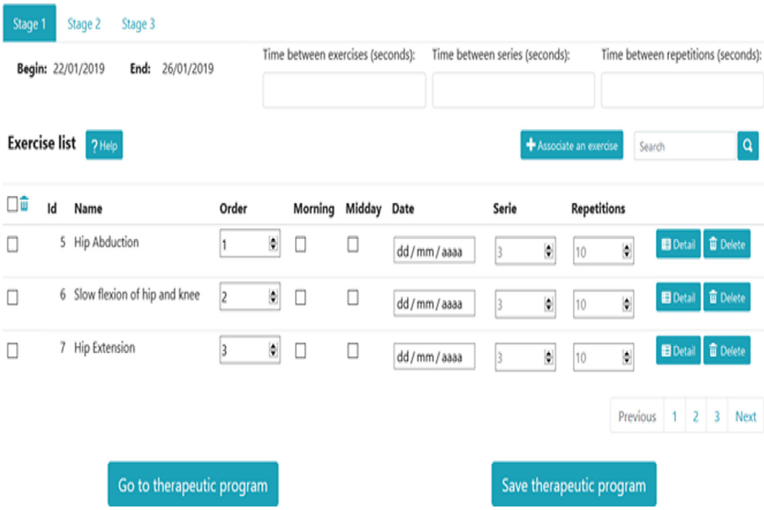


Fig. 5. Proposal for the UI to define the stages of a therapeutic program

6 Recommendations

Figure 5 shows some modifications performed to the user interface to assign exercises for each stage of a therapeutic program, as an alternative to solve the problems encountered by the usability experts. Modifications like a pagination to navigate between the exercises of each stage, an input field to incorporate the date for each exercise, a button to disassociate exercises from a stage and a search box to identify specific exercises and to allow the user filtering and/or sorting the elements of the tables present in the interfaces considering a specific text. Also, the input field to incorporate the order of the exercises in the stage. It allows alter the order of apparition of the stages or exercises.

On the other hand, the text associated to the button *"Therapeutic program"* was changed by the text *"Go to the therapeutic program"* while the button *"Save"* was changed by the text *"Save the therapeutic program"*. It is a better description of what it means due to be more understandable for the user.

7 Conclusions

In this paper, we carried out a usability inspection with the *ePHoRt* platform. Our objective was to explore the usability concerns of the application, but also, we intended to investigate the extensions of this platform. The results obtained suggest improving existing functionalities, and also, development perspectives. Future efforts should focus on conducting subsequent usability evaluations, once the suggestions and changes in the platform are completed. As a priority we consider carrying out user experiences involving physiotherapists in clinical environments.

Acknowledgements. The work is supported by Universidad de Las Am´ericas (UDLA) under Grant No.: SIS.JPM.18.01.

References

1. Rybarczyk, Y., Pérez-Medina, J.L., Leconte, L., Jimenes, K., González, M., Esparza, D.: Implementation and assessment of an intelligent motor tele-rehabilitation platform. Electronics **8**(1), 58 (2019)
2. Rybarczyk, Y., Deters, J.K., Gonzalvo, A.A., González, M., Villarreal, S., Esparza, D.: ePHoRT project: a web-based platform for home motor rehabilitation. In: World Conference on Information Systems and Technologies, pp. 609–618. Springer (2017)
3. Pérez-Medina, J.L., Gonzalez, M., Pilco, H.M., Jimenes, K., Acosta-Vargas, P., Sanchéz-Gordon, S., Calle-Jimenez, T., Esparza, D., Rybarczyk, Y.: Usability study of a web-based platform for home motor rehabilitation. IEEE Access **7**(1), 58 (2019)
4. Pilco, H., Sanchéz-Gordon, S., Calle-Jimenez, T., Pérez-Medina, J.L., Rybarczyk, Y., Jadan-Guerrero, J., Guevara-Maldonado, C., Nunes, I.L.: An agile approach to improve the usability of a physical telerehabilitation platform. Appl. Sci. **9**, 480 (2019)
5. Rybarczyk, Y., Kleine Deters, J., Cointe, C., Esparza, D.: Smart web-based platform to support physical rehabilitation. Sensors **18**(5), 1344 (2018)
6. Rybarczyk, Y., Deters, J.K., Gonzalo, A.A., Esparza, D., Gonzalez, M., Villarreal, S., Nunes, I.L.: Recognition of physiotherapeutic exercises through DTW and low-cost vision-based motion capture. In: International Conference on Applied Human Factors and Ergonomics, pp. 348–360. Springer (2017)
7. Deters, J.K., Rybarczyk, Y.: Hidden markov model approach for the assessment of tele-rehabilitation exercises. Int. J. Artif. Intell. **16**(1), 1–19 (2018)
8. Rybarczyk, Y., Vernay, D.: Educative therapeutic tool to promote the empowerment of disabled people. IEEE Latin Am. Trans. **14**(7), 3410–3417 (2016)
9. Stewart, T.: Ergonomic requirements for office work with visual display terminals (VDTs): part 11: guidance on usability. International Organization for Standardization ISO 9241 (1998)

10. Tullis, T., Albert, B.: Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (2008)
11. Nielsen, J.: Usability engineering. Elsevier, Amsterdam (1994)
12. Quiñones, D., Rusu, C.: How to develop usability heuristics: a systematic literature review. Comput. Stand. Interfaces **53**, 89–122 (2017)
13. Pribeanu, C.: A revised set of usability heuristics for the evaluation of interactive systems. Informatica Economica **21**(3), 31 (2017)
14. Bastien, J.C., Scapin, D.L.: A validation of ergonomic criteria for the evaluation of human-computer interfaces. Int. J. Hum.-Comput. Interact. **4**(2), 183–196 (1992)
15. Pérez-Medina, J.L., Vanderdonckt, J.: Sketching by cross-surface collaboration. In: Rocha, Á. et al. (eds.) 2019 International Conference on Information Technology Systems, ICITS 2019. AISC, vol. 918, pp. 1–12. Springer (2019)
16. Paterno, F.: Model-Based Design and Evaluation of Interactive Applications. Springer Science & Business Media, Cham (1999)

Evaluation of Smart Phone Open Source Applications for Air Pollution

Rasa Zalakeviciute[1](✉), Katiuska Alexandrino[1], Patricia Acosta-Vargas[2], Jorge-Luis Pérez-Medina[2], and Wilmar Hernandez[3]

[1] Biodiversidad Medio Ambiente y Salud (BIOMAS), Universidad de Las Américas, calle José Queri y Av. de los Granados/Bloque 7, Quito 170125, Ecuador
{rasa.zalakeviciute, katiuska.alexandrino}@udla.edu.ec

[2] Intelligent and Interactive Systems Lab (SI2 Lab), Universidad de Las Américas (UDLA), Campus Queri, calle José Queri y Av. de los Granados/Bloque 7, Quito 170125, Ecuador
{patricia.acosta, jorge.perez.medina}@udla.edu.ec

[3] Universidad de Las Americas, Campus Queri, calle José Queri y Av. de los Granados/Bloque 7, Quito 170125, Ecuador
wilmar.hernandez@udla.edu.ec

Abstract. Global industrialization, urbanization and technological development have been rapidly changing the atmospheric composition. Air pollution is one of the most critical concerns to human and environmental health, causing respiratory, cardiovascular diseases and incidents of cancer, responsible for seven million premature deaths every year. While most of the major cities have some level of air quality monitoring, the data reported on the official websites are not easily accessed, and are rarely communicated to the citizens, mainly limited to the newspaper and television reports of extreme pollution events. A more efficient way to stay informed of the risks of exposure to bad air quality is a mobile application that can be accessed from anywhere. In this study, we evaluate available open source mobile phone applications for Android and iOS systems. Specifically, we analyze the usability and accessibility functions of the applications in the most populous and most contaminated South American capital cities.

Keywords: Evaluation · Mobile applications · Air pollution

1 Introduction

Rapid global industrialization, urbanization and technological development have been increasingly changing the composition of the atmosphere. Air pollution is becoming one of the most critical concerns to human and environmental health [1, 2]. An array of atmospheric pollutants are causing respiratory, cardiovascular diseases and incidents of cancer, costing seven million of premature deaths every year [3]. Unfortunately, these numbers are forecasted to double in the next few decades [4].

Governmental and non-governmental organizations have demonstrated the need to develop action plans aimed to reduce air pollution [5–7]. While this is difficult to

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 474–484, 2020.
https://doi.org/10.1007/978-3-030-20040-4_43

achieve immediately, on the other hand, avoiding the exposure to critical concentrations could be a feasible solution for the current circumstances [8, 9]. The awareness of the air quality and the exposure to toxic atmospheric pollutants relies on the existing urban networks of air quality monitoring. This highly depends on the dedicated funding for air quality monitoring purposes, and often works discriminatory against poorer countries and cities due to high investments to infrastructure. While most of the major cities have some level of air quality monitoring, the data reported on the official websites are not easily accessed, and are rarely communicated to the citizens, mainly limited to the newspaper and television reports of extreme pollution episodes. A more efficient way to inform the population of the risks of being exposed to bad air quality in their city is a mobile application that can be accessed from anywhere [10, 11]. Recent developments of mobile phone applications have been quickly saturating the market. Several existing mobile phone applications share the information on the air quality in the main urban conglomerates. However, not all of them offer the same quality of information, and the necessary parameters, crucial in helping the user to plan or avoid their daily activities based on the health risks [12].

Therefore, in this first of the kind study, we propose an evaluation of available open source mobile phone applications for Android and iOS systems. This work aims to classify the selected applications by the list of the necessary mobile application parameters (e.g. Air Quality Index (AQI), map, historical trend of pollution, health recommendations, etc.) to raise the awareness of the public about the current air quality conditions in the region. In addition, for our case study of usability of the mobile applications, we choose to evaluate the functionality provided by the tools in order to recommend the most complete application offering all functionalities essential for the users. Finally, most of the mobile applications are not accessible or inclusive, so the Web Content Accessibility Guidelines 2.1 are applied in order to achieve a higher level of accessibility for users with disabilities. The results of this study will enable the development of more usable and accessible mobile applications to ensure the availability of crucial information in order to make educated decisions to prevent health risks.

2 Methodology

To evaluate the mobile phone applications, 12 applications were downloaded onto the android (Samsung Galaxy J7, OS 6.0.1) and Apple (i-Phone 7, iOS 12.1.2) cellular phones. Selected applications include Air Visual (IQ*Air*), Air Matters (Air-Matters), BreezoMeter (Eureka Forbes), Air Bubbles (Wynd Technologies, Inc.), Sh**t! I Smoke (A.Martiny, M.Celho), Plume (Plume Labs), Air Quality - Monitoring Global Air Quality (Mingwei Zhang), HazeToday (Emir Fithri Bin Samsuddin), Global Air Quality Index (Mai Huy Toan, (c)WMOApp), Air quality – AQI PM2.5 Checker (Amber Mobile Limited), EPA AIRNow (EPA), AirQualityEx. Table 1 presents the list of features that were evaluated for both operating systems (A-Android and I-IOS).

Based on the importance of the parameters for human health considerations, we have retained three features to evaluate the usability of each air quality application: risk/color scale, map and health recommendations. The traffic light-based risk/color scale was selected in order to understand the air quality conditions, which is an intuitive way to interpret otherwise unfamiliar concentrations of the pollutant or AQI. The availability of a map of the area was also chosen to help the user know how close they are to the monitoring data point reporting the air quality. And finally, health recommendations are also extremely important for the user, to help them plan their daily activities in terms of health risk prevention.

Table 1. Features evaluated for availability for mobile phone air quality applications.

#	Feature	Description
1	Air quality index (AQI)	
2	Concentration of pollutant	
3	The scale of risk of the pollution	Based on the colors of traffic light (most common – green (safe) to red (dangerous)) or other (lighter to darker grey, blue to red)
4	Pollutant information	
5	Weather forecast	
6	AQI forecast	
7	Graphs	
8	Map	
9	Health recommendations	
10	Comparison with other cities	
11	Historical data	

The complexity of the feature was evaluated by three interaction criteria: (i) direct interaction of the end-user selecting an element (for example clicking on AQI button); (ii) the scrolling quantity that the end-user has to perform to visualize desired information; and (iii) the number of navigations necessary between the screens to complete a specific task.

Mobile applications [13] must be accessible to all types of users, to guarantee this requirement, we applied the Web Content Accessibility Guidelines (WCAG) 2.1. According to the World Wide Web Consortium (W3C), accessibility refers to a mobile application design that allows people to perceive, understand, navigate and interact with the application while providing content [14]. More detail of the method is described in a previous study [12].

3 Results and Discussion

The completeness of the 12 mobile phone applications for air quality was evaluated by estimating an inclusiveness of the characteristics presented in Table 1 available as features. Table 2 presents these findings for Android (A) and Apple (I) mobile phone applications. The reviewed free access applications include between 11 and 2 parameters. More of the applications are freely accessible for Apple (12 applications) if compared to the Android (7 applications) operating system. The most complete in terms of parameters are Air Matters (11 features), AirVisual (9 features) and Plume (9 features), Air Checker (8 features, but only available for Apple) (see Table 2).

Table 2. Features evaluated for availability in 12 mobile phone applications: for Android (A) and i-Phone (I). (1) Air quality index (AQI); (2) Concentration of pollutant; (3) Risk/color scale; (4) Pollutant information; (5) Weather forecast; (6) AQI forecast; (7) Graphs; (8) Map; (9) Health recommendations; (10) Comparison with other cities; (11) Historical data.

App name	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
AirVisual	A/I	A/I	A/I	A/I	A/I	A/I	A/I	A/I	A/I		
Air matters	A/I	A/I	A/I	A/I	A/I	A/I	A/I	A/I	A/I	A/I	A/I
Air bubbles	A/I	I	A/I					A/I			
BreezoMeter	A/I	I	A/I	I		I		A/I	A/I		
Plume	A/I	A/I	A/I	A	A/I	I		A/I	A/I	A/I	A/I
Sh**t! I smoke! smoke		A/I						A/I			
Air Quality	I		I				I	I		I	I
Haze today	I		I					I			
Gobal Air Quality Index	I		I					I	I		
Air Checker	I	I	I		I	I	I	I			I
AirNow			A/I						I		
Air Quality	I	I	I						I		

To focus on the usability of the applications, we reviewed the availability of the three basic components of the air quality applications: Risk/color scale, Map, and Health recommendations. Three mobile applications contained all these three parameters: Air matters, Plume, and BreezoMeter. A brief description of these applications is shown below.

3.1 Air Matters

Air Matters mobile phone application is available in both the Android and iOS platforms and is the most complete in terms of the number of features (11) for both operating systems. This application allows the user to consult the weather and concentrations of atmospheric pollutants in most urban centers around the world, as shown in Fig. 1 for South America. It displays the air quality index and indicates a specific risk/color scale (from green - good (AQI: 0) to purple – bad (AQI: 500)) on the first loaded screen (Fig. 2a). In addition, it shows the concentration values for each criteria contaminant: CO, NO_2, SO_2, $PM_{2.5}$, PM_{10} and O_3. If clicked on each of the atmospheric pollutants, the information about the sources and ranges of risk pop up (U.S. standards) (Fig. 2b). If returned to the main page, this application also provides health recommendations for the conditions in more detail for different activities. For example, recommendations are given in relation to the necessity of purifier use, if it is advisable for children and elders to do outdoor activities, ventilation, mask use and outdoor activities. It forecasts the air quality and displays historical data for hours, days or months (Fig. 2c). Finally, to get to the map, it is necessary to scroll to the bottom of the page (Fig. 2d). In addition, Air Matters allows observing the places with the worst and best AQI and pollutant concentrations. This application also has the peculiarity that it can be connected to a purifier or to a monitoring equipment to check the internal air quality.

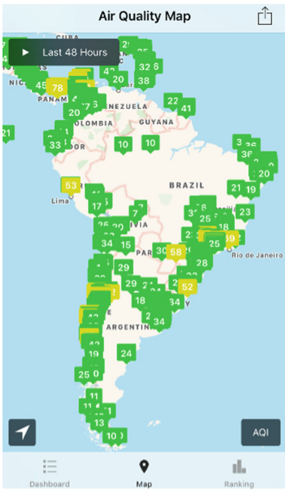


Fig. 1. Map of South American continent AQI's displayed by Air Matters mobile phone application for the last 48 h. Colors represent the severity of AQ conditions (green-good, yellow-elevated pollution).

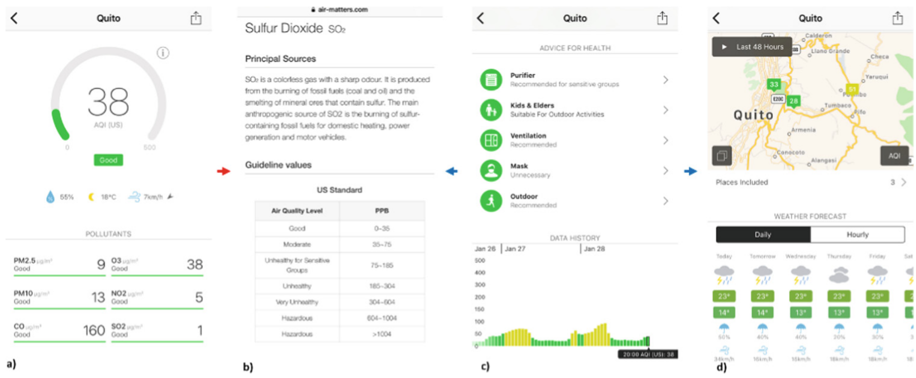


Fig. 2. The screen evolution of Air Matters application: (a) first image of the loaded application displaying the AQI and concentrations of each of the contaminants, general evaluation of AQ conditions is given (good); (b) by clicking on a pollutant of interest the information and risk ranges are provided, (c) in addition to the historical data diagram, health recommendations are given, in the case of returning to the main page and scrolling down; and (d) Map of the location and weather forecast are provided.

3.2 Plume

Plume mobile phone application is also available for both the Android and iOS systems. While it is not as complete in terms of the number of features (total of 9) as the previously discussed Air Matters, it is one of the more dynamic applications. Here in a very visual manner, the important user information is displayed on the first page. As the application is loaded, the first thing that pops up is the air quality index (from blue – clean (AQI: 0) to black – very polluted (AQI: 200)) positioned in the comparative scale of hazard and a relative evaluation of pollution conditions, for example “High” (Fig. 3a). If clicked on the circle, the more detailed information pops up with the AQI for the most common air pollutants: PM_{10}, $PM_{2.5}$ and Ozone, depending on the location (Fig. 3b). Below the risk circle, the temperature, wind speed, UV index and humidity values are displayed as well. At the bottom of the screen an interactive historical (solid line) and forecast (broken line) curve is shown. Finally, the health recommendations for outdoor activities and health problem populations are provided if consulting a few easy graphical icons (Fig. 3a).

To get to the map, the top left corner icon must be selected (Fig. 3a), and the option of world map must be chosen (Fig. 3c). Once the map is loaded, an interpolated world map appears centered on Europe. If this is not the location of your interest, you must scroll to find your region (for example, Ecuador, (Fig. 3d). While, it takes a few extra steps to get to the map, it is one of the very few applications that has an interpolation of the air pollution map. This feature is very important in case there are no monitoring points near the location of your interest. However, there is no guarantee that the interpolation is highly reliable, as usual for this type of air pollution maps [15].

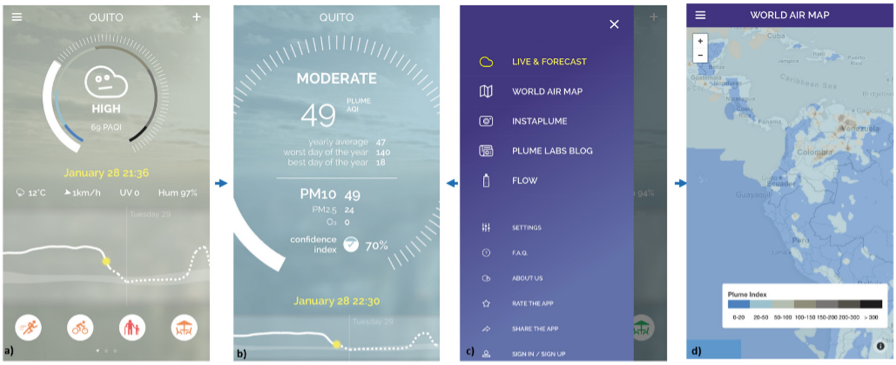


Fig. 3. The screen evolution of Plume application: (a) first image of the loaded application; detailed pollution information image appears if clicked on the pollution circle (b); another click takes you to the main page, where health recommendations can be viewed by consulting any of the action icons on the bottom of the main screen; click on the upper left corner icon takes you to more options window (c); selecting map a user can consult interpolated map of air pollution anywhere in the world (d).

3.3 BreezoMeter

BreezoMeter mobile phone application is also available for both the Android and iOS systems. While it is not as complete in terms of the number of features (4 for Android and 7 for i-Phone) as the previously discussed Air Matters and Plume, it is one of the most logically configured applications. Here all the important user information is displayed on the first page. As you load the application, the first thing that pops up is the air quality index (from green – clean (AQI: 100) to red – very polluted (AQI: 0)) positioned in the comparative scale of risk and a relative evaluation of pollution conditions, for example “Moderate Air Quality” (Fig. 4a). Not only that, but also which pollutant is the most present and responsible for the air pollution conditions at the moment. At the same screen, the recommendations for planning activities outside are given. These recommendations are suggested for: General population, People with heart diseases, People with lung diseases and asthma, Children, Elderly, Active outdoors, and Pregnant women. If you scroll down you can consult pollutant concentrations of each of the criteria pollutants (CO, NO_2, SO_2, $PM_{2.5}$, PM_{10} and O_3) (Fig. 4b), and learn about their sources and specific health issues by clicking on each of the pollutants (Fig. 4c). Right below you can consult the next 6 h of forecast for the AQI and current weather conditions (Fig. 4b). Finally, at the bottom of the main screen the map is displayed, centered on your location, and indicating air quality color in the scale of health risk, intuitively based on traffic light (Fig. 4d).

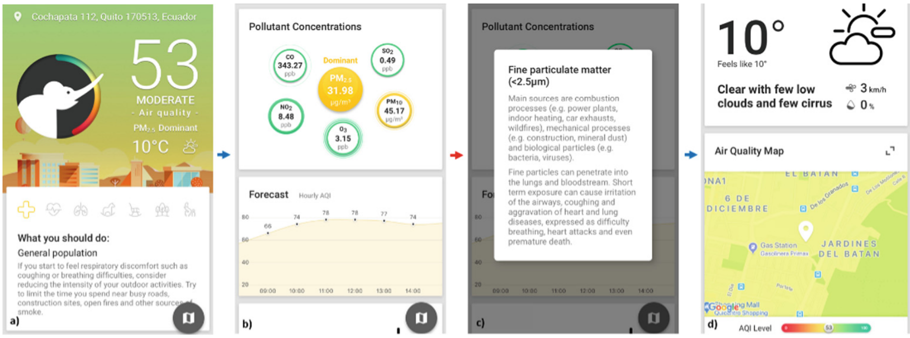


Fig. 4. The screen evolution of BreezoMeter application: (a) first image of the loaded application; (b) scrolled down image of the pollution concentrations and next 6 h forecasting of AQI for the location of interest; if curious for more information on pollutants click to appear the image (c); and at the bottom of the screen if keep scrolling a map of the location of the user is shown (d).

4 Recommendations

Based on the usability evaluation we propose a new user interface combining the three features used to evaluate the applications Air matters, Plume and BreezoMeter. The characteristics were designed taking into account the improvement of the three interaction criteria described at the beginning of Sect. 3. We also incorporate the recommendations suggested by the accessibility study. Figure 4 presents the recommended user interface (Fig. 5).

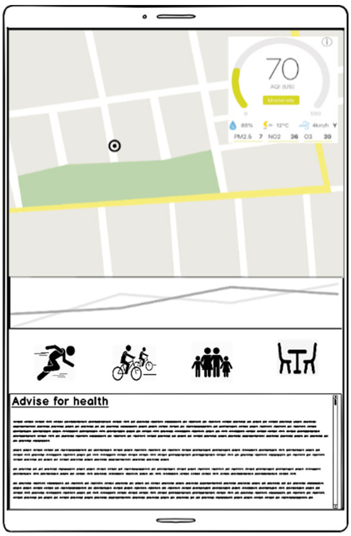


Fig. 5. The recommended user interface.

Results of a previous study using Accessibility Scanner Tool [12], where accessibility assessment was applied to 10 air quality mobile applications, including AirVisual, BreezoMeter, and Plume, revealed that the most of the mobile applications did not achieve an acceptable level of accessibility. It was identified that the most frequent failures are related to "Text Contrast" followed by "Tactile Target". As mobile devices are more likely to be used in a variety of environments, including outdoors, where strong light sources are more likely to occur, this scenario increases the importance of using a good contrast for all users and can aggravate the challenges that low vision users face in accessing low contrast content on mobile devices.

Therefore, we suggest that the contrast (minimum) which requires a contrast of at least 4.5:1 (or 3:1 for large-scale text) and contrast (enhanced) which requires a contrast of at least 7:1 (or 4.5:1 for large-scale text) are used (Fig. 6). This problem is resolved by applying the tool Contrast Ratio Calculator[1]. For example, Fig. 6a shows an example of applying the minimum 3:1 contrast. Where the background has the color #990 and the text #fff, it does not comply with the WCAG 2.1. Figure 6b shows an example of the application of the enhanced 7:1 contrast. Where the background has the color #60f and the text #fff, it complies with WCAG 2.1.

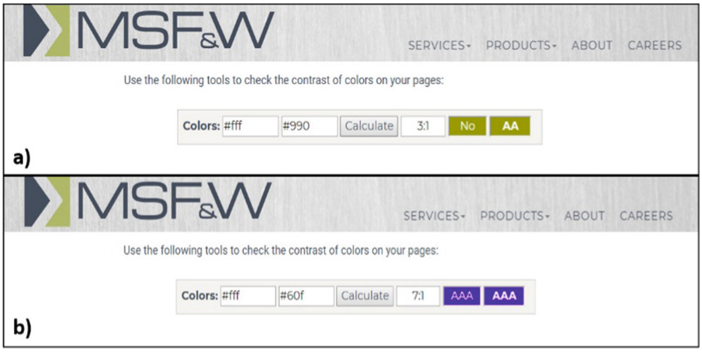


Fig. 6. (a) Contrast (minimum) 3:1; (b) contrast (enhanced) 7:1.

5 Conclusions

More of the applications are freely accessible for Apple (12 applications) if compared to the Android (7 applications) operating system. Even though a number of applications contained more features, in order to focus on the usability of the applications, we reviewed the availability of the three basic components of the air quality applications: Risk/color scale, Map, and Health recommendations. Eventually, retaining three mobile applications – Air matters, Plume, and BreezoMeter – including all three features. In quite a few cases, the user must keep scrolling or clicking to find the essential information, and possibly not finding it due to a number of steps leading to it.

[1] http://www.msfw.com/Services/ContrastRatioCalculator.

Therefore, according to the findings of this study we suggest improvements of the mobile phone applications to provide the key information as soon as the application is loaded. The best way to present the information would be to center the map to the location of the user and allow interactively move around the map of interpolated air pollution map, preferably indicating the risk of the exposure to the pollution by the color scale. The same initial screen should contain the information of the relative risk/color scale of the pollution in terms of danger to the health. The initial page must also contain an instant action recommendation. The latter could be directed to the risk groups such as children and elders, and people that may be performing physical activities. Finally, to assure the accessibility of the mobile application, the interface has to be readable, with enough contrast and fair size text, especially for the health recommendations. Even though there are number of mobile phone applications, not all of them are organized in the most efficient manner, thus the findings of this work may help raise the air quality awareness more inclusively and more effectively.

Acknowledgments. The funding for this research is provided by CEDIA-CEPRA projects: CEPRA XII-2018-13 and is supported by UDLA internal project ERI.WH.18.01.

References

1. World Health Organization: Air pollution levels rising in many of the world's poorest cities. WHO (2017)
2. UNECE: Air pollution and health. https://www.unece.org/environmental-policy/conventions/envlrtapwelcome/cross-sectoral-linkages/air-pollution-and-health.html. Accessed 17 Jan 17 2019
3. WHO: 7 million premature deaths annually linked to air pollution page 1 of 2. WHO: 7 million premature deaths annually linked to air pollution page 2 of 2. http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/#.WqBfue47NRQ.mendeley. Accessed 7 Mar 2018
4. Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature **525**, 367 (2015)
5. European Environment Agency: Air quality in Europe—2018 report (2018)
6. European Environment Agency: Air quality in Europe—2017 report (2017)
7. Zalakeviciute, R., Rybarczyk, Y., López-Villada, J., Diaz Suarez, M.V.: Quantifying decade-long effects of fuel and traffic regulations on urban ambient PM2.5 pollution in a mid-size South American city. Atmos. Pollut. Res. (2017). https://doi.org/10.1016/j.apr.2017.07.001
8. Singla, S.: Air quality friendly route recommendation system. In: Proceedings of the 2018 Workshop on MobiSys 2018 Ph. D. Forum, MobiSys Ph. D. Forum 2018, pp. 9–10. ACM, New York (2018)
9. Zalakeviciute, R., Buenaño, A., Rybarczyk, Y., Sannino, D.: Urban air pollution mapping and traffic intensity: active transport application. In: Air Pollution. InTech (2019)
10. Pérez Medina, J.L., Zalakeviciute, R., Rybarczyk, Y.: Assessing the usability of air quality mobile application. In: 2019 International Conference on eDemocracy & eGovernment (ICEDEG). IEEE (2019)
11. Garzon, S.R., Walther, S., Pang, S., Deva, B., Küpper, A.: Urban air pollution alert service for smart cities. In: Proceedings of the 8th International Conference on the Internet of Things, IOT 2018, pp. 9:1–9:8. ACM, New York (2018)

12. Acosta-Vargas, P., Zalakeviciute, R., Luján-Mora, S., Hernandez, W.: Accessibility evaluation of mobile applications for monitoring air quality, vol. 918, pp. 1–11. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11890-7_61
13. Mobile Accessibility at W3C: Web accessibility initiative (WAI). W3C
14. World Wide Web Consortium (W3C) Web Accessibility Perspectives: Explore the impact and benefits for everyone. Web Accessibility Initiative (WAI). W3C
15. De Mesnard, L.: Pollution models and inverse distance weighting: some critical remarks. Comput. Geosci. **52**, 459–469 (2013). https://doi.org/10.1016/j.cageo.2012.11.002

Heuristic Method of Evaluating Accessibility of Mobile in Selected Applications for Air Quality Monitoring

Patricia Acosta-Vargas[1](✉), Luis Salvador-Ullauri[2], Jorge Luis Pérez-Medina[1], Rasa Zalakeviciute[1], and Wilmar Hernandez[1]

[1] Intelligent and Interactive Systems Lab, Universidad de Las Américas, Quito, Ecuador
{patricia.acosta, jorge.perez.medina, rasa.zalakeviciute, wilmar.hernandez}@udla.edu.ec
[2] Centro de Educación Continua, Escuela Politécnica Nacional, Quito, Ecuador
lsalvador@cec-epn.edu.ec

Abstract. At present, advances in technology, the use of smartphones and access to the Internet pose significant challenges in the area of accessibility of mobile applications. This study identified the lack of clear policies and methods for assessing accessibility in apps as a significant challenge for web accessibility experts.

On the other hand, the heuristic evaluation concept proposed by Nielsen and Molich is an inspection method based on the evaluation of an interactive system. To reduce accessibility barriers, in this research we propose a heuristic method based on Brajnik's Barrier Walkthrough. The method helps to propose some improvements to evaluate accessibility in mobile applications between a heuristic method and the Web Content Accessibility Guide 2.1. As a case study, we applied this evaluation to users with low vision who use mobile applications under the Android operating system that helps users monitor the air we breathe. The results of this research can help future works about the development of accessible mobile applications.

Keywords: Accessibility · Air · Applications · Evaluation · Heuristic method · Monitoring · Mobile · Quality · WCAG 2.1

1 Introduction

Currently, the way society connects and interacts on the web has changed a lot. It was a great challenge the technological advance during the evolution of the computer to the present with smartphones, which have become an excellent alternative to navigate cyberspace and infinite content.

Today, the number of mobile phone users was 5 billion in 2017, but that number will increase to 5.5 billion users in the next five years according to the report from networking technology provider Cisco [1].

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 485–495, 2020.
https://doi.org/10.1007/978-3-030-20040-4_44

First, according to the Convention on the Rights of Persons with Disabilities (CRPD) [2, 3], it establishes the right of access to "information intended for all users with appropriate technologies and accessible formats for different disabilities." As a result, accessibility has become one of the main concerns of our society.

Second, Bourne et al. [4] argue that globally, we can find some form of visual impairment in 1.3 billion people. Concerning farsightedness, 188.5 million individuals show a mild visual impairment, 217 million present variations from moderate to severe visual impairments, and 36 million people are entirely blind [4]. Of which approximately 80% of all vision-related impairments are preventable[1]. According to the authors, most visually impaired people are approximately 50 years of age [4].

Therefore, it can be inferred that users with disabilities are not the only ones excluded by technology. This condition indicates that developers must consider accessibility parameters; during the development of mobile applications [5], when generating a high impact, for the elderly population that is acquiring disabilities related to old age.

Consequently, this study firstly proposes to apply a set of heuristics to achieve accessibility in mobile applications aimed at users with low vision. For which we propose a checklist as a result of a literature review study and the guidelines that are part of this area of knowledge. Second, it suggests using the low vision accessibility checklist. Our research presents a case study of mobile applications related to computer programs that allow us to monitor the kind of air we inhale.

Nielsen and Molich proposed the concept of heuristic evaluation; it is an inspection method based on the evaluation of an interactive system [6]. This research proposes Brajnik's heuristic method based on the Barrier Walkthrough (BW) method for evaluating the accessibility of websites [7].

Therefore, we will use a variation of Brajnik's Barrier Walkthrough[2] method. In which, it is essential to define two concepts, (i) the heuristic method which is an analytical process based on trial and error inspections with several accessibility barriers and (ii) the accessibility barrier which is the limitation that prevents the user from interacting with an application. Accessibility barriers include parameters according to the user, the context of use, the purpose, and the website. From which information [8] is extracted to establish relationships between user satisfaction, effectiveness, and productivity.

The proposed heuristic method for evaluating accessibility in mobile applications has the following phases:

- randomly select applications
- to select the category of users
- identify the scenarios
- evaluate the application with the barrier travel method
- save the results
- analyze the results
- suggest improvements to assessment accessibility

[1] https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] https://users.dimi.uniud.it/~giorgio.brajnik/projects/bw/bw.html.

The rest of this document is structured as follows: Sect. 2 presents the background and related work, Sect. 3 details the case study and method, Sect. 4 presents the discussion and results, and finally, Sect. 5 details the conclusions and future work of this research.

2 Background and Related Work

Currently, the "Web for all" proposed by the World Wide Web Consortium (W3C) is essential knowledge when applying the Web Content Accessibility Guidelines (WCAG 2.0) [9] which were replaced in June 2018 by WCAG 2.1, to develop accessible and inclusive Web sites for all users. Therefore, some studies related to accessibility heuristics applied to mobile devices apply WCAG 2.1 [10] based on best practices.

The accessibility [11] refers to the process of removing barriers that prevent communication and interaction between the mobile application and users. It refers to a design of mobile applications that allow users to understand, navigate, perceive, and interact with the application while providing content.

Web Content Accessibility Guidelines (WCAG) 2.1 [10], consist of four principles, 13 guidelines and 76 compliance criteria.

Perceivable - The first principle argues that people should be able to perceive content visually, sonorously and tactilly [10].

Operable - The second principle of WCAG 2.1 states that users should be able to use and navigate the interface components [10].

Understandable - The third principle indicates that both the content and the interface controls for its management must be understandable to the user [10].

Robust - The fourth principle is that the content must be robust enough to be reliably interpreted by the users of the application, and support work with current and future technologies [10].

There are associated with one of the following levels of compliance:

Level "A" is the minimum accessibility level [12].

Level "AA" is the intermediate accessibility level [12].

Level "AAA" is the maximum accessibility level. A mobile application that reaches level "AAA" is an application that can be accessed by all people [10].

Previous research by several authors related to accessibility in mobile applications was reviewed:

Acosta-Vargas et al. [14–16] applied Accessibility Scanner tool with WCAG 2.1 in mobile applications; results reveal that not all of these mobile applications are accessible. It recommends correcting the barriers identified in order to improve the level of accessibility.

Grellmann et al. [17] argue that WCAG 2.0 is a standard for the evaluation of web accessibility. They propose an adaptation of the standards to the mobile version applied to people with aphasia (PWA). The results describe the effectiveness of the guidelines in the broader context of addressing various disabilities and offer recommendations for the design of users with aphasia.

According to Wilson et al. [18] point out that mobile applications in video games have evolved. However, they present a continuing problem related to accessibility barriers that users face when they play. The authors propose to apply the W3C guidelines for building accessible mobile applications.

The study of Yan and Ramachandran [19] share the analysis of the accessibility status in mobile applications based on a graphical user interface (GUI) study. The authors tested the compliance of the applications with the accessibility guidelines in 479 applications for Android. The authors propose to apply guides and tools to reduce accessibility problems in mobile applications.

In this research, we apply the [23] "Barrier Walkthrough" (BW) method. BW is an analytical technique based on a Brajnik [20] heuristic evaluation. This method consists of the inspection of the mobile application to identify possible accessibility barriers. A barrier is an affectation that prevents people from achieving the objectives when they navigate through a mobile application [21]. Barrier Walkthrough's heuristic method aims for the evaluator to identify scenarios composed by different categories of users, configurations, objectives and possible tasks. Also, it is essential to define the barriers related to a type of user in a specific context regarding efficiency, productivity, safety, and user satisfaction, identifying the degree of severity of each barrier [22].

Brajnik et al. [23] argue that manual accessibility assessment plays a vital role in validating the accessibility of web pages. As a result, the authors applied methods to manually evaluate the accessibility of applications in different categories of users, including those with hearing, visual, cognitive, motor and other disabilities. As a result, both the effectiveness and reliability of the experts are significantly superior to those of non-experts.

3 Method and Case Study

3.1 Case Study

In this research, four mobile air quality monitoring applications running on the Android system were evaluated. Data collection took place on January 6, 2019. Table 1 summarizes the free applications used to assess the kind of air we inhale, contains the identifier assigned to the application, the tool name, the logo, the update date, the current version, the company responsible for the application and the number of downloads.

3.2 Method

Figure 1 contains a summary of the heuristic method to evaluate accessibility in mobile applications.

Phase 1: Select the applications to evaluate; in this first phase we select four mobile applications for air quality monitoring. The number of discharges was an essential indicator in the selection of mobile applications. Table 1 contains the information of the mobile applications evaluated with the heuristic method.

Phase 2: Select the type of users, in this second phase, is essential to select the category of users, in this case, applied to users with low vision. This phase is vital because it defines the category of users that use the Android applications selected in

Table 1. Mobile applications evaluated with the heuristic method

Id	Tool	Logo	Updated	Offered by	Download
A	AirVisual		June 1, 2018	AirVisual	**1.000.000+**
B	Plume Air Report		July 9, 2018	Plume Labs	**100.000+**
C	Air Quality Index BreezoMeter		July 11, 2018	BreezoMeter	**100.000+**
D	Air! World Air Quality		June 18, 2016	Inside	**100.000+**

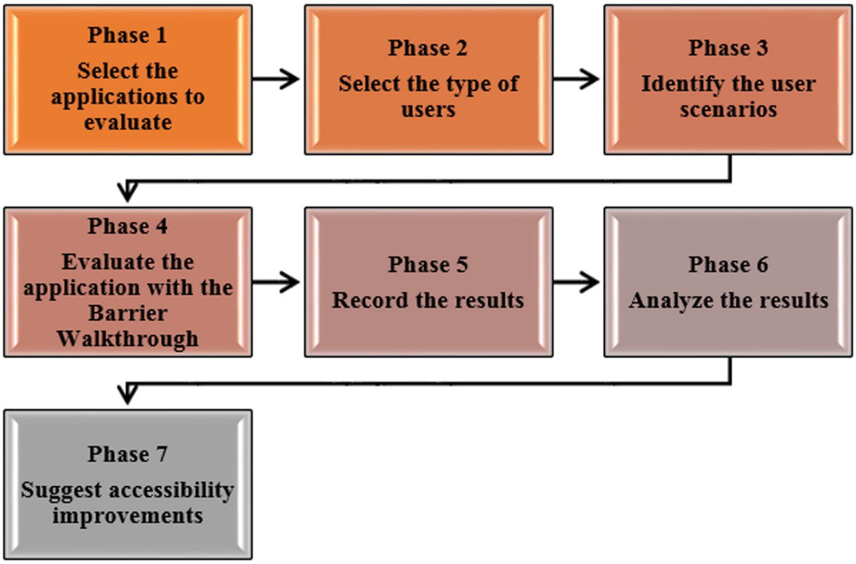


Fig. 1. A heuristic method for accessibility assessment

phase 1. Application functions include parameters such as font size increase, color polarity, contrast levels, and reduced screen resolution.

Phase 3: Identify the user scenarios, in this third phase, the context of the evaluation is defined, we applied this case study only to the first interface of each mobile application with which the user interacts. Table 2 contains the principles of WCAG 2.1 that are related to each of the 18 barriers for low vision users.

Phase 4: Evaluate the application with the Barrier Walkthrough (BW), In this fourth phase, we apply the Barrier Walkthrough method proposed by Brajnik [20] with the modification indicated above by the authors. In such a way that for cases that do not present serious problems, we record the value of zero. Finally, when the problem is

Table 2. Barriers for low vision users

Id	WCAG 2.1 (principle)	Barrier type
1	Operable	Functional images (background) [20]
2	Operable	Functional images (lacking text) [20]
3	Operable	Dynamic menus in JavaScript [20]
4	Operable	Internal links are missing [20]
5	Perceivable	Rich images (poorly positioned)
6	Perceivable	Rich images (background)
7	Perceivable	Color is necessary [20]
8	Perceivable	Insufficient visual contrast [20]
9	Perceivable	Moving content [20]
10	Perceivable	Too long tooltips
11	Perceivable	Missing layout clues
12	Perceivable	Dynamic changes
13	Robust	Text cannot be resized [20]
14	Robust	Inflexible page layout [20]
15	Robust	Skip links not implemented [20]
16	Understandable	Too long lines of text [20]
17	Understandable	Too many links [20]
18	Understandable	Image used as titles

critical, we give the value of 3. Table 3 contains information on the values for calculating the severity score of each barrier.

Table 3. Values for calculating the severity

Impact	Persistence	Severity
0	0	Non-severity [20]
1	1	Minor [20]
1	2	Minor [20]
1	3	Significant [20]
2	1	Significant [20]
2	2	Significant [20]
2	3	Critical [20]
3	1	Critical [20]
3	2	Critical [20]
3	3	Critical [20]

Phase 5: Record the results, in this phase records the results of the heuristic evaluation of AirVisual, Plume Air Report, Air Quality Index BreezoMeter, and Air World Air Quality. Table 4 contains the evaluation values, recording the impact and persistence of each of the 18 barriers related to low vision users. To find the degree of severity, the values recorded in Table 3 apply.

Table 4. Heuristic evaluation of mobile applications

Barrier type	AirVisual			Plume Air			BreezoMeter			Air Quality		
	I	P	S	I	P	S	I	P	S	I	P	S
Functional images embedded	1	1	m	1	1	m	0	0	N-s	0	0	N-s
Functional images lacking text	2	3	c	2	2	s	2	2	s	2	2	s
Dynamic menus in JavaScript	1	1	m	1	1	m	1	1	m	1	1	m
Internal links are missing	1	1	m	1	1	m	1	1	m	1	1	m
Rich images (poorly positioned)	2	2	s	2	2	s	1	1	m	1	1	m
Rich images (background)	1	1	m	1	1	m	1	1	m	1	1	m
Color is necessary	2	2	s	2	2	s	2	2	s	2	2	s
Insufficient visual contrast	3	3	c	2	3	c	2	2	s	3	3	c
Moving content	1	1	m	1	1	m	1	1	m	1	1	m
Too long tooltips	1	2	m	1	1	m	1	1	m	1	1	m
Missing layout clues	1	1	m	1	1	m	1	1	m	1	1	m
Dynamic changes	1	1	m	1	1	m	1	1	m	1	1	m
Text cannot be resized	2	2	s	2	2	s	1	1	m	1	1	m
Inflexible page layout	1	2	m	1	2	m	1	1	m	1	1	m
Skip links not implemented	1	2	m	1	1	m	1	1	m	1	1	m
Too long lines of text	2	2	s	1	1	m	1	1	m	2	2	s
Too many links	1	1	m	1	1	m	1	1	m	1	1	m
Images used as titles	2	1	s	0	0	N-s	0	0	N-s	0	0	N-s

I=Impact, P=Persistence, S=Severity, N-s=Non-severity, m=minor, s=significant, c=critical

Phase 6: Analyze the results; at this phase, the most severe tool is AirVisual, followed by Plume Air and BreezoMeter. Table 5 contains the barriers identified in the heuristic evaluation, as well as the degree of severity identified. We have detailed further analysis of the results and discussion section.

Table 5. Types of barriers identified in the heuristic evaluation

Barrier type	AirVisual	Plume Air	BreezoMeter	Air Quality
Functional images lacking text	critical	significant	significant	significant
Rich images that are poorly positioned	significant	significant	minor	minor
Color is necessary	significant	significant	significant	significant
Insufficient visual contrast	critical	critical	significant	critical
Text cannot be resized	significant	significant	minor	minor
Too long lines of text	significant	minor	minor	significant
Images used as titles	significant	Non-severity	Non-severity	Non-severity

Phase 7: Suggest accessibility improvements, in this phase the barrier faced by the user is identified to suggest a possible improvement. Table 6 presents the principle related to WCAG 2.1, the type of barrier and the improvement that can be applied to solve the problem of users with low vision.

Table 6. Barriers and possible improvements

Principle	WCAG 2.1	Barrier type	Problem
Operable	1.1, 1.1.1, 1.4, 1.4.5	Functional images (lacking text)	Include the attribute ALT. Informative text that allows the user to understand the effects
Perceivable	1.3, 1.3.1, 1.3.2	Rich images that (badly positioned)	Reduction of effectiveness
Perceivable	1.4: 1.4.1	Color is necessary	Reduction of effectiveness
Perceivable	1.4: 1.4.3, 1.4.6	Insufficient visual contrast	Reduction of effectiveness
Robust	1.4, 1.4.4	Text cannot be resized	Reduction of productivity and satisfaction
Understandable		Too long lines of text	Reduction of productivity and effectiveness
Understandable	1.4, 1.4.5	Images used as titles	Reduction of effectiveness

4 Results and Discussion

Figure 2 presents the severities identified for the mobile applications evaluated, and it is observed that the application with the highest number of severities of the "critical" type corresponds to AirVisual, the applications with "significant" severities correspond to AirVisual and Plum Air. Among the four applications evaluated, the one with the lowest number of severities is BreezoMeter, although not in its entirety with the guidelines for acceptable accessibility. The data set and analysis is available in Mendeley Data[3].

From the data recorded in Table 6, we can observe that 55.2% presents a severity of the "significant" type, 20.7% presents a severity of the "minor" type, 13.8% presents a severity of the "critical" type; finally, 10.3% do not present serious problems.

From the analysis obtained, it is concluded that the principle that is most violated is "perceivable" with 42.9%, followed by "understandable" with 28.6%, "operable" and finally, "robust" with 14.3%.

From results it can be concluded that: (i) the Brezzo Meter tool is the most accessible, it does not present "critical" severity, which means that users can access the mobile application with fewer problems. (ii) not all users can perceive the content in a

[3] https://data.mendeley.com/datasets/kxnsjkddwv/4.

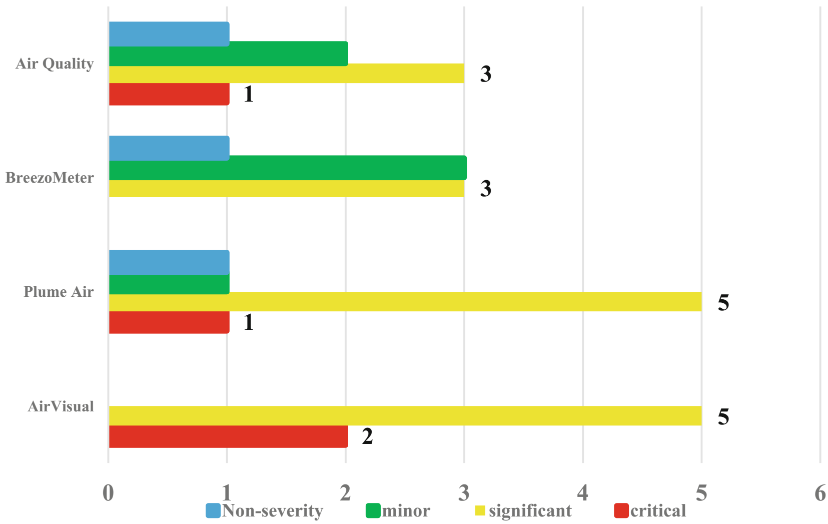


Fig. 2. Severities identified in mobile applications

visual. (iii) users may have problems using and interacting with interface components. (iv) the content and interface controls for the management of mobile applications are not completely understandable by the user, and (v) assistive technologies can not interpret the content because of the lack of robustness.

5 Conclusions and Future Works

The inclusion of low vision users can be achieved by reducing the number of barriers identified in mobile applications, which must be corrected before the application goes into production.

The results of the evaluated mobile applications reveal that the applications do not reach an acceptable level of accessibility. Therefore, we suggest correcting errors that affect users with low vision in order to reach the "AA" level of accessibility recommended by the W3C. In order to solve the "operable" principle with severity "critical," it is suggested to avoid using empty ALT for clickable images. It is important that functional images without text include the "ALT" attribute and write the text in a way that allows the user to understand the effects of link activation. About the "perceivable" principle, with the types of severity "critical" and, "significant," the problem of rich images that are poorly positioned, color and insufficient visual contrast, produce a reduction in effectiveness. Therefore, it is recommended to change the design of the application so that it does not provide erroneous visual clues. We recommend using redundant means to differentiate information elements when applying typographical conventions that do not only set colors.

On the other hand, it is suggested that any background image is removed or changed so that it never interferes with the perception and interpretation of foreground content. For the "robust" principle, with the severity of the "significant" type, with the problem of not being able to change the size of the text, which causes a reduction in productivity and user satisfaction. It is recommended to specify the font size, always use the CSS properties. To give a solution to the "Understandable" principle, with a "significant" severity, related to too long lines of text, or with images as titles. We suggest implementing a design so that the text automatically adapts to the width of the window.

In the future, it is suggested to combine the automatic and heuristic method to achieve a complete accessibility assessment. The evaluation method applied can serve as a reference in future scenarios for mobile applications, taking into account the type of barrier and accessibility for users with different disabilities.

Acknowledgments. The researchers thank Universidad de Las Américas (Ecuador) for funding this study through the project "CEPRA XII-2018-13".

References

1. Cisco: Cisco Visual Networking Index: Global mobile data traffic forecast update, 2017–2022 white paper. Cisco. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html
2. ESCAP: Disability at a glance 2015: strengthening the evidence base in Asia and the Pacific (2015). https://doi.org/10.1787/eag-2015-en
3. World Health Organization (WHO): United Nations expert group meeting on building inclusive society and development through promoting ICT accessibility: emerging issues and trends (2012)
4. Bourne, R.R., Flaxman, S.R., Braithwaite, T., Cicinelli, M.V., Das, A., Jonas, J.B., Naidoo, K.: Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob. Health **5**(9), e888–e897 (2017). https://doi.org/10.1016/s2214-109X(17)30293-0
5. Díaz-Bossini, J.M., Moreno, L.: Accessibility to mobile interfaces for older people. Procedia Comput. Sci. **27**, 57–66 (2013). https://doi.org/10.1016/j.procs.2014.02.008
6. Nielsen, J., Rolf, M.: Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 249–256 (1990)
7. Brajnik, G., Yesilada, Y., Harper, S.: Guideline aggregation: web accessibility evaluation for older users. In: International Cross-Disciplinary Conference on Web Accessibility, pp. 127–135 (2009). https://doi.org/10.1145/1535654.1535686
8. Yesilada, Y., Brajnik, G., Harper, S.: Barriers common to mobile and disabled web users. Interact. Comput. **23**, 525–542 (2011). https://doi.org/10.1016/j.intcom.2011.05.005
9. World Wide Web Consortium (W3C): Web content accessibility guidelines (WCAG) 2.0. https://www.w3.org/TR/WCAG20/
10. World Wide Web Consortium (W3C): Web content accessibility guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/
11. Acosta-Vargas, P.: Framework for accessibility evaluation of hospital websites. In: International Conference on eDemocracy & eGovernment (ICEDEG), pp. 9–15. IEEE (2018)

12. Acosta-Vargas, P., Lujan-Mora, S., Salvador-Ullauri, L.: Web accessibility polices of higher education institutions. In: 2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET), pp. 1–7 (2017)
13. World Wide Web Consortium (W3C): Introduction to Web Accessibility—Web Accessibility Initiative (WAI). https://www.w3.org/WAI/fundamentals/accessibility-intro/
14. Acosta-Vargas, P., Zalakeviciute, R., Luján-Mora, S., Hernandez, W.: Accessibility evaluation of mobile applications for monitoring air quality, pp. 1–11. Springer, Cham (2019)
15. Acosta-Vargas, P., Luján-Mora, S., Acosta, T., Salvador-Ullauri, L.: Toward a combined method for evaluation of web accessibility. In: Advances in Intelligent Systems and Computing, pp. 602–613 (2018)
16. Acosta-Vargas, P., Acosta, T., Luján-mora, S.: Challenges to assess accessibility in higher education websites: a comparative study of Latin America Universities. IEEE Access **6**, 36500–36508 (2018). https://doi.org/10.1109/ACCESS.2018.2848978
17. Grellmann, B., Neate, T., Roper, A., Wilson, S., Marshall, J.: Investigating mobile accessibility guidance for people with aphasia. In: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 410–413. ACM (2018)
18. Wilson, A., Crabb, M.: W3C accessibility guidelines for mobile games. Comput. Games J. **7**, 49–61 (2018). https://doi.org/10.1007/s40869-018-0058-7
19. Yan, S., Ramachandran, P.G.: The current status of accessibility in mobile apps. ACM Trans. Access. Comput. **12**, 1–31 (2019). https://doi.org/10.1145/3300176
20. Brajnik, G.: Barrier Walkthrough. https://users.dimi.uniud.it/~giorgio.brajnik/projects/bw/bw.html
21. Brajnik, G.: Ranking websites through prioritized web accessibility barriers. In: Technology and Persons with Disabilities Conference, pp. 1–8 (2007). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.79.6853&rep=rep1&type=pdf
22. Ferreira, C., Ferreira, S., Sacramento, C.: Mobile application accessibility in the context of visually impaired users. In: Symposium on Human Factors in Computing Systems, p. 32. ACM (2018)
23. Yesilada, Y., Brajnik, G., Harper, S.: How much does expertise matter? In: Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2009, p. 203 (2009). https://doi.org/10.1145/1639642.1639678

Remote Monitoring Air Quality in Dangerous Environments for Human Activities

David Pozo[1(✉)], Santiago Solórzano[1], Jorge-Luis Pérez-Medina[2], Kevin Jaramillo[1], Ricardo López[1], and Rasa Zalakeviciute[2]

[1] Universidad de Las Américas (UDLA), Quito, Ecuador
{david.pozo, santiago.solorzano, kevin.jaramillo, ricardo.lopez.molina}@udla.edu.ec
[2] Intelligent and Interactive Systems Lab (SI2 Lab), Universidad de Las Américas (UDLA), Quito, Ecuador
{jorge.perez.medina, rasa.zalakeviciute}@udla.edu.ec

Abstract. This work describes the architecture details of a remote- controlled mobile robot prototype that allows to monitor the real-time concentrations of Sulfur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), tropospheric Ozone (O_3), carbon dioxide (CO_2) and methane (CH_4) that could seriously affect human's health after a violent natural or human caused phenomena. Parameters such as temperature and humidity will also be monitored in order to understand the conditions of the people to be rescued from dangerous environments. In order to establish the air quality inside a collapsed structure after a disaster and how it could affect the health of rescuers and the people at risk, the measurements will be displayed inside a user-friendly interface which will be able to generate real-time graphics of these parameters and the possible consequences of being exposed to high concentrations of pollution or heat. Additionally, a task model is provided to describe the interaction among the system and the final users. The data acquired by the robot are transmitted through a point to point network based on elements of low consumption and cost. These elements let the robot transmit the information to a computer, which through the graphic user- friendly interface will indicate the parameters of air quality in a friendly way for the end user.

Keywords: Remote · Monitoring · Robotic · Platform · Air quality · Collapsed structures · User interface

1 Introduction

Every day the world's population is threatened by disasters that can be caused by violent phenomena of natural or anthropogenic origin. Earthquakes, hurricanes, landslides or collapse of structures can cause serious material damage and mainly injuries to the physical integrity of people. In all these events, emergency personnel must be able to respond promptly in order to safeguard the lives of people at risk, however, this can put their own lives in danger. For instance, a hurried entry to a disaster zone can cause the loss of lives among the rescuers, especially due to the lack of knowledge of the environment with which they will encounter. In the case of collapsed structures, the

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 496–507, 2020.
https://doi.org/10.1007/978-3-030-20040-4_45

risk of inhalation of toxic substances that may affect the health of the rescuers must be considered. In addition, it must be taken into account that a previous analysis of environmental characteristics can help make an evaluation of rescue strategies and times, which can be crucial when safeguarding the lives of people at risk.

The World Health Organization (WHO) considers that Sulfur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO) and tropospheric Ozone (O_3) and particulate matter (PM) are dangerous for human health when its concentrations are above of established safe levels. For example, (SO_2) above 30 ppm can cause death after 10 min of exposure; while (NO_2) above 500 ppm is a lethal dose due to the developing lung edema [1, 2]. CO is a common silent killer in domestic gas leaks, concentrations above 800 ppm which initially may cause dizziness, nausea, within 45 min cause convulsions and within 2 h leave a person insensible. Meanwhile CO concentrations twice as high will result in death within 2 h due to tissue hypoxia and organ failure [3]. However, in a mild to heavy smoke CO concentrations may reach from 2000 ppm to 10000 ppm, respectively, which may cause death within a few minutes in the worst-case scenario. As for particulate matter, smoke inhalation is the principal cause of death in fires by asphyxiation (choking) and lack of oxygen. Moreover, the carbon dioxide (CO_2) methane (CH_4) could affect human health quite considerably, as increased CO_2 concentrations increase respiratory distress. Also, CH_4 and CO_2 are known greenhouse gases that may absorb the infrared radiation and further increase temperatures. Additionally, parameters such as temperature and humidity should be reviewed in order to establish the health conditions of the people to be rescued. While the comfort level is defined by a very narrow window of variation in temperature and relative humidity, extreme heat and humidity may expose a person to dangerous conditions, mostly due to the overheating, causing heat cramps, heat exhaustion or even heat stroke, in which case, unless treated promptly, it is a life-threatening emergency (coma, stroke).

The remainder of this manuscript is organized as follows. Next section introduces several related works and background. Third section introduces the main functionalities of air quality robotic platform. Fourth section presents details of the user-friendly interface and data transmission architecture. Section five discusses a preliminary experiment. Finally, section six shows conclusions and future work.

2 Related Work

Natural disasters could seriously affect the local infrastructure of a city and could be really critical when it affects industrial zones where harmful pollutants could be released. In the case of hurricane Harvey in 2018, the air quality measurements were important during and after the natural disaster in order to establish the air pollutants which people faced [4]. For example, disasters like hurricanes influence the air quality due to the damages that they provoke; or the effects accumulated over three years of seismic activity in Canterbury, New Zealand. In the latter case, the liquefaction process released high concentrations of air pollutants that along with the dust, generated from building destruction induced by earthquakes, caused damages for people's health [5].

On the other hand, the Health - Related Quality of Live parameter (HRQoL) has been analyzed in order to establish the physical and psychological impact on the rescuers after a disaster, which showed that the risk of exposure to a potential toxic substance and the previous preparation to a specific situation could affect the HRQoL of the emergency personal in long term [6, 7].

A series of investigations related to air pollution monitoring systems have been done, especially in urban areas establishing parameters of measurements, schemes, case studies and sensor technologies [8–10]. The wireless network for a real time air quality monitoring system is one of the best options because of its portability, fast deployment and low cost. The reliability of these monitoring networks depend of the sensors to be mounted and the specific application to be used [11, 12]. A natural disaster could cause damages in local network monitoring systems. This was the case of Puerto Rico, where after the Hurricane Maria the main air monitoring network suffered damages. Based on previous lessons, a case study was presented in order to deploy a low cost air quality system based on Real Time Affordable Multi-Pollutant (RAMP) and a black carbon (BC) monitoring system [13].

Finally, the way how the air quality parameters are visualized over a human-machine interface, is important as it can help the final users understand and keep attention on the data that is being visualized. Some studies have been developed in order to evaluate air quality interfaces. A previous case study worth mentioning, is related to the development of inclusive mobile application based on an accessibility test [14]. In [15] a model of a mobile interface for air quality monitoring is presented, which provides air pollution data in the whole city of Quito, Ecuador, and gives a series of recommendations to develop an interface and a method to interpolate measurements.

3 The Air Quality Sensor Monitoring Application

The Air Quality Sensor Monitor (AQSM) application is a solution for air quality monitoring in hazardous environments for people. The global vision of the application is to offer a mobile robotic platform and software for visualization to support the monitoring of pollution parameters in conditions of risk to human health (Fig. 1).

3.1 Hardware Architecture

The AQSM Hardware architecture consist of 3 subsystems: Mobile Robotic platform, Platform Movement Control and Receiver System. All these parts let the final user have a total control over the robot and let him see the data acquired over the main user interface. The general scheme of the hardware architecture can be seen in Fig. 2.

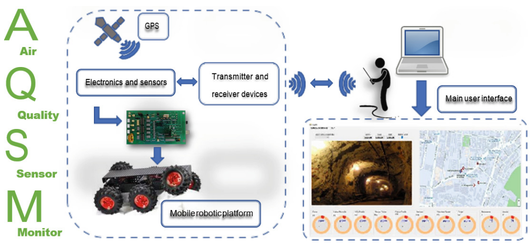


Fig. 1. General scheme of the AQSM

Mobile Robotic Platform. The mobile robotic platform prototype works thanks to a commercial structure consisting of six motors over which the electronic and sensing air system is implemented. The motors are controlled by the use of a radio control receiver with a motor driver device. The electronic system allows to use the Global Positioning System (GPS) in order to get the location of the mobile robot. Also, it is possible to acquire the data incoming from 9 sensors that provide information about environmental air conditions. Some of these sensors require the use of a vacuum pump in order to get the sufficient quantity of air to perform the measurements. All these acquired parameters are embedded and sent by a transmitter system through a 2.4 GHz band and power of 60 mW based on a Xbee module.

In order to recognize the environment where the robot is moving around, a video camera system of 5,8 GHz with a power of 2000 mW system is used. Moreover, because the prototype must function in dark areas, it is provided with a light system that is controlled through the main user interface.

The energy system is spread in two stages. The first one dedicated to the electronic and air sensing system while the second one is necessary for the use of the elements that consume great amount of energy such as motors, vacuum pump, lights and motor controllers.

Platform Movement Control and Receiver System. The end user must be capable to control the robot in agile and effective way in order to position it in a desired location. To accomplish this objective, the operator uses a commercial radio-frecuency control with a wide operation range so it can command over large distances. This is a specialized device that is used commonly in applications like unmanned aerial vehicle (UAV) control or hobby activities, thanks to robustness and fiability. In order to accomplish missions, the movement control must be combined with the receiver system that provides the user with information about the GPS position, path and terrain conditions through the transmitted video signal from the mobile robot. The receiver system works with two principal elements. The first one is specialized in data acquisition, like GPS information or sensor measurements by the use of a Xbee receiver. While the second one is used with a video real time receiver system. Both are connected to the PC main user interface by USB ports.

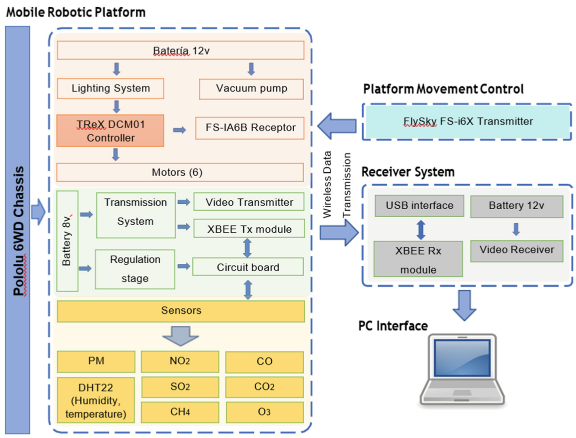


Fig. 2. General hardware architecture scheme of the mobile robot platform

3.2 AQSM User Interface

Figure 3(a) shows the main user interface of AQSM. Basically, the user screen is composed of five key areas. In order to give the reader a good understanding of AQSM, we describe the main features of the application.

We adopt the *User-Centered agile Development (UCD)* approach [16] to specify and design the concepts and the main functionalities to create therapeutic programs (TPs) with the platform in a flexible and efficient manner. This process is described by using the task tree model. It was inspired by the *Concur Task Trees (CTT)* notation described in [17]. We have adopted this representation because it supports the design of interactive applications for user interface model-based design. It generates a tree graphical representation of the hierarchical decomposition of the tasks provided by the platform. Figure 4 shows the tasks of the application.

3.3 Application's Tasks

The acquisition process of air pollution and meteorological parameters is initiated by the tasks of connecting to a mobile robotic platform. To start the connection, the operator must assign a connection port. Once the connection is established, the operator can start capturing the data.

Once the connection with the mobile robotic platform has begun, the operator can visualize the exact location where the device is located. The operator can modify the parameters map display, change its orientation, zoom level, etc. through the interface. Also, the operator can observe the level of autonomy of the mobile robotic platform at all times so the platform can be recovered in case of a drop in autonomy levels.

Once the process of capturing the pollution values has started, the operator disposes of a visualization area of the image captured by the mobile robotic platform. This image depends on the orientation and/or location of the camera built into the mobile robotic platform.

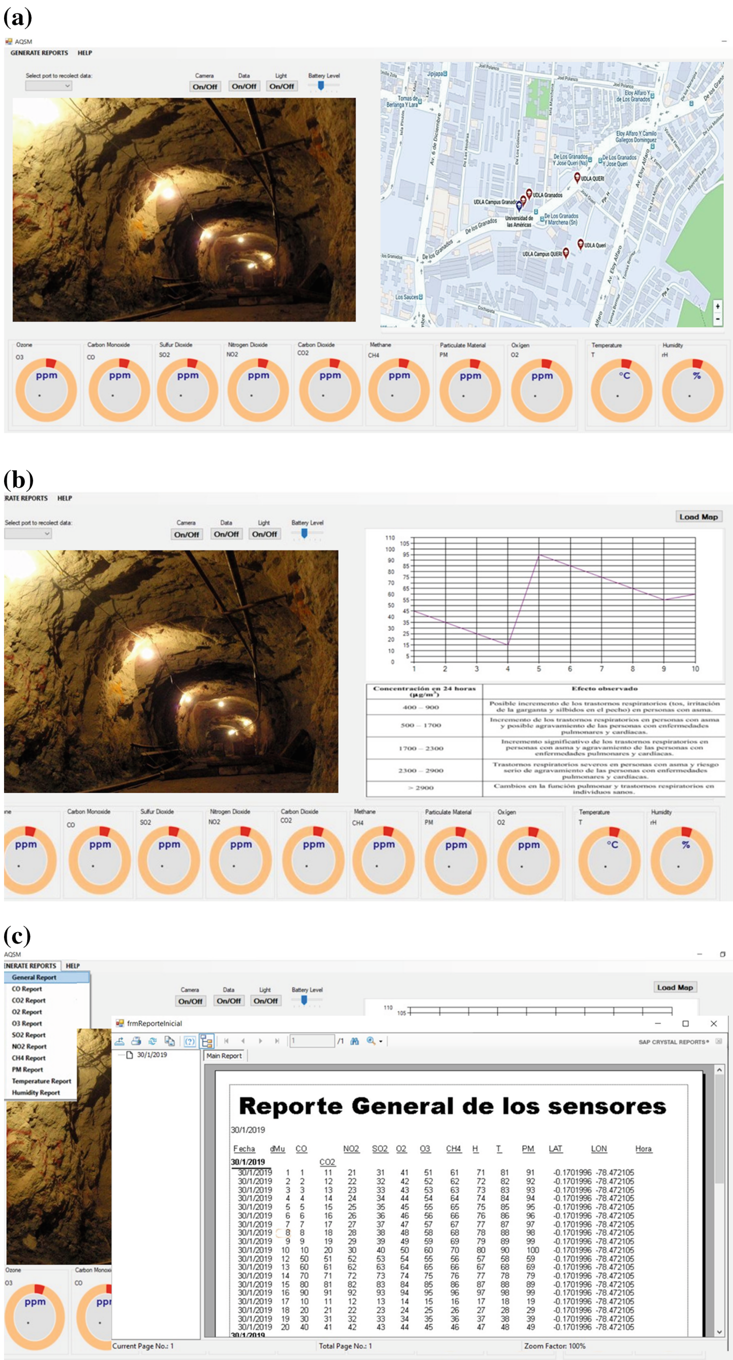


Fig. 3. Interface of AQSM. (a) Main user interface of AQSM system. (b) Parameters' information. (c) Parameters' report.

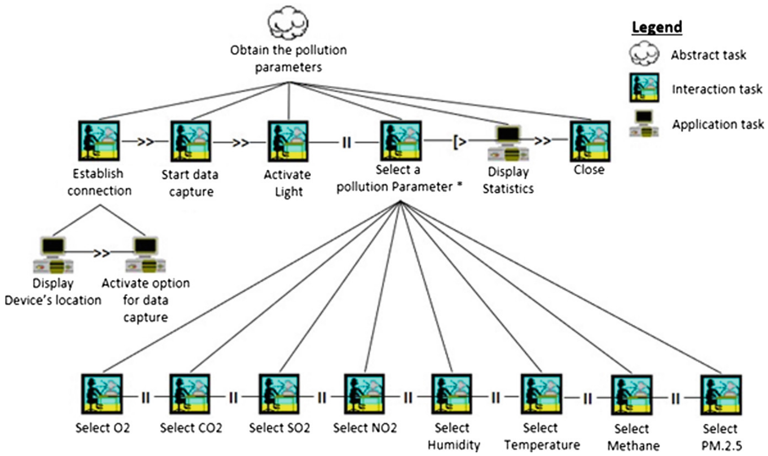


Fig. 4. The application's tasks.

The operator can activate or deactivate the obtaining of the images of the camera. Also, manipulate the activation or deactivation of a light to support the required visualization in dark spaces.

The system shows the pollution and meteorological parameters. For each pollution value the operator may be informed of warnings by change in the color of each parameter. The operator may consult the details of any pollution parameter at any time (see Fig. 3(b)). From a selection the system shows a descriptive graph, associated with a scale of interpretation. Finally, the operator can see the history of the statistics captured by the device (see Fig. 3(c)). This report can be generated from the all sensors or individually. Once the report is generated it has a series of functions, like print onto PDF and filter the informations by dates.

4 Preliminary Experiment

4.1 The Plan

We performed a heuristic evaluation in order to detecting usability concerns. Four experts tested the mobile application in order to find the difficulties end-users might have when using the application. Each expert received a user manual for the application, including the evaluation tasks, the set of usability heuristics (according to [18]), a series of steps to perform the evaluation, and an evaluation template. The usability experts used the set of 14 heuristics presented in Table 2. Usability problems are evaluated based on the severity listed in Table 1. The levels of severity range from 1 to 5. The value of 1 represents the lowest level of severity while a value of 5 represents the most serious problem of usability.

Table 1. The list of severities

Reference	Description
1	A usability problem is not considered, in its entirety
2	Barely aesthetic problem: it does not need to be modified, unless there is time available
3	Less usability problem: the solution to this problem should have low priority
4	Major usability problem: it is important to solve it, for this high priority must be given
5	Usability catastrophe: it is mandatory to resolve it, before the product is disclosed

The task to be evaluated was the consultation of the information on a contamination parameter present in the AQSM system. Each heuristic presented in Table 2 was used to explain and comment the usability problems found by experts when analyzing the task to be evaluated. The evaluation procedure was as follows:

1. Each expert became familiar with the application.
2. Each expert made an individual evaluation considering the set of heuristics evaluation and the severity's levels.
3. The results of each expert were merged for further analysis.

4.2 The Results

The number of problems detected by each expert varied between 6 and 13. Table 2 shows the collaborative consolidation of usability problems, grouped by categories and severities. Overall, a total of 37 usability concerns where found. Most of the usability problems detected are related to:

1. There is no help option and documentation for the user.
2. The user does not have a guide that identifies where it is when selecting a component of the main interface.
3. There are many elements that can confuse the user.
4. The information in the recommendation table presents a lot of information. In it, a vocabulary based on iconography could be used.
5. The contamination parameter selected by the user is not highlighted.
6. It is not indicated what is the pollution value belonging to the statistics.
7. It is not allowed to filter the information in the sensor report. For example, the operator may need to indicate a range of dates.

Table 2. Usability problems and severity for all tasks

Usability heuristics			Severity					Total Prob.
Group	ID	Description	1	2	3	4	5	
User guidance	H1	Prompting	0	2	1	0	1	4
	H2	Feedback	0	0	3	0	1	4
	H3	Information architecture	2	1	1	0	0	2
	H4	Grouping/distinction	2	0	1	1	0	2
Total - user guidance			4	3	6	1	2	12
User effort	H5	Consistency	2	1	1	0	0	2
	H6	Cognitive workload	0	2	0	2	0	4
	H7	Minimal actions	3	1	0	0	0	1
Total - user effort			5	4	1	2	0	7
User control and freedom	H8	Explicit user actions	3	0	1	0	0	1
	H9	User control	2	1	0	0	1	2
	H10	Flexibility	0	3	1	0	0	4
Total - user control and freedom			5	4	2	0	1	7
User support	H11	Compatibility with the user	2	2	0	0	0	2
	H12	Task guidance and support	0	0	2	1	1	4
	H13	Error management	2	0	1	0	1	2
	H14	Help and documentation	1	0	2	0	1	3
Total - user support			5	2	5	1	3	11
Totals			19	13	14	4	6	37

Table 2 shows the usability problems per usability heuristics. Most of the usability problems found are related to the *"user guidance"* (which 2 catastrophic and 6 minor usability problems), *"user support"* (with 3 catastrophic, 1 mayor and 5 minor usability problems). User support and are mainly related to all heuristics. The user interfaces does not have help and documentation. Also, error messages are non-existent. User guide problems are mainly related to the fact that it is difficult to have the feed-back of the user tasks and the does not help the user to understand the result of their task. These problems are aligned with user support. This is mainly because the system does not offer a manner to identify when a contaminant parameter is selected to visualize their statistics and recommendations. The problems obtained in the category of user effort suggests that the user interface have many elements that can confuse the user.

4.3 Discussion

Based on the usability problems detected, a final main user interface version was developed (see Fig. 5). The actions taken to correct the principal concerns found by the test are presented in Table 3.

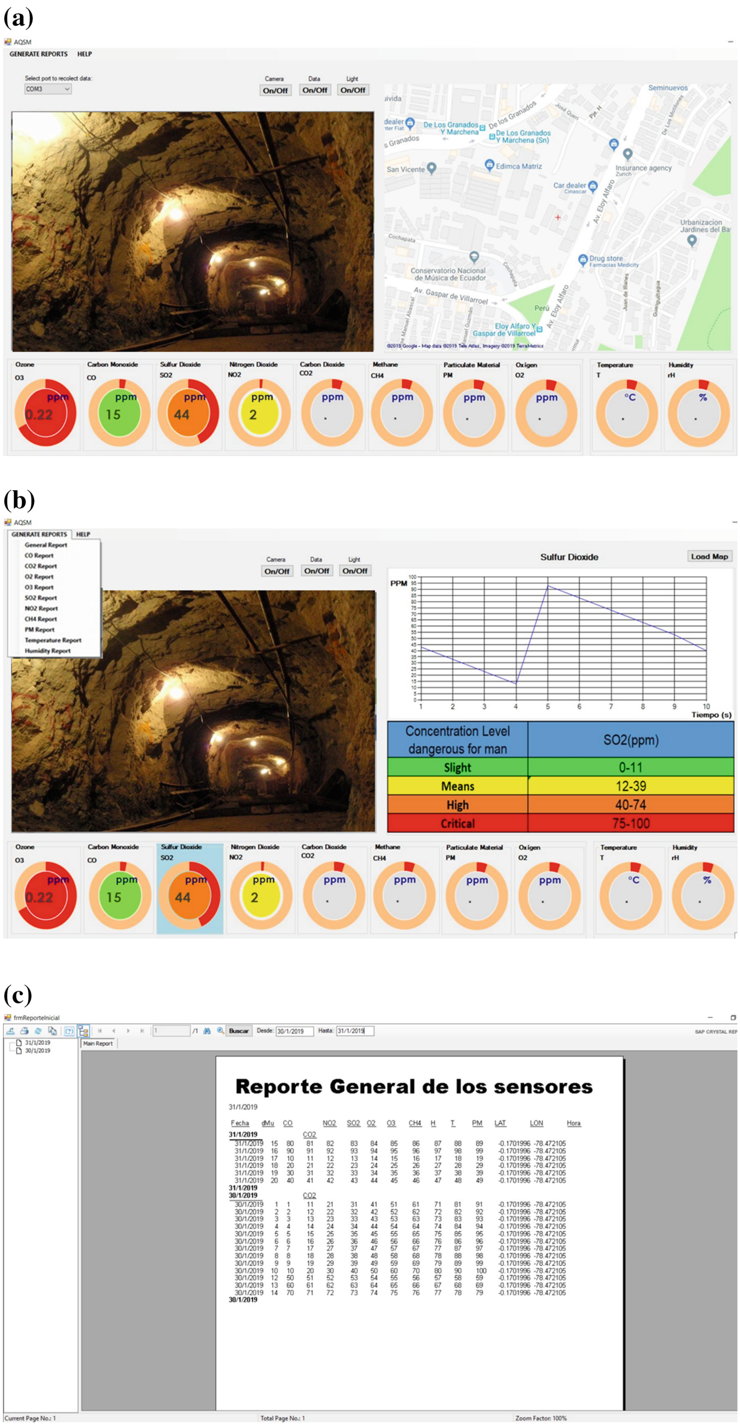


Fig. 5. Interface of AQSM improved. (a) Main user interface of AQSM system. (b) Parameters' information. (c) Parameters' report.

Table 3. The list of concerns and actions

Concern	Solution
There is no help option and documentation for the user	There was a help button, but maybe it couldn't have been seen by its size by the tested user
The user does not have a guide that identifies where it is when selecting a component of the main interface	Once the parameter is clicked, appears the information associated with it
There are many elements that can confuse the user	Due to the amount of data is necessary the visualization of all the parameters
The information in the recommendation table presents a lot of information. In it, a vocabulary based on iconography could be used	The information was shortened and a color code was implemented
The contamination parameter selected by the user is not highlighted	A color over the parameter selected was implemented
It is not indicated what is the pollution value belonging to the statistics	The information associated appears in the statistics
It is not allowed to filter the information in the sensor report For example, the operator may need to indicate a range of dates	This option was included, however didn't well explained to the users in the test

5 Conclusions

In this paper we present an Air Quality Sensor Monitor system for human activities in dangerous environments. We propose a hardware architecture so a mobile robotic platform could acquire environmental parameters and send it to a PC interface. Moreover, a friendly-user interface was developed in order to show the parameters collected by the robot. The user interface was submitted to a heuristic evaluation in order to detect usability concerns. The results showed that the system proposed has a potential application. Likewise, several recommendations made by experts were taken in to account in order to improve the user interface usability. Finally, the original and the improve interface with the expert's recommendations were presented.

Acknowledgements.. The authors thank the staff of Unidad de Innovación y Tecnológia (UITEC) for their support, equipment and infrastructure.

References

1. National Research Council (US) Committee on Acute Exposure Guideline Levels: Volume 8. Sulfur Dioxide Acute Exposure Guideline Levels. In: Acute Exposure Guideline Levels for Selected Airborne Chemicals. National Academies Press (US), Washington (DC) (2010)

2. National Research Council (US) Subcommittee on Rocket-Emission Toxicants: Assessment of Exposure-Response Functions for Rocket-Emission Toxicants. Appendix E, ACUTE TOXICITY OF NITROGEN DIOXIDE. National Academies Press (US), Washington (DC) (1998)
3. Demling, R.H.: Smoke inhalation lung injury: an update. Eplasty **8**, e27 (2008)
4. Phillips, A., Clark-Leach, G., McGraw, K., J.L., A., C., B., K., K.: Preparing for the next storm learning from the man-made environmental disasters that followed hurricane harvey. Eip rep., Environmental Integrity Project, Washington (August 2018)
5. Somervell, E., Aberkane, T.: The effects of on-going seismic activity on air quality in canterbury, New Zealand. Open Atmos. Sci. J. **8**, 1–6 (2014)
6. Tang, B., Ge, Y., Liu, X., Kang, P., Liu, Y., Zhang, L.: Health-related quality of life for medical rescuers one month after ludian earthquake. Health Q. Life Outcomes **13**, 88 (2015)
7. Slottje, P.I., Twisk, J.W.R., Smidt, N., Huizink, A.C., Witteveen, A.B., van Mechelen, W., Smid, T.: Health-related quality of life of firefighters and police officers 8.5 years after the air disaster in amsterdam. Q. Life Res. **16**(2), 239–252 (2007)
8. Khare, M.: Air pollution –monitoring, modelling and health. InTech, Janeza Trdine 9, 51000 Rijeka, Croatia (2012)
9. Mayor of London The Queen's Walk SE1 2AA, London: Guide for Monitoring air Quality in London (2018)
10. Sharma, J., John, S.: Real time ambient air quality monitoring system using sensor technology. Int. J. Adv. Mech. Civil Eng. **4**(1), 72–74 (2017)
11. Khedo, K.K., Perseedoss, R., Mungur, A.: A wireless sensor network air pollution monitoring system. CoRR abs/1005.1737 (2010)
12. Mead, M., Popoola, O., Stewart, G., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J., McLeod, M., Hodgson, T., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J., Jones, R.: The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. **70**, 186–203 (2013)
13. Subramanian, R., Ellis, A., Torres-Delgado, E., Tanzer, R., Malings, C., Rivera, F., Morales, M., Baumgardner, D., Presto, A., Mayol-Bracero, O.L.: Air quality in puerto rico in the aftermath of hurricane maria: a case study on the use of lower cost air quality monitors. ACS Earth Space Chem. **2**(11), 1179–1186 (2018)
14. Acosta-Vargas., P., Zalakeviciute., R., Luján-Mora, S., Hernandez, W.: Accessibility evaluation of mobile applications for monitoring air quality. In: International Conference on Information Technology and Systems (2019, In Press)
15. Pérez-Medina., J.L., Zalakeviciute., R., Rybarczyk, Y.: Assessing the usability of air quality mobile application. In: The International Conference on e-Democracy and e-Government (2019, In Press)
16. Pérez-Medina, J.L., Vanderdonckt, J.: A tool for multi-surface collaborative sketching. In: WorkShop Cross-Surface 2016: Third International Workshop on Interacting with Multi-Device ecologies in the wild (2016)
17. Paterno, F.: Model-Based Design and Evaluation of Interactive Applications. Springer, Heidelberg (1999)
18. Pribeanu, C.: A revised set of usability heuristics for the evaluation of interactive systems. Inf. Econ. **21**(3), 31 (2017)

Improving Web Accessibility: Evaluation and Analysis of a Telerehabilitation Platform for Hip Arthroplasty Patients

Patricia Acosta-Vargas[1(✉)], Luis Salvador-Ullauri[2], Jorge Luis Pérez-Medina[1], Mario Gonzalez[1], Karina Jimenes[1], and Yves Rybarczyk[1]

[1] Intelligent and Interactive Systems Lab, Universidad de Las Américas, Quito, Ecuador
{patricia.acosta, jorge.perez.medina, mario.gonzalez.rodriguez, karina.jimenes, yves.rybarczyk}@udla.edu.ec

[2] Centro de Educación Continua, Escuela Politécnica Nacional, Quito, Ecuador
lsalvador@cec-epn.edu.ec

Abstract. Currently, the accelerated evolution of technology and the adoption of new platforms in the field of telerehabilitation pose considerable challenges in the field of web accessibility. Web accessibility aims to make web pages usable by the maximum number of people, regardless of their knowledge, personal abilities or the technical characteristics of the equipment used to access the Web. This paper presents the accessibility evaluation of ePHoRt, a web platform that aims to improve the rehabilitation of patients after a total or partial hip replacement operation known as arthroplasty. The accessibility method is based on the Website Accessibility Conformance Evaluation Methodology 1.0 that considers the Web Content Accessibility Guidelines 2.1. We evaluate first the accessibility of ePHoRt using the WAVE automatic tool. Then, manual inspection was performed. Based on the results, we present recommendations to ensure that the application is accessible and inclusive. The results can serve for future work related to web accessibility in telerehabilitation platforms.

Keywords: Accessibility · Analysis · Evaluation · e-Health · Hip arthroplasty patients · Improving · Platform · Telerehabilitation · Web Content Accessibility Guidelines (WCAG) 2.1

1 Introduction

World Wide Web Consortium (W3C) [1] defines that a website is accessible if people with disabilities can use it as effectively, safely and securely as people without disabilities. Web accessibility means that people with disability will be able to perceive, understand, navigate and interact with the web. Web accessibility also benefits other people, including older people who have seen their skills diminish as a result of age.

According to the guidelines proposed by the Web Accessibility Initiative (WAI) [2], for a website to be accessible it must contain a content that is easily

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 508–519, 2020.
https://doi.org/10.1007/978-3-030-20040-4_46

understood and navigable by all people, regardless of the physical, intellectual or technical disability they present.

Following the World Health Organization, about 15% of the world's population lives with some form of disability [3]. The global estimate of disability is increasing due to population aging and the rapid spread of chronic diseases.

More than a billion people around the world have some disability [4], in future years, disability will be an even higher concern, as its prevalence is increasing. At present, [4] the high rate of elders' population in developed countries is an essential socio-health concern for governments.

The accelerated advance of information, communication, and medical technologies have the development of telemedicine platforms. However, not all platforms can be used by all people, because the factors of inclusion and access to them are not considered. As a result, existing platforms are not accessible [5], which creates a problem of interaction between users and platforms.

The ePHoRt is a Web platform that aims to improve the rehabilitation of patients after a total or partial hip replacement operation known as arthroplasty [6, 7]. The goal of the application is to reduce recovery costs as well as provide access to health at a low cost.

On the other hand, the new challenges of the platform and accessibility are primarily related to the development of a system to support the physiotherapist during the process of planning, monitoring, and evaluation of rehabilitation exercises [8]. In order for the telerehabilitation platform to be inclusive for all types of users, according to previous studies carried out for the patient [9], this study must also be accessible for the physiotherapist.

In this research, the standards of the Web Content Accessibility Guidelines (WCAG) 2.1 [10] and an adaptation of the Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0 [11] were applied. The study proposes to apply an interactive evaluation method for the telerehabilitation platform that consists of nine phases: (i) define the scope of the evaluation, (ii) explore the destination website, (iii) select the representative sample, (iv) evaluate the selected sample, (v) record the results of the evaluation, (vi) analyze the results, (vii) manually check for errors, (viii) suggest improvements, and (ix) correct errors.

In future work, it is suggested to look for new methods, techniques and combined strategies to evaluate and improve web accessibility in the telerehabilitation platform. This study suggests to website developers and designers to apply WCAG 2.1 in order to generate more inclusive web platforms.

The rest of this article is structured as follows: Sect. 2 details the background and related work, Sect. 3 presents the method used and the case study, Sect. 4 presents the results and discussion, and finally, Sect. 5 presents the conclusions and future work of this research.

2 Background and Related Work

At present, the trend in rehabilitation medicine is distance therapy platforms. Patients can perform rehabilitation therapies [12] from home through a web platform, which brings several advantages regarding cost reduction, time, and access to health services.

Rybarczyk et al. [8] initiated the ePHoRt project for remote monitoring of telerehabilitation in patients after hip replacement. In another study [13] used hidden Markov models to assess the quality of patient movements, and a Kinect camera was used to capture the movements.

The implementation [14] of an appropriate telerehabilitation web platform is essential so that patients can follow the physiotherapist's instructions after arthroplasty. It is, therefore, necessary for the physiotherapist to have an accessible web platform.

Previous studies by several authors related to web accessibility in telerehabilitation applications were reviewed, described below:

Calle-Jiménez et al. [15] propose an iterative method to improve the level of accessibility through automatic evaluation tools. The results indicate that the studied platform does not comply with all WCAG 2.0 accessibility standards.

Acosta-Vargas et al. [16] present an accessibility study performed on a Web platform to promote physical rehabilitation of patients with an arthroplasty. The authors used tools through plugins installed in the Web browser. WAVE[1], Siteimprove[2], Open WAX[3], and Tenon[4] were used. The results indicate that the tele-rehabilitation platform requires improvements to reach an appropriate level of Web accessibility.

Also, Bargagna et al. [17] describe a set of computerized exercises designed for cognitive training of adults with Down Syndrome. The objective of the study was to develop a web telerehabilitation platform by following WCAG 2.0.

Smeddinck et al. [18] argue that motion-based video games may have a variety of benefits to apply to physiotherapy, rehabilitation, and prevention in older adults. The authors suggest that aspects of accessibility and immersion should be considered to harness the potential benefits of motion-based video games.

Winters [19] proposes a novel framework for the development of interfaces leading to universal access to telerehabilitation for persons with disabilities. This study makes use of some guidelines related to WCAG and Section 508.

Finally, Bąkała et al. [20] explain the concept of online health services in Poland, which tends to include people with disabilities. The researchers present a comparative analysis of e-health solutions in Poland and other European countries.

We focus on using an adaptation of Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0 and, Web Content Accessibility Guidelines (WCAG) 2.1 [10], contain four principles, 13 guidelines and 76 compliance criteria (success), plus an indeterminate number of sufficient techniques and counseling techniques. The principles are the following:

Principle 1: Perceivable - All users must be able to perceive the content in a visual, sound, tactile way.

Principle 2: Operable - Users must be able to use and navigate the interface components.

1 https://wave.webaim.org/.

2 https://siteimprove.com/.

3 https://github.com/goonoo/OpenWAX.

4 http://tenon.io/.

Principle 3: Understandable - Both the content and the controls of the interface for its management must be understandable to the user.

Principle 4: Robust - Content must be robust enough to be reliably interpreted by a variety of application users, and to work with current and future technologies.

3 Method and Case Study

3.1 Case Study

The case study was applied in ePHoRt [8] which is a project that aims to develop a web-based system for remote monitoring of rehabilitation exercises in patients after hip replacement surgery. The platform uses Kinect as an appropriate tool for the therapeutic purpose of the project. In this case, six web pages of the platform were evaluated. The evaluation was applied to the pages related to the management of the physiotherapist. It contains functions for the management of patients, physicians, messages, monitoring of therapeutic programs, management of learning resources and results of therapeutic programs.

3.2 Method

This study applies an adaptation of the WCAG-EM 1.0 and the Web Content Accessibility Guidelines (WCAG) 2.1.

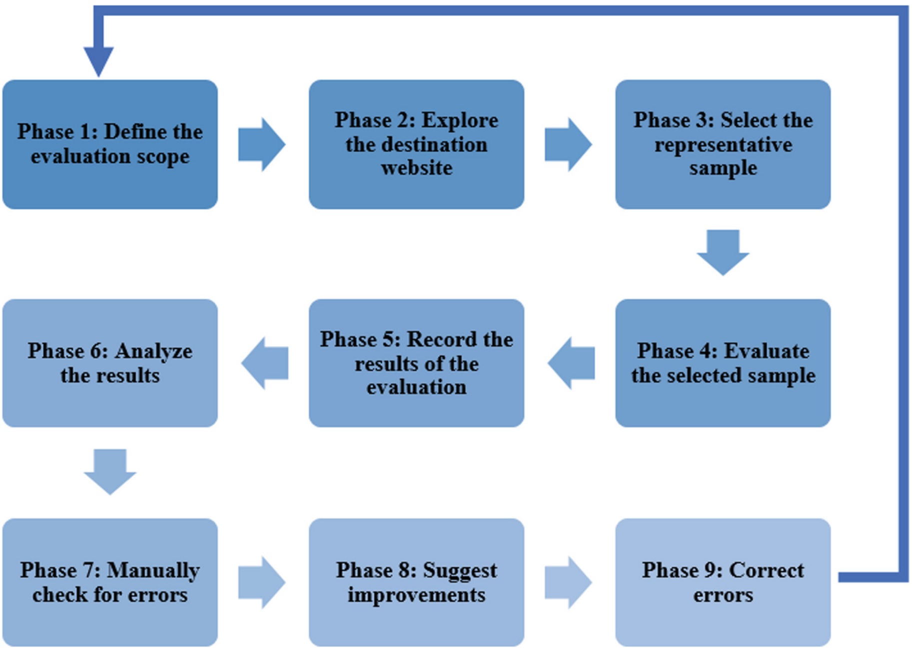


Fig. 1. Method for accessibility assessment in telerehabilitation platform

The study proposes to apply a combined evaluation method between the WAVE (Web Accessibility Evaluation tool) used in previous studies by the authors [16] and the manual revision until the telerehabilitation platform is accessible, according to the level of accessibility defined by the evaluator, this method consists of nine phases, according to Fig. 1.

WCAG-EM 1.0 is a methodology adapted to different situations, such as self-evaluation and third-party evaluation. However, accessibility should not be applied only in the evaluation phase, but throughout the development cycle of the platform. All phases are applied in the same way as proposed in WCAG-EM 1.0, with a variation from phase 4.

Phase 1: Define the Evaluation Scope. In this phase, six web pages were randomly selected from the telerehabilitation system, including physiotherapist pages. At this stage, the adequacy level (A, AA, AAA) to be assessed must also be defined. It is generally applied up to level AA, but it is often useful to broaden the scope to make the site more accessible. It is essential to define a list of browsers with which to navigate, support products or other user agents with which accessibility features must be compatible. Table 1 contains the identifier of the screen, the name, and the description.

Table 1. Web pages of the telerehabilitation platform to evaluating the process

ID	Screen	Description
A	Physiotherapist Menu	It presents the options that the physiotherapist has in the platform like patient management, doctor management, messaging, suspended therapeutic programs management, therapeutic programs management, exercise management, learning resources management, and results of the therapeutic programs
B	Therapeutic Program Management	It allows managing the current therapeutic program considering the associated patient and doctor. It also duplicates a previous therapeutic program and shows a button to check the details of each one
C	New Therapeutic Program	Add and edit information about exercises and resources in each stage. Three stages are defined by default considering a period. The physiotherapist can add new ones
D	Doctor Management	It allows managing the information of the patient's doctors. This information is essential in the case; the physiotherapist needs to talk with the patient or if he needs to send a report of the patient's therapeutic program
E	Therapeutic Program | Learning Resources	It allows managing a set of learning resources for the patient and letting the physiotherapist planning the therapeutic programs by using them. Also, a learning resource can be duplicated and associated with another therapeutic program as required
F	Therapeutic Program | Asign Exercises	It allows for creating a new exercise in order to use later in the planning therapeutic programs process. It requires a name, a set of instructions, a video and a skeleton file as reference. It also allows associating a therapeutic program previously defined in with the stages. The uploaded skeleton file is mandatory for the evaluation of the exercises

Phase 2: Explore the Destination Website. In this phase, the features of the pages evaluated are studied, such as styles, structures, designs and the functions they perform. The technologies used are identified, such as HTML, CSS, and JavaScript. In this phase also identifies the statuses of the relevant pages for people with disabilities, as well as the preferences and options of the site to support instructions. The evaluator can access all the pages, at this phase, the use, purpose, and functionality of the site are better understood, and it is also essential to identify the relevant or problematic pages that can be included in the sample to be analyzed.

Phase 3: Select the Representative Sample. In this third phase, it is essential to include the complete processes to achieve more reliable results. The idea is to evaluate the entire web platform. It should be noted that in this phase the screens were taken randomly. The sample size depends on several factors, such as site size, complexity, functionality, technologies used, coding styles, processes and functions performed by users.

Phase 4: Evaluate the Selected Sample. In this phase, the automatic WAVE tool (version: 1.0.9, updated on November 17, 2017) was selected for evaluation, with extensions in Google Chrome due to the ease of evaluation on websites that require authentication. This guarantees accessibility reports 100% private and secure. The extension can verify intranet web pages, including password, protected web pages, dynamically generated or sensitive. This tool was evaluated in previous studies by the authors [16]. The tool helps to verify each page according to the compliance criteria of WCAG 2.1.

Phase 5: Record the Results of the Evaluation. In this phase, the *Therapeutic Program Management* screen was used as a case study; Fig. 2 presents the errors that exist on the page.

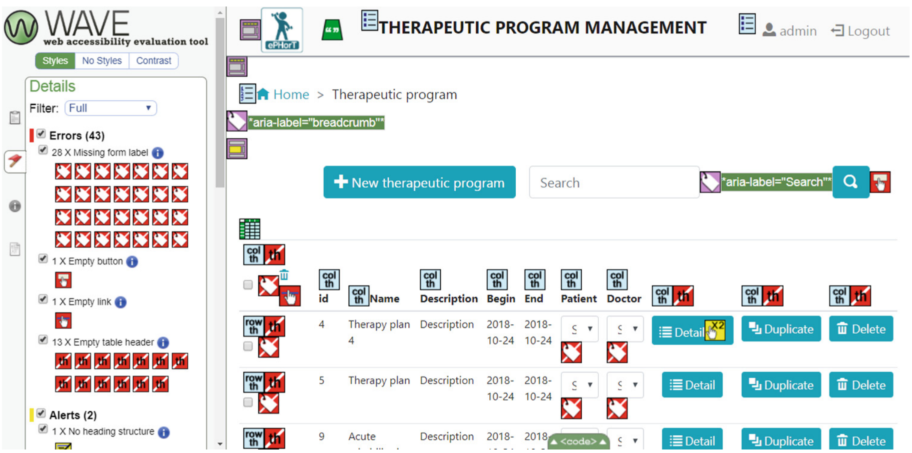


Fig. 2. Screenshot Therapeutic program management with errors

Figure 2 presents errors related to *missing form label*, implying that a form control does not have an appropriate text label for the function of the form, which is an issue because it is not presented to users using screen readers. This problem can be corrected by placing a text label for a form control that is visible, using the <*label*> element associated with the corresponding form control.

In the fifth phase, the results are recorded in a spreadsheet. Table 2 contains the data obtained in the evaluation of the web pages recorded in Table 1.

Table 2. Evaluation results with WAVE tool

ID	Errors	Alerts	Features	Structural elements	HTML5 and ARIA	Contrast errors
A	1	1	0	3	14	9
B	43	2	1	24	8	30
C	38	2	6	34	15	7
D	10	4	0	14	8	7
E	26	3	5	19	15	17
F	10	3	1	5	9	3
Total	128	15	13	99	69	73

Phase 6: Analyze the Results. In this phase, a detailed study of the errors identified with the WAVE tool is carried out. Table 3 shows the errors with the success criteria of WCAG 2.1.

Table 3. Evaluation results with WAVE tool

ID	Errors	Styles						Contrast
		Multiple form labels	Missing form label	Empty button	Empty link	Empty table header	Linked image missing alternative text	Very low contrast
A	**1**						1	9
B	**43**		28	1	1	13		30
C	**38**	14	11		3	10		7
D	**10**		3	1	1	4	1	7
E	**26**	5	10		1	10		17
F	**10**		8	1			1	3

Phase 7: Manually Check for Errors. In this phase, the errors identified on each page are checked manually. Table 4 records the detail of errors, standards and guidelines and level that violate some accessibility principles.

The error that is repeated most frequently corresponds to 1.1.1 with 26.7%, followed by 1.3.1 and 2.4.4 with 20%, with this it can be inferred that the most affected principle is Perceptible, where all content that is not text must place an alternative text to indicate to the user about a specific purpose.

Table 4. Errors vs. standards and guidelines

Detail of errors	Standards and guidelines	Level
Multiple form labels	1.1.1 Non-text Content, 1.3.1 Info, and Relationships 2.4.6 Headings and Labels 3.3.2 Labels or Instructions	A A AA A
Missing form label	1.1.1 Non-text Content 1.3.1 Info and Relationships 2.4.6 Headings and Labels 3.3.2 Labels or Instructions	A A AA A
Empty button	1.1.1 Non-text Content 2.4.4 Link Purpose (In Context)	A A
Empty link	2.4.4 Link Purpose (In Context)	A
Empty table header	1.3.1 Info and Relationships	A
Linked image missing an alternative text	1.1.1 Non-text Content 2.4.4 Link Purpose (In Context)	A A
Very low contrast	1.4.3 Contrast (Minimum)	AA

Phase 8: Suggest Improvements. In this phase, each page is corrected according to the suggestions identified. Table 5 contains the details of the errors and suggestions for correcting them.

Table 5. Suggestions for fixing errors

Detail of errors	Suggest improvements
Multiple form labels	Ensure that at most one label element is associated with the form control. If multiple form labels are necessary, use aria-labelledby This attribute is used to specify the name or label of the current item
Missing form label	For a visible form control, use the <label> element to associate it with the respective form control If no label is visible, provide an associated label, add a descriptive title attribute to the form control, or reference the label using aria-labelledby Labels are not required for an image, send, reset, button, or hidden form controls
Empty button	Place text content within the <button> element or give the <input> element a value attribute
Empty link	Remove the empty link or provide text within the link that describes the functionality and target of that link
Empty table header	If the table cell is a header, provide text within the cell that describes the column or row. If the cell is not a header or should remain empty, make the cell <td> instead of <th>
Linked image missing an alternative text	Add appropriate alternative text that presents the content of the image and the function of the link
Very low contrast	Increase the contrast between the foreground (text) color and background color. Large text (larger than 18 points or 14 points bold) does not require as much contrast as smaller text

Phase 9: Correct Errors. In this phase, the errors presented in the evaluation were corrected with the automatic tool. Figure 3 presents the corrected errors in the Therapeutic Program Management website. If the website meets the acceptable level defined by the evaluator, the process is finalized, otherwise returns to phase 1 to repeat the process until it reaches an acceptable level. According to WCAG 2.1 the definition, an acceptable level is considered if it reaches level AA [10].

4 Results and Discussion

Tables 2 and 3 contain details of the errors found in the evaluation of web pages. The analysis can help to understand the applicability of WCAG 2.1. The information found can be useful for designers, developers, and evaluators of web applications. Figure 3 presents a summary of the errors found on the pages evaluated. The screen with the most significant number of errors corresponds to the *Therapeutic Program Management*, followed by the *New Therapeutic Program* and *Therapeutic Program|Learning Resources*.

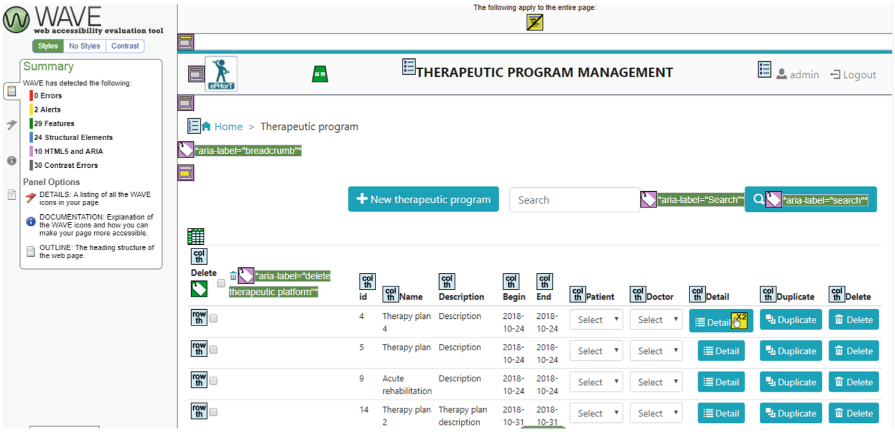


Fig. 3. Screenshot Therapeutic program management without errors

Besides, Fig. 3 shows that the most frequent error is *very low contrast* with 73 errors and corresponds to 36.3%, followed by *missing form label* with 60 errors that correspond to 29.9%, finally, *empty table header* with 37 errors that corresponds to 18.4%. Table 6 presents a summary of the principles, guidelines, success criteria and levels identified in the evaluation process.

Table 6. Summary of accessibility principles identified in the evaluation

Principle	Guideline	Success criteria	Level
1. Perceivable	1.1 Text Alternatives	1.1.1 Non-text Content	A
1. Perceivable	1.3 Adaptable	1.3.1 Info and Relationships	A
1. Perceivable	1.4 Distinguishable	1.4.3 Contrast (Minimum)	AA
2. Operable	2.4 Navigable	2.4.4 Link Purpose (In Context)	A
2. Operable	2.4 Navigable	2.4.6 Headings and Labels	AA
3. Understandable	3.3 Input Assistance	3.3.2 Labels or Instructions	A

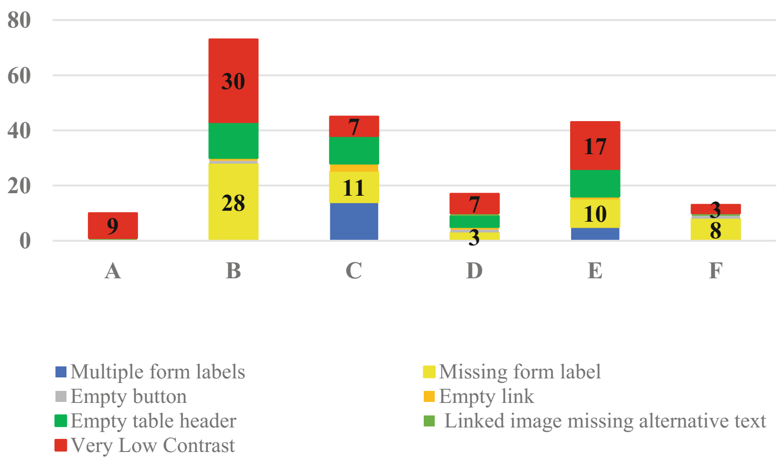


Fig. 4. Errors found in the evaluation

It is observed that 50.0% have faults related to the principle of *perceivable*, followed by *operable* with 33.3% of faults, finally, *understandable* with 16.7%.

In order to resolve the most frequently recurring faults in this telerehabilitation platform, the following improvements are suggested:

For the error related to *very low contrast*, it is suggested to increase the contrast between the foreground (text) color and the background color. Large text (more than 18 dots or 14 dots in bold) does not require as much contrast as smaller text.

The error related to the *missing form label* can be corrected by using the <*label*> element associated with the control of the respective form. If there is no visible label, it is necessary to provide a descriptive label with the form control.

To correct the error related to the *empty table header*, if the table cell has a header, it is suggested to provide a text within the cell that describes the column or row. If the cell is not a header or should remain empty, it is recommended to keep cell <*td*> instead of <*th*>.

5 Conclusions and Future Work

Evidently, from the experiment, we can conclude that it is not possible to measure the level of accessibility of a website in a fully automated way. Compliance criteria require that human judgment is validated. The tools are useful for a quick evaluation and to find common accessibility errors. However, they are not enough to determine if a website is 100% accessible, which will require a manual evaluation, despite being more expensive.

It is essential that the information reaches all people without any discrimination. It is essential to raise awareness and apply the principles of web accessibility to demonstrate and generate social responsibility in our society, while at the same time positively reinforcing the corporate image.

In future work, it is suggested to continue looking for new methods, techniques and strategies to evaluate and improve web accessibility on the telerehabilitation platform. Finally, this study suggests raising awareness among web developers and designers to apply the WCAG 2.1 guidelines during the application development cycle to create accessible and inclusive platforms.

Acknowledgments. The authors thank UDLA for funding this study through the project "SIS. JPM.18.01".

References

1. World Wide Web Consortium (W3C): Web Accessibility Perspectives: Explore the Impact and Benefits for Everyone—Web Accessibility Initiative (WAI)—W3C. https://w3c.github.io/wai-website/perspective-videos/
2. World Wide Web Consortium (W3C): Introduction to Web Accessibility—Web Accessibility Initiative (WAI). https://www.w3.org/WAI/fundamentals/accessibility-intro/
3. World Health Organization (WHO): WHO—World report on disability. https://www.who.int/disabilities/world_report/2011/report/en/
4. World Health Organization (WHO): World report on disability 2011 (2011)
5. Leporini, B., Hersh, M.: Games for the rehabilitation of disabled people. In: Proceedings of 4th Workshop on ICTs Improving Patients Rehabilitation Research Techniques – REHAB 2016, pp. 109–112 (2016). https://doi.org/10.1145/3051488.3051496
6. Rybarczyk, Y., Jimenes, K., Leconte, L., Medina, J.P., Acosta-Vargas, P.: Telerehabilitation platform for post-arthroplasty recovery : a dynamic time warping approach. In: ETCM (2018)
7. Rybarczyk, Y., Villarreal, S., González, M., Acosta-Vargas, P., Esparza, D., Sanchez-Gordon, S., Calle-Jimenez, T., Jadán, J., Nunes, I.L.: Interaction with a tele-rehabilitation platform through a natural user interface: a case study of hip arthroplasty patients. In: International Conference on Applied Human Factors and Ergonomics, pp. 246–256 (2018)
8. Rybarczyk, Y., Deters, J.K., Gonzalvo, A.A., Gonzalez, M., Villarreal, S., Esparza, D.: ePHoRt project: a web-based platform for home motor rehabilitation. In: Advances in Intelligent Systems and Computing, pp. 609–618 (2017)
9. Acosta-Vargas, P., Luján-Mora, S., Acosta, T., Salvador-Ullauri, L.: Toward a combined method for evaluation of web accessibility. In: Advances in Intelligent Systems and Computing, pp. 602–613 (2018)

10. World Wide Web Consortium (W3C): Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/
11. World Wide Web Consortium (W3C): Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0. https://www.w3.org/TR/WCAG-EM/
12. Rybarczyk, Y., Kleine Deters, J., Cointe, C., Esparza, D.: Smart web-based platform to support physical rehabilitation. Sensors **18**, 1344 (2018). https://doi.org/10.3390/s18051344
13. Deters, J.K., Rybarczyk, Y.: Hidden Markov model approach for the assessment of tele-rehabilitation exercises. Int. J. Artif. Intell. **16**, 1–19 (2018)
14. Van Der Weegen, W., Kornuijt, A., Das, D.: Do lifestyle restrictions and precautions prevent dislocation after total hip arthroplasty? A systematic review and meta-analysis of the literature. Clin. Rehabil. **30**, 329–339 (2016). https://doi.org/10.1177/0269215515579421
15. Calle-Jimenez, T., Sanchez-Gordon, S., Rybarczyk, Y., Jadán, J., Villarreal, S., Esparza, W., Acosta-Vargas, P., Guevara, C., Nunes, I.L.: Analysis and improvement of the web accessibility of a tele-rehabilitation platform for hip arthroplasty patients. In: International Conference on Applied Human Factors and Ergonomics, pp. 233–245. Springer, Cham (2019)
16. Acosta-Vargas, P., Rybarczyk, Y., Pérez, J., González, M., Jimenes, K., Leconte, L.: Towards web accessibility in telerehabilitation platforms. In: ETCM (2018)
17. Bargagna, S., Bozza, M., Buzzi, M.C., Buzzi, M., Doccini, E., Perrone, E.: Computer-based cognitive training in adults with down's syndrome. In: International Conference on Universal Access in Human-Computer Interaction, pp. 197–208. Springer, Cham (2014)
18. Smeddinck, J., Gerling, K.M., Tiemkeo, S.: Visual complexity, player experience, performance and physical exertion in motion-based games for older adults. In: Proceedings of 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS 2013, pp. 1–8 (2013). https://doi.org/10.1145/2513383.2517029
19. Winters, J.M.: Telerehabilitation interface strategies for enhancing access to health services for persons with diverse abilities and preferences. In: Telerehabilitation, pp. 57–78. Springer (2013)
20. Bąkała, A., Korczak, K.: Accessibility of e-health services for people with disabilities. Pr. Nauk. Uniw. Ekon. we Wrocławiu. **18**, 30–31 (2010)

Advances in User Experience, Affordance and Technology

Pharmaceutical Online Store Project: Usability, Affordances and Expectations

Marcelo A. Guimarães[1(✉)], Rodrigo P. Zisman[1], and Adriano B. Renzi[2]

[1] Original.io, Rio de Janeiro, Brazil
{marcelo.guimaraes,jesus}@original.io
[2] Universidade Federal Fluminense – Industrial Design – TDT, Niterói, Brazil
adrianorenzi@gmail.com

Abstract. Pharmaceutical business has started reaching out for e-business in Brazil, trying to better fill the digital gap and reach a higher audience. This type of business doesn't show the same relation between digital and physical stores, as the Chris Anderson' long tail theory perceived in bookstore business, but the recent increase of the delivery business has open possibilities and urged the pharmaceutical enterprises to follow. Based on precedent data, a high fidelity prototype was developed and a series of tests were applied to verify its usability and task flow. Conclusion discusses the relation of proposed interaction with users' expectations, the usability tests with the prototype, the relation between stakeholders' interests in the project, the influence of this factor in the applied methods and the final outcome, merging research science with market and business demands.

Keywords: Human factors · Human-computer interaction · Usability

1 Introduction

Pharmaceutical business has been reaching out for e-business in Brazil, trying to better fill the digital gap and reach a higher audience. This type of business doesn't show the same relation between digital and physical stores, as the Anderson' long tail theory [1] perceived in bookstore business [2], as it its digital branch doesn't focus on products of the far tail. The recent increase of the delivery business [3] has open possibilities and urged the pharmaceutical enterprises to follow. In the 3rd wave of computing era, where daily chores are being fulfilled more often with the help of a digital devices in pervasive actions, it becomes imperative to any business to consider the digital interaction as part of the business model and flow.

This paper presents the phases related to the design revision of an online pharmaceutical store, developed with the 2 wave of computing thinking into a pervasive experience journey interaction system. The process merges the understanding of users' needs and expectations with the stakeholders' premises and strategies.

Based on business directives, initial phases of the project included detailed analysis from Google Analytics to understand the online store's sales performance, conversion rate, bounce rate as well as most visited categories, followed by Card Sorting [5, 17] to

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 523–534, 2020.
https://doi.org/10.1007/978-3-030-20040-4_47

help establish the structure and information architecture of the store and its place in the interactive journey. Based on these precedent data, a high fidelity prototype was developed and a Heuristic Evaluation [6] analysis based on Nielsen and Molich 10 heuristics with 3 UX experts, was applied to verify the system usability and task flow. The project is presented through its sequential UX testing phases and validation of the updated prototype using Think-aloud Protocol [7, 8] with final users.

Conclusion discusses the final results, validation and next steps as well as the relation of stakeholders' strategies with users' mental model and expectations, from collected data and usability tests, to analyze the full research process and its impact on stakeholders' interests and the final outcome, merging research science with market and business demands.

As technology advances and gets to be part of pharmaceutical business growth (stock management, larger product offer, consumer behavior data, digital stores), since the early 1980's, Brazil has the 6th biggest pharmaceutical market of the world – according to IQVIA, with profits reaching US$ 33.1 billions in 2017 (Fig. 1). The EMIS Insights Market Analysis puts Brazil surpassing France for the 5th place on 2022. Interfarma (Pharmaceutical Industry Research Association) shows a total of US$ 43.7 billions which in the next 5-year period will increase 60%.

The Venancio Drugstore chain has focused its business for 39 years exclusively on Rio de Janeiro. The business model acts from service to products, based on three major lines: beauty, health and well-being. The first store was founded in Tijuca, a middle class neighborhood, and the business has gradually spread to 67 stores, from very poor areas to high per capita income neighborhoods of the city. The business attends not only people as final consumers, but also hospitals, government and private medical clinics. Attempting to reach a broader audience of consumers, the drugstore takes its first step towards e-business and delivery system, offered exclusively through phone sales until recently.

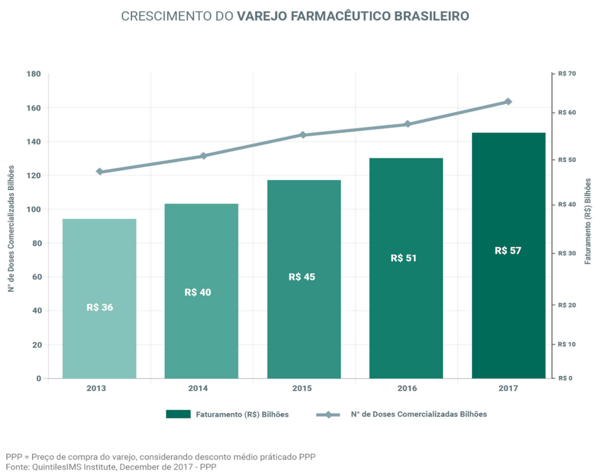


Fig. 1. Pharmaceutical market growth in the last 5 years (IQVIA).

Although the drugstore offers a diverse and broad range of services and products, the increasingly presence of online businesses since the 1990's makes the e-business move seems a little late in the 3rd wave of computing. However, the proposal intends to offer availability of connection and purchase from different channels, expecting at least to provide responsive design and usability to easier the user experience and meet users' expectations and mental models.

The online stores, as oppose to brick and mortar stores, do not inflict social embarrassment during search-purchase procedures and users feel free to give up products in the middle of the experience [9, 10]. Usability still plays a major role to help users fulfill tasks, find products and influence positively purchase decision and the reliability of the store. According to Banati *et al*. [11] and Renzi and Freitas [12], usability is closely related to users' trust on an e-business. The authors (*ibidem*) exemplify the relation as auto-stimulatory as the good reputation of an e-business attracts users to visit the website for the first time, either by recommendation, references or exploratory search. The good experience using the e-store, with good usability to build a satisfactory experience journey, will have users coming back for future consulting and purchase, even if nothing was bought in the first time. Users ponder over risks and benefits associated to online shopping before finalizing the purchase [12], from security and privacy to improper billing not fully disclosed from the beginning of the user experience.

The main goal established by Venancio Drugstore was to develop a better user experience and interconnect both desktop and smartphone channels in one continuous interaction, since the system has been previously observed as confusing and with many usability problems. The whole strategy starts with re-thinking the e-commerce structure and extends further to its physical stores, putting together an omnichannel experience. This is the first phase of the project focused on the digital branch, but understanding the whole integration of the business.

For this part of the project Original.io, a UX design agency located in Rio de Janeiro, had the primal challenge to reach the balance between Venancio Drugstore's interests on the growth of the brand's position in the market and users' needs and expectations using the drugstore website. The agency calls this balance challenge a Relation Triad, which involves (1) the stakeholders' interests and business strategies, (2) users' needs and expectations and (3) UX design directives – one nurturing and influencing the other.

On the e-commerce business segment, it is important to reach a responsible relation between brand strategies and users' needs, as developing empathy for stakeholders means looking at the project from their perspective and take to action, understanding the stakeholder's pain as well as the user's.

During the development of interactive products it is important to understand users using research methods and test the project itself throughout the whole process. Startups, when conceptualizing new products or services, need to find balance between technology, process and market expectations. They need to identify and involve people to find what delights them and then transform them in users. The UX design approach could diminish the risks of new enterprises, financially and timely, and can be the key to make users fall in love with new products and help filter ideas that seemed incredible at first [13]. Since the beginning of the user centered design [14] it is recommended that testing methods throughout the whole development process could bring products closer to users' expectations as well as minimize errors and loss of investments. According to Abras *et al*., the direct benefits of using usability testing during the creation process are: (1) usability increase of the product, (2) involve real users in the process, (3) give real tasks for users to test, (4) enable researchers to observe and record users' actions with the product and (5) enable researcher analyze collected data and make proper changes to the project.

Szuc and Wong [15] principals and cofounders of Apogee Usability Asia Ltd, point a few factors that should be considered to create great UX design:

– meets user needs that a business fully understands and nurtures
– maps to business needs, whose improvement we can track over time
– connects to data points that speak to the product or service story
– uses a well-defined and well-understood design framework that scales well and promotes consistent and usable interactions
– leverages design patterns that promote useful, usable, and delightful interactions
– maps to well-defined design principles that connect to brand principles and business goals and directions
– undergoes continuous improvement through customer and business feedback
– tries out new ideas and conducts experiments that do not disrupt the core value
– is led and owned by people who are well educated and grounded in deep knowledge of design foundations
– is visible and improves through structured and balanced critique

2 Analytics and Card Sorting

After a few meetings to better understand stakeholders' expectations, the first location visit had the objective to analyze similarities between the information architecture of the drugstore e-business and the physical stores. The two branches of the business were identified as very different proposals with just a few similarities, lacking connections for an integrated user experience. All categories of products were analyzed and the best selling products were identified, having medicine as the biggest seller, followed by dermatological cosmetics. Although the pharmaceutical market is very inconsistent regarding product categorization, based on the 5 major pharmaceutical companies, a series of cards were created for the Card Sorting execution.

A preliminary analysis of Venancio Drugstore older website, using Google analytics, helped understand the basics of their online consumers. This phase surfaced which users are commonly attracted to the e-store, the most visited pages, the most common paths used to explore and origins of connection. This data helped select users to participate in the Card Sorting phase.

According to Sherwin [17], from NN group (https://www.nngroup.com/articles/card-sorting-definition/), Card Sorting is "a UX research method in which participants group individual labels, written on notecards, according to criteria that make sense to them. This method uncovers how the target audience's domain knowledge is structured, and it serves to create an information architecture that matches users' expectations". The method has 4 phases:

1. *Choose a set of topics.* The set should include 40–80 items that represent the main content on the site. Write each topic on an individual index card.
2. *User organizes topics into groups.* Shuffle the cards and give them to the participant. Ask the user to look at the cards one at a time and place cards that belong together into piles. Some piles can be big, others small.
3. *User names the groups.* Once the participant has grouped all the cards to her satisfaction, give her blank cards and ask her to write down a name for each group she created. – *Debrief the user*. (optional, but highly recommended.) Ask users to explain the rationale behind the groups they created.
4. *Analyze the data.* Once you have all the data, look for common groups, category names or themes, and for items that were frequently paired together.

For this project it was used the Closed Card Sorting [4, 17] with representation cards of the main categories, since Venancio Drugstore has in total 619 different product categories. Closed Card Sorting is a variation where users are given a predetermined set of category names, and they are asked to organize the individual cards into these predetermined categories. Closed Card Sorting does not reveal how users conceptualize a set of topics. Instead, it is used to evaluate how well an existing category structure supports the content, from users' perspective.

Participants between ages 20 and 35 from male and female genders were selected to participate. The invitees represented users who frequently buy drugstore products through Internet, users who buy sporadically and users who have never bought any product on any e-drugstore.

There were many similar information structures between the participants, but a few particularities called attention: different genders with different points of view regarding beauty products and daily care categorization; part of the participants preferring to have vitamins & supplements apart from medicine, but some suggesting that everything should be under medicine; users got confused with the term convenience and what should be part of it.

From users' feedback, a basic structure of the system (Fig. 2) was proposed and a high fidelity prototype was developed for usability evaluation. Keeping each department colors from the physical stores helped users identify easier the different sessions of products. The informational structure had in consideration the unification of duplicate categories from the older website version, in order to each product be linked to only one category.

Fig. 2. Information structure based on Pharmaceutical business analysis, online analytics and Closed Card Sorting [4, 16].

3 Heuristic Evaluation and Think-Aloud Protocol

The Heuristic Evaluation consists of 3–5 experts going through a system to analyze and evaluate its interactions based on a set of heuristics. According to Nielsen's experiment in 1992 [6] with 19 usability experts, the use of heuristic evaluation showed that using

3–5 experts could bring up around 75% of the total usability errors. For this project, the Heuristic Evaluation using desktop/laptop encompassed 3 usability specialists. Each specialist accessed the online prototype using their own computer, as a different device could be unfamiliar and bring factors to distort the results.

The first phase consists of each expert going through the system to analyze and evaluate its interactions based on Nielsen and Molich set of heuristics. The experts evaluate the system separately, with no exchange of information between them, and present a report with all notified issues. Each notified usability problem is related to one or more heuristics. The ten heuristics, created during the 2nd wave of computing, are: (1) Visibility of system status; (2) Match between system and the real world; (3) User control and freedom; (4) Consistency and standards; (5) Error prevention; (6) Recognition rather than recall; (7) Flexibility and efficiency of use; (8) Aesthetic and minimalist design; (9) Help users recognize, diagnose, and recover from errors; and (10) Help and documentation.

The 3 experts evaluations are compiled in one final report by the main researcher and the pointed heuristic problems are categorized by each expert in degrees of gravity, from 0 to 4:

- Gravity 0 (zero): It is not a problem. Usually pointed to parts of the system or interface that could be considered awkward, but does not get in the way of users.
- Gravity 1 (one): It is an aesthetic problem. Pointed mostly to parts of the system related interface layout that may not follow users expectations or are considered an aesthetic flaw, but do not affect users interactions or interoperability of information.
- Gravity 2 (two): It is a minor problem. These are problems that cause discomfort to users, may bring misunderstandings of information, result in interaction or action confusion, oblige users to re-do or take longer actions, or distort the notion of place-making. These problems do not impede procedures nor may cause serious interaction problems, but certainly are problems that bring discomfort and irritability to users.
- Gravity 3 (three): It is a major problem. These are problems that may diverge users from their objectives. They can take forms such as misleading information or interactions, actions requirements that are not obvious or are completely different from other channels, lack of interoperability between channels, information or action break, functionalities that are far from users expectations or do not follow a set pattern or have low affordance. These problems can make users give up using the system if there is any other option out of it.
- Gravity 4 (four): It is a catastrophic problem. These are problems that may impede users to reach objectives or compromise the whole system structure or interfere seriously in users' sense of Place-making. Mostly, these are problems instated in the system's concept, structure, taxonomy, interoperability and information system that affects the whole idea of cross-channel interaction, or even just isolated interactions. Serious problem can bring companies to re-think the whole idea of a product.

Although it is expected that evaluators explore the whole system to search and analyze problems, it was suggested to all evaluators to fulfill the task: search and buy a pain killer and execute payment through invoice – a bank slip commonly used in Brazil for people who do not wish to use credit cards for online purchases.

From the heuristic pointed problems, the prototype was updated and 4 users were invited to be part of a Think-aloud Protocol in order to validate the changes [8, 9]. The invitees were representatives of final users, based on precedent user mapping in the beginning of the project. The same task of "buying a pain killer and paying through bank slip" was presented to the invited users. The observation focused specially on updated parts from the Heuristic Evaluation report: information understanding during check out procedure, payment action flow and recognition of products and labels.

According to Villanueva [8], the method consists of a researcher observing users doing specific tasks within a controlled environment. The users' actions and thoughts are to be described verbally aloud on real time. The researcher records the users' actions by written notifications, video or voice recording. Filming and voice recording have the advantage of capturing the exact steps and descriptions of users, while written notifications depends on the researcher experience with observing reactions and quickness in writing down relevant actions of the experiment [9]. When noticing some reluctance from the users in verbalizing actions and thoughts during the Think-aloud Protocol, questions related to the users' actions should be placed to keep the flow of verbalization of their thoughts.

Villanueva [4] indicates the use of Think-aloud Protocol in 4 steps

1. Organize a small number of users (around 4)
2. The researcher meets each user separately
3. The researcher provides a prototype for users to execute a list of tasks
4. Users verbalize thoughts and actions while executing tasks
5. Researcher makes notes on improvements to be considered for the system when analyzing results.

4 Results

The evaluators reported a total of 83 usability problems using the high fidelity prototype. The most number of occurrences were related to heuristic 4 (consistency and standards) with 25 errors. Error prevention (h5) in second was cited to 15 usability errors, followed by heuristics 3 (user control and freedom) and 7 (flexibility and efficiency of use) to 11 problems, each (Fig. 3). No errors were linked to heuristic 9 (help users recognize, diagnose and recover from errors) – a difficult heuristic to verify in a prototype.

As an example of the pointed problems, the overlapping label with discount information on product's information averts users to see properly its characteristics (h2, h4, h8). The noted problem led to the creation of a specific area in each product for sales and discount labels.

Evaluators also pointed the lack of option (h7) to print an invoice for payment – a bank slip commonly used in Brazil for online purchases. Many errors related to heuristic 5 (error prevention) was expected as the prototype still showed incompletion and malfunction in certain areas.

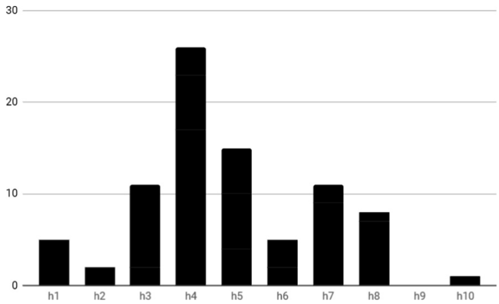


Fig. 3. Usability problems related to each of the10 heuristics of Nielsen and Molich.

Major problems were marked more frequently (32 problems), mostly related to heuristics h3, h4, h5 and h10. The amount of errors related to heuristic 3 surpasses the sum of all the other heuristics' major issues (Fig. 4). The lack of users' control while trying to choose the payment option before finalizing the purchase process is one example of many problems related to the heuristic. From users' point of view, the payment method selection should be before finishing the product order. The confusing documentation regarding the company's policy, statements and how to proceed was also a common pointed issue.

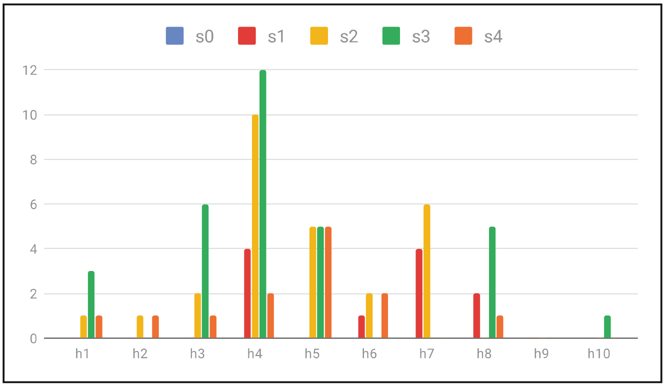


Fig. 4. Relation between usability erros, heuristics and gravity

Problems categorized as minor totaled 27 notifications, mostly related to heuristic 4 (10 errors). Heuristics 7 and 5 also show significant number of notifications by the evaluators, such as in products description page, the correlated suggestions of other products automatically increase the size of the page, taking down the Finalize Order button, reachable only by scrolling down (h7). The sudden disappearance of the button could bring confusion to users and impact negatively the flow process.

The cited catastrophic problems were numerous (18), considering the usual low rate found in Heuristic Evaluations, mostly related to the malfunction of the system (h5) and heuristics 4 ad 6, such as “no breadcrumbs nor category title when entering any department page”.

5 Conclusion

The sequential tests with users participation and UX experts approximated the project to users’ mental model. Each step brought up new marks to the project to be reshaped, fixed and rethought in a continuous refining. The Card Sorting helped understand the users’ product organization point of view and clarified the middle point between users’ thinking and stakeholders strategies. The information architecture was a turning point to establish the base for the prototype development. Even though the information structure is the foundation of the project, the creation of the interactive prototype still showed many glitches to be fixed and rethought. Specially related to consistency and standards of icons, types, information and actions.

Heuristic 4 was related to a significant amount of errors during the evaluation, and too present on major problems – problems that may diverge users from their objectives and can make users give up using the system if there is any other option out of it. Errors linked to heuristic 3 also called attention, as it showed many occurrences considered major and one catastrophic. Since the lack of user control and freedom is usually considered frustrating, by users, rethinking the approach for the noted errors was considered urgent.

Often related to minor and aesthetic severities, heuristic 8 surprisingly was linked to major problems and one catastrophic issue on this project. Although with relative urgency, the errors to be fixed usually do not present complexity.

Most of the fixed problems reported by the Heuristic Evaluation were validated with users going through the system on Think-aloud Protocol, but new problems were pointed during the new testing. Users showed to have problems finding the specific medication, with the bank slip printing process and errors in buttons related to the purchase process: “the biggest problem was not being able to find the medication using the search tool (homepage). But I found all information clear and succinct”, “the print slip button is not clickable. The cursor did not change to the hand-cursor shape” and “I tried to click on buy many times, but it seems that I should click on ‘one-click purchase’, instead. I though it was a problem with my browser, but no”.

Although many updated were validated, new recommendations surfaced on the last testing phase (Think-aloud Protocol) for the refinement of the system: (1) revise the vocabulary broadness of the search tool and optimize the informational structure, (2) make easier to users to recognize the function mini-cart, where costumers can view all their selected items before purchase, (3) limit to one button option at product description page to minimize purchase action doubt to users and (4) shipping cost have to be clear at all times. But (5) the checkout process follows a standard flow in the system and can’t be redesigned. The best direction is to design it more related to the brand standards.

The development of the project involving different research methods with experts and final users on diverse phases proved to be efficient to avoid development misleads and it saved unnecessary expenses during the project. Although the development through integrated phases might seem to take longer time, the whole process was essential to approximate users' expectations to stakeholders' strategies and the brand. Results from each phase were presented to stakeholders for a closer integration of the Relation Triad and be sure that taken steps were validated by all actors involved. In order to compensate the number of phases, the whole proceedings were agile executed and each method was applied in a faster pace to attend the market timing expectations. The future steps of the project intend to expand interaction propositions to the physical branch of the business.

Acknowledgements. For this project, all research phases were executed in partnership with UFF's Industrial Design research lab – TDT – in attempt to bring together a more integrated relation between business, research and education. Original.io is an UX Design agency located in Rio de Janeiro since 2011, with projects focused on UX Design and Omnichannel strategies for retail companies of diverse areas (fashion, wellness, supermarkets, malls etc.). The agency encourages close relations with local universities, such as ESPM, PUC-RJ and UFF.

References

1. Anderson, C.: The Long tail. Wired magazine (2004). http://www.wired.com/wired/archive/12.10/tail.html
2. Renzi, A.B.: Usabilidade na procura e compra de livros em livrarias online. Master's Dissertation, PPDESDI, UERJ, Rio de Janeiro, RJ, p. 298 (2010)
3. Guimaraes, M., Renzi, A.B.: Usability and interaction evaluation on breakfast delivery mobile app: users' experience expectations. In: Advances in Usability, User Experience and Assistive Technology, pp. 564–573. Springer, Orlando (2018)
4. Mucchielli, R.: O questionário na pesquisa psicossocial. Martins Fontes publishing, São Paulo (2004)
5. Padovani, S., Ribeiro, M.A.: Card sorting: adaptation of the technique for application to non-digital information system design. Revista Brasileira de Design da Informação/Braz. J. Inf. Des. **10**(3), 293–312 (2013)
6. Agner, L.: Ergodesign e arquitetura de Informação – trabalho com usuário. Ed. Quartet, Rio de Janeiro, RJ (2012)
7. Nielsen, J.: 10 usability heuristics for interface design. www.nngroup.com/articles/ten-usability-heuristics (1995)
8. de Villanueva, R.A.: Think-aloud protocol aril heuristic evaluation of non-immersive, desktop photo-realistic virtual environments 2004. Dissertation (Master of Science) – University of Otago, Dunedin – New Zealand (2004)
9. Renzi, A.B., Freitas, S.: Aplicação de Think-aloud Protocol em teste de usabilidade na procura de livros em livrarias online: recomendações de leitores. In: 10º USIHC – Congresso Internacional de ergonomia e usabilidade, Conference Proceedings, Rio de Janeiro (2010)
10. Miranda, F.: Estudo ergonômico de websites de comércio eletrônico: seleção do produto pelo usuário no processo de compra, Master's dissertation – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 313 p. (2005)

11. Banati, H., Bedi, P., Grover, P.S.: Evaluating web usability from the user's perspective. J. Comput. Sci. **2**(4), 314–317 (2006)
12. Renzi, A.B., Freitas, S.: Usabilidade e fatores de confiança na procura e compra de livros em livrarias on-line. In: Textos Selecionados de Design, PPDESDI, Rio de Janeiro, RJ, vol. 3. pp. 9–40 (2013)
13. Chau, P.Y.K., Hu, P.J., Lee, B.L.P., Au, A.K.K.: Examining Customers' trust in online vendors and their dropout decisions: an empirical study. Sci. Direct Electron. Commer. Res. Appl. **6**, 171–182 (2007)
14. Renzi, A.B., Chammas, A., Agner, L., Greenshpan, J.: Startup Rio: user experience and startups. In: Design, User Experience, and Usability: Design Discourse, Los Angeles, CA, pp. 339–347. Spring (2015)
15. Abras, C., Maloney-Krrichmar, D., Preece, J.: User centered design. In: W Encyclopedia of Human-Computer Interaction. Sage Publications, Thusnd Oaks (2004)
16. Six, J.M.: Fundamental principals of great UX design—How to deliver great UX design. In: UX matter (2014). http://www.uxmatters.com/mt/archives/2014/11/fundamental-principles-of-great-ux-design-how-to-deliver-great-ux-design.php
17. Sherwin, K.: Card sorting definition. In: NN Group (2018). https://www.nngroup.com/articles/card-sorting-definition/

Social Event Management System: Users' Recommendations

Thiago Gomes da Silva[1(✉)] and Adriano B. Renzi[2]

[1] Serviço Nacional de Aprendizagem Comercial, Rio de Janeiro, Brazil
tgsarvore@gmail.com
[2] Industrial Design – TDT, Universidade Federal Fluminense, Niterói, Brazil
adrianorenzi@gmail.com

Abstract. Social meetings and events pose a major part in Brazilian culture and have great importance throughout various interaction contexts: from daily neighborhood social activities to workplace relations. In a country well known by its festivities, to gather people, from small parties to huge events, can bring a certain social impact. Since the beginning of the 3rd wave of computing, the use of digital apparatuses and social medias became gradually part of the Brazilian social processes and directly affect their daily social life and work [1]. Technology has been changing social behavior and the way people interact with each other, with less direct interaction, but never the less connected. This research is part of the project "RatatApp", a system to help organize collaborative events that can encompass multiple administrators and manage multiple events, users, schedules and finance. Although the whole research comprises many phases, this paper presents the high fidelity prototype-testing phase.

Keywords: User experience · Management systems · Human-computer interaction

1 Social Meetings and Digital Culture

Social meetings and events pose a major part in Brazilian culture and have great importance throughout various interaction contexts: from daily neighborhood social activities to workplace relations. In a country well known by its festivities, music and carnival, to gather people, from small parties to huge events, can bring social impact, in diverse proportion. And to plan and organize an event that involves different people, with different interests, can be a hard task, even if managing just an alumni barbecue reunion.

Friendships are built with time, and according to Silva [1], besides "permanency, affinity and affection, the three basic factors for long-term friendships, to have shared experiences with friends is a key coefficient to make the bond stronger". In which, Norman [2] add: "we are social creatures, biologically prepared to interact with others, and the nature of these interactions depends on understanding others' emotions".

From the beginning of the 3rd wave of computing (many computer for one user), the use of digital apparatuses and social medias become gradually part of the Brazilian social processes and directly affect their daily social life and work [3]. Technology is

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 535–546, 2020.
https://doi.org/10.1007/978-3-030-20040-4_48

changing social behavior and the way people interact with each other, with less direct interaction, but never the less connected. As Norman [4] indicates, "The confusions foisted upon us by technology are increasing at a faster pace than ever before. Today's heavy usage of the Internet, cellular telephones, portable music players, and the wide variety of portable, wireless message and e-mail systems shows just how important these technologies have become to our lives".

As interaction becomes more and more pervasive, every touchpoint with a digital apparatus is part of a whole user experience journey [5], integrating people within contexts of use. More than ever before, digital interactions are part of users' story and technology shapes users' behavior, as much as users shape technology advancements. From Turkle's [6] view of this relation, when people exchange instant messages, e-mails and tweets, the technology reshape the boundaries of intimacy and solitude. Youngsters avoid phone calls afraid of revealing too much of themselves and prefer text messages to each other. Adults also prefer to use the keyboard instead of their voice, substantiated of being more efficient to communicate, as real time actions take too much time. As people gets more attached to technology, the unplugged world means less and unsatisfying.

Silva [1] steps up the hypothesis that face-to-face gatherings are of utmost importance to friendship maintenance, as 3 factors seem to help determine a long-lasting friendship: permanence, affinities and affection. Permanency is how long friends know each other; Affinity is about friends shared experiences and their story together; and Affection the nurture of stronger caring and love, built from joined experiences. Nevertheless, as technology has gradually been growing into our daily interactions, modifying the way we work, perceive our environment and socially relate, more customized communication tools surface to providing broader possibilities to gather people. As Castelli [7] indicates, what the Internet does is process the virtuality and transforms it into our reality, constituting the network society, which is the society in which we live.

The advancements of technology through the 3rd wave of computing bring forward new possibilities of interaction, where computers are integrated in one dynamic system ecology and transforming human-computer interaction into human-information interaction [5]. Digital artifacts are not isolated interactive devices, as each one is part of an ecosystem and the user experience is build through pervasive interaction with diverse touchpoints, part of the system ecology, and blended spaces [8, 9].

This research is part of the project for the event management system "RatatApp", a system to help organize collaborative events. The system can encompass multiple administrators and helps manage multiple events, and their related users, schedules and finance. Based on previous interviews with users and mental model mapping, the use of smartphones was presented as the most common artifact used for social interaction. And so, for this research, the prototype and usability testing were all focused on smartphone usage.

2 RatatApp: Mental Model and Structure

Based on precedent studies regarding social technology relations, a semi-structure questionnaire was applied to users in order to better understand their mental model, necessities and expectations for an event management system. All selected participants were users with frequent experience in organizing social events.

The questionnaire was structured based on Mucchielli's [10] studies on psycho-social surveys and focused on mapping the whole managing process, the factors of influence, the difficulties involved in event administration and users' needs (managers and invitees).

The answers from the selected participants surfaced important points to be considered for the development of the project:

- The finance involved must be clear and suitable for all involved, as this is considered a factor that could determine the number of participants. For a non-profit event (a friends' reunion for example) the numbers should be clear to everyone.
- The process of finance administration (for friends events) was unanimous to have organizers centralizing the full expenses, based on the number of confirmed participants, and later collect dividends from the participants.
- The conciliation of time availability among participants is a big issue and the system should present schedule easily manageable by either the main organizers or all involved (if the event if for a small number of people).
- The easiest device to check and participate in the process of preparing the event is the smartphone, so the system should be developed for smartphone use.
- Minimize the steps of the event process as much as possible. The system has to be direct to the point and easy to access.

The interaction structure of the smartphone app was based on similar event management systems, information architecture theories of Resmini and Rosatti [11] and Agner [12] and interview results (Fig. 1).

From the information structure strategy, a high fidelity prototype was developed to experiment and further analyze users' interactions. The prototype was built with InVision in order to provide easy online access through smartphones and better simulate the final product.

The testing of the high fidelity prototype usability was conducted using two simultaneous methods: Heuristic Evaluation and Cooperative Evaluation, as these two different methods could provide important data for comparison. The Heuristic Evaluation, as it follows a set of usability criteria, allows a broader vision of the project based on experts' point of view. The Cooperative Evaluation brings users' point of view from direct experience with the app to help analyze the project's usability and interaction problems.

Fig. 1. Basic structure of the app: group and event creation tasks and finance and schedule management.

3 Heuristic Evaluation and Cooperative Evaluation

When new necessities derived from the third wave in computing, researchers adapted the 10 usability heuristics of Nielsen and Molich [13] to new contexts of interaction and new devices. As smartphones becomes more present in people's daily routine, Inostroza *et al*. [14], in 2012, presents twelve heuristics for smartphone interaction. The researchers adapt the precedent Nielsen's ten usability heuristics for desktop to the new interactive perspectives and display size of the smartphone. Most of the 10 heuristics are easily applied to the smaller device, but three heuristics had to be more focused to the new challenge: the heuristic seven (flexibility of use) is expressed as "customizations and shortcuts", a new eighth heuristic focus on efficiency and performance of use, and a twelfth heuristic is added for the smaller display size of smartphones and the context of interaction with fingers: physical and ergonomic interaction (Table 1).

Table 1. Comparison of the heuristic sets for desktop and smartphone – Nielsen and Molich and Inostrozza *et al*.

Nielsen e Molich (1990)	Inostroza (2012)
Desktop	Smartphone
Usability of graphic interfaces	Usability of graphic interfaces
1. Visibility of system status	1. Visibility of system status
2. Match between system and the real world	2. Match between system and the real world
3. User control and freedom	3. User control and freedom
4. Consistency and standards	4. Consistency and standards
5. Error prevention	5. Error prevention
6. Recognition rather than recall	6. Minimize the user's memory load
7. Flexibility and efficiency of use	7. Customization and shortcuts
8. Aesthetic and minimalist design	8. Efficiency of use and performance
9. Help users recognize, diagnose and recover from errors	9. Aesthetic and minimalist design
10. Help and documentation	10. Help users recognize, diagnose and recover from errors
	11. Help and documentation
	12. Physical interaction and ergonomics

The Heuristic Evaluation consists of 3–5 experts going through a system to analyze and evaluate its interactions based on a set of heuristics. According to Nielsen's experiment in 1992 [15] with 19 usability experts, the use of heuristic evaluation showed that using 3–5 experts could bring up around 75% of the total usability errors (Fig. 2). For this project, the Heuristic Evaluation using smartphone [14] encompassed 3 usability specialists. Each specialist accessed the online prototype using their own smartphone, as a different device could be unfamiliar and bring factors to distort the results. Since the project is based on smartphone interaction, the evaluation was based on Inostroza's proposal of 12 heuristics for smartphone.

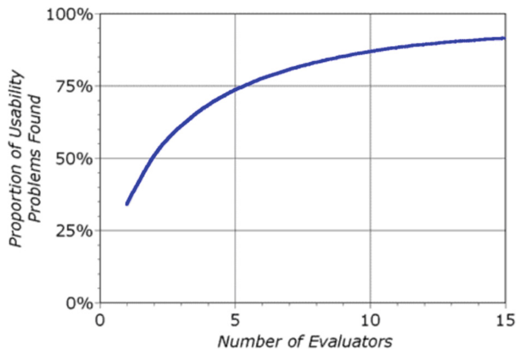


Fig. 2. Nielsen's relation between number of experts and percentage of errors found [13].

The first phase consists of each expert going through the system to analyze and evaluate its interactions based on Inostroza's set of heuristics. The experts evaluate the system separately, with no exchange of information between them, and present a report with all notified issues. Each notified usability problem is related to one or more heuristics. The 3 experts evaluations are compiled in one final report by the main researcher and the pointed heuristic problems are categorized by each expert in degrees of gravity, from 0 to 4:

– Gravity 0 (zero): It is not a problem. Usually pointed to parts of the system or interface that could be considered awkward, but does not get in the way of users.
– Gravity 1 (one): It is an aesthetic problem. Pointed mostly to parts of the system related interface layout that may not follow users expectations or are considered an aesthetic flaw, but do not affect users interactions or interoperability of information.
– Gravity 2 (two): It is a small problem. These are problems that cause discomfort to users, may bring misunderstandings of information, result in interaction or action confusion, oblige users to re-do or take longer actions, or distort the notion of place-making. These problems do not impede procedures nor may cause serious interaction problems, but certainly are problems that bring discomfort and irritability to users.
– Gravity 3 (three): It is a big problem. These are problems that may diverge users from their objectives. They can take forms such as misleading information or interactions, actions requirements that are not obvious or are completely different from other channels, lack of interoperability between channels, information or action break, functionalities that are far from users expectations or do not follow a set pattern or have low affordance. These problems can make users give up using the system if there is any other option out of it.
– Gravity 4 (four): It is a catastrophic problem. These are problems that may impede users to reach objectives or compromise the whole system structure or interfere seriously in users' sense of Place-making. Mostly, these are problems instated in the system's concept, structure, taxonomy, interoperability and information system that affects the whole idea of cross-channel interaction, or even just isolated interactions. Serious problem can bring companies to re-think the whole idea of a product.

In Cooperative Evaluation, the researcher selects users to analyze a system. Each user, separately, fulfills pre-defined tasks by the researcher and verbalizes issues and doubts during the process [16]. The surfaced issues are discussed with the researcher, who usually sums questions with the purpose of deepening further discussed subjects and gather data for the analysis. It is important to be full aware of users' mistaken actions and misunderstanding perception in order to pinpoint problems and elucidate solutions from users' perspective. User and researcher evaluate the system together and the participants are encouraged to ask questions about the purpose and interaction of the product.

Miranda and Moraes [17] recommend the use of Cooperative Evaluation in products that need improvement and in prototypes while in intermediary and final phases. According to the authors, the evaluation follows 4 basic steps:

- 1. Recruit participants who are representative of final users
- 2. Plan and prepare tasks for the participants to execute in order to let them explore the whole system or at least the parts considered most relevant
- 3. Record and make notes of all actions, questions and discussions during the execution of tasks by users
- 4. Analyze the collected results.

The Cooperative Evaluation was conducted with the help of 3 users, with their actions written and audio recorded. All 3 selected participants had frequent experiences in organizing events. The context of use was presented to all participants and all doubts clarified before starting each session. There were 4 proposed sequential tasks:

- 1. Create a new group
- 2. Create an availability schedule
- 3. Create a new event
- 4. Create a new expense in the group and divide the finance among group participants.

4 Results

The data analysis compiled both Heuristic and Cooperative Evaluations. Any user observation, collected though Cooperative Evaluation, that haven't been noted by the experts on Heuristic Evaluation, were classified based on Inostroza's heuristics for better comparison and analysis.

A total of 55 problems were reported (Fig. 3) and most problems pointed by the specialists were related to heuristics 4 (consistency and standards), 5 (error prevention), 6 (Minimize the user's memory load) and 7 (Customization and shortcuts), linked to 9–10 errors, each. Heuristics 2 (Match between system and the real world) and 3 (User control and freedom) were related to 4–6 errors. The rest of the heuristics were associated to small number of occurrences. The heuristic 12 (Physical interaction and ergonomics) was not associated to any of the pointed problems.

Catastrophic problems (Fig. 4) were mostly related to heuristics 3 (user control and freedom) and 7 (Customization and shortcuts). As an example of catastrophic problem

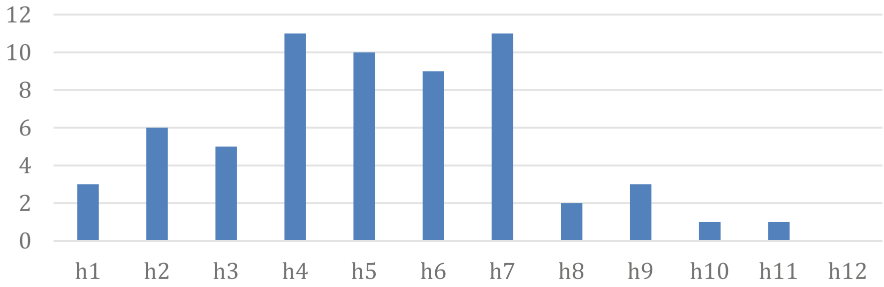


Fig. 3. Reported usability problems related to 12 heuristics for smartphones [14].

related to heuristic 7, while users were using the creating group page it was hard to understand the difference between the options "quantity" and "per person". The confusion surfaced doubts about the necessity at all of having a quantity field, as the action of including new members, the system automatically adds to the quantity of participants of the group.

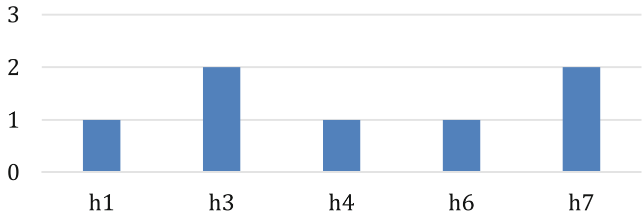


Fig. 4. Catastrophic problems related to heuristics for smartphone [14].

Although heuristic 3 was related to a small number of problems, these few problems were considered to be issues that could impede users to fulfill tasks or increase the chances of quitting, as the heuristic is cited for catastrophic and major problems: as users open the lateral menu, they could not close it back (heuristic 3). Users also accused frustration with the lack of back option and cancel an action while interacting with the system (heuristic 3).

The problems considered major issues were mostly related to heuristics 5 (error prevention) and 7 (Customization and shortcuts), such as the lack of examples in how the functionalities worked in the app (heuristic 5) and the creation of events not presenting an area to determine the hour and duration of the event (heuristic 7). Heuristics 2, 4 and 6 have also been associated to major problems, such as the order of the items in the event creation area, whose specialists suggested it should be changed (heuristic 2), the color inconsistency between buttons of same function throughout the app (heuristic 4) and misunderstanding the label of some buttons in the Sign In area (heuristic 6) – Fig. 5. Regarding major problems, no issue was reported related to heuristic 8.

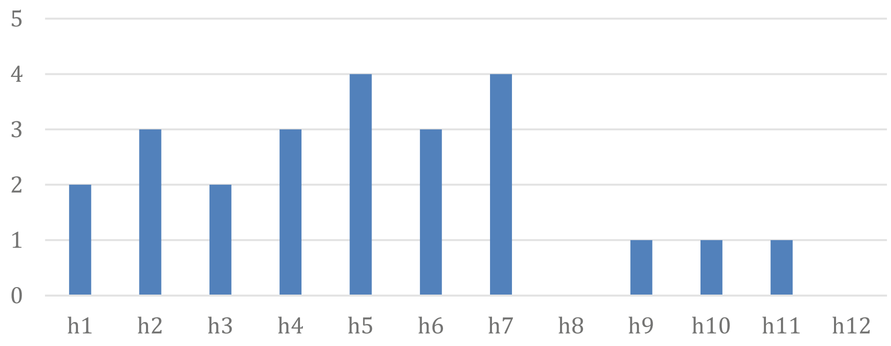


Fig. 5. Major problems related to 12 heuristics for smartphones [14].

The 21 problems considered minor issues (Fig. 6) were mostly related to heuristics 4 (consistency and standards) and 5 (error prevention). Variation of font size identified at the home screen (heuristic 4) and the presence of both "check" and "confirm" buttons on different pages, which could confuse users (heuristic 5) are examples of he cited heuristics.

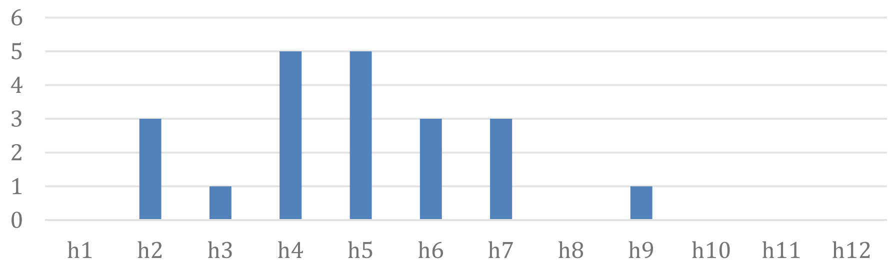


Fig. 6. Minor problems related to 12 heuristics for smartphones [14].

The majority of reported usability issues were classified as major and minor problems, as normally expected on Heuristic Evaluations. But, for this evaluation, a high number of major problems surpassed the total of minor problems.

5 Conclusion

The reported problems from both methods helped fix important issues in the high fidelity prototype. Although results from the interviews and benchmark research lead the basic structure and visual design of the project, only after the evaluations was possible to identity usability issues and interaction misleads that were not considered before. The results surfaced the necessity of changes in the project to better reach users' expectations. Icons and fonts were fixed to keep the consistency and bring a better perception of functionality, as well as, all considered dead end interactions were modified for better users' control. Features of creation and management of events were reformulated to make the process more obvious and easier to edit.

Customization and shortcuts (heuristic 7) was the most cited heuristic, related to both catastrophic and major problems. The constant relation emphasizes users' interest in fulfilling tasks and reaching objectives with few steps through their own interaction and decision paths. The constant appearance of the heuristic in reports and feedbacks from specialists and users helped turn the experience process more intuitive and more direct to the point. Figure 7 presents two examples of modifications from heuristic 7 problems: the staring point to create a new group or event, or edit the program schedule became more intuitive direct and with less steps to conclude the tasks; the creation of a new event has less information to fill out and became easier to manage, reducing also the steps of the process.

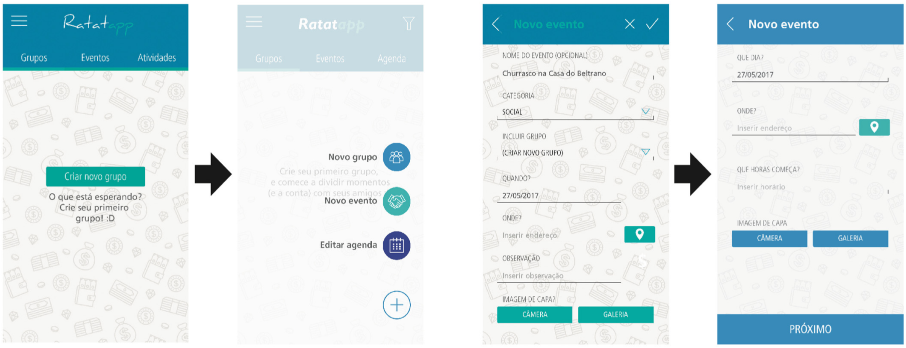


Fig. 7. Reports of problems related to heuristic 7 (customization and shortcuts) brought changes to the project to make the user experience more intuitive and direct to the point, diminishing the number steps to fulfill tasks, considered too much by users and specialists.

The two different evaluating methods showed different perspectives for the prototype testing: (1) the Heuristic Evaluation, with usability specialists, collected technical inputs for the structure, context of use and system general functionality. The results surfaced problems that affected the system interaction in the whole experience strategy; (2) the Cooperative Evaluation, with final user, brought up points important to users, more focused on interaction easiness and experience flow, evincing points that should be more intuitive. Users' collaboration and their reported issues indorsed the importance of fixing the errors related to the heuristic 7, in order to make the whole experience direct to the point and flexible to customization.

Although the Heuristic Evaluation found more problems than the Cooperative Evaluation, the choice of using both methods was assertive, as their results showed to be complementary. The 2 evaluations results helped fix many usability problems that were not predicted in earlier stages of the project and better approach the app to users' expectations for a better experience journey construction. However, although with problems fixed, the high fidelity prototype was once more evaluated with users to validate the modifications and analyze new reactions with the app.

After the prototype was adjusted to be closer to users' expectations and needs, new tests were conducted to validate the applied changes and check for new usability problems. As expected, new interesting feedback surfaced from further testing, keeping the project development process dynamic, as pointed by Resmini and Rosatti's manifest [11]. Although this paper focus on the prototype testing phase, the development of the project had a broader range of action to understand the role of the app and digital devices within the full experience journey [3, 5, 8] of planning and managing a social event with friends. The most lucrative aspect to all students involved in the project certainly was to experience the whole process from start to finish, encompassing research, development, testing, adjustments and validating.

References

1. da Silva, B.G.R.S.: A Amizade em Tempos de Tecnologia. Paco Editorial, Jundiaí (2016)
2. May, P., Ehrlich, H.C., Steinke, T.: ZIB structure prediction pipeline: composing a complex biological workflow through web services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)
3. Renzi, A.B.: Experiência do ususário: a jornada de Designers nos processos de gestão de suas empresas de pequeno porte utilizando sistema fantasiado em ecossistema de interação cross-channel. These de doutorado, 239 p. Escola Superior de Desenho Industrial, Rio de Janeiro (2016)
4. Czajkowski, 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)
5. Renzi, A.B.: Experiência do usuário: construção da jornada pervasiva em um ecossitema. In: Proceedings SPGD 2017, Rio de Janeiro, vol. 1 (2017)
6. Turkle, S.: Alone Together: Why We Expect More from Technology and Less from Each Other. Basic Books, New York (2011)
7. Castelli, M.: Internet e sociedade em rede. In: MORAES, D. (org.) Por uma outra comunicação: mídia, mundialização cultural e poder. Record, Rio de Janeiro (2003)
8. Sande, A., Schinaider, S., Renzi, A.B.: Experience, usability and the sense of things. In: Design, User Experience, and Usability: Designing Pleasurable Experiences. HCI International 2017 Proceedings, Part II, Vancouver, BC, pp. 77–86 (2017)
9. Renzi, A.B.: The use of task-flow observation to map users' experiences and interaction touchpoints. In: Proceedings of the Advances in Human Factors and Systems Interaction, AHFE 2019. Springer, Orlando (2018)
10. Mucchielli, R.: O questionário na pesquisa psicossocial. Martins Fontes Publishing, São Paulo (2004)
11. Resmini, A., Rosatti, L.: Pervasive Information Architecture – Designing Cross-Channel User Experiences. Morgan Kaufmann – Inprint of Elsevier, Burlington (2011)
12. Agner, L.: Ergodesign e arquitetura de Informação – trabalho com usuário. Ed. Quartet, Rio de Janeiro (2012)
13. Nielsen, J.: 10 usability heuristics for interface design (1995). www.nngroup.com/articles/ten-usability-heuristics
14. Inostroza, R., Rusu, C., Roncaglioso, R., Rusu, V.: Usability heuristics for touchscreen-based mobile devices. In: 9th International Conference on Information Technology. IEEE Computer Society, Temuco (2013)

15. Nielsen, J.: How to conduct a heuristic evaluation (1995). https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/
16. Teixeira, E., de Moraes, A.: Avaliação cooperativa da interface de sites hipermídias focados na “banda larga”. In: Congresso Internacional de Ergonomia e Usabilidade, Design de Interfaces e Interação Humano-Computador, Rio de Janeiro, vol. 3 (2004)
17. Miranda, F., Moraes, A.: Avaliação da interface de um site de comércio eletrônico através da técnica avaliação cooperativa. In: Anais 2º USIHC, Rio de Janeiro (2003)

User Experience Interaction in Cross-Channel Systems: Comparing Heuristic Evaluations

Adriano B. Renzi[1(✉)], Olnei Almeida[2], and Marcos Cesar[3]

[1] Universidade Federal Fluminense Industrial Design – TDT, Niterói, Brazil
adrianorenzi@gmail.com
[2] UX Department, Banco Original, São Paulo, Brazil
nei.fda@gmail.com
[3] Digital Department, Packaging Brands, Rio de Janeiro, Brazil
marcccos@gmail.com

Abstract. Differently to usability metrics, which tends to focus on specific tasks with specific digital apparatuses and contexts, the pervasive experience permeates different digital devices to build an experience journey with many interaction touchpoints. Each interactive device play a part, sometimes complementary, to help users fulfill their needs or just generate an experience. As user experience develops to pervasive cross-channel interaction, it becomes difficult to relate ecology systems UX problems to formal usability heuristics. A set of UX heuristics was proposed in 2016 and first tested in 2018 with the multi-channel Airbnb system through pervasive testing on 3 different devices: desktop, tablet and smartphone. This paper presents a comparison analysis of heuristic evaluations, using Nielsen's desktop/tablet heuristics, Inostroza's smartphone heuristics and Renzi's UX cross-channel heuristics.

Keywords: User experience · Pervasive interaction systems · Cross-channel

1 Pervasive Interaction

Since the 3rd wave of computing has emerged in the beginning of this century, designing interactivity has been gradually changing, as new possibilities of interactions and technology have surfaced and mental models and expectations have evolved. The diversity of digital apparatuses in our daily lives has been aiming and directed for an interoperability merge in an ecology of systems [1], where the ecosystem can be one system, integrating many devices and many systems, to create a journey of sequential events. Digital devices are more integrated into our daily chores as wearables and interactive objects, in a internet of things (phones, watches, glasses, refrigerators, cars, makeup, papers, toys, home, etc.), suiting the development path expected for the next 10 years [2]: "In 5 years, technology will start to disappear from sight and be more integrated to objects and accessories around us. Wearables will be more and more common (glasses, watches, key holders, bracelets, shoes, hats and coats). The processing will be on the server's side and the portable artifacts will serve as data inputs and outputs. In 10 years, all objects and artifacts will be imbued with digital interaction. Smart glasses and wearable accessories will be commonly used among people. The

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 547–558, 2020.
https://doi.org/10.1007/978-3-030-20040-4_49

computer ubiquity will have reached its maturity and smart/electronic homes (houses, building, offices) will be an important tool for this evolution, saving time, resources and energy. The paradigm of product purchase will gradually be substituted by the idea of product use (rental), as well as the notion of acquiring being substituted by the notion of usage. Walls around us will be touch screen displays, creating the possibility of total customization of environments. Everything that can be digitally interactive, it will be. Interaction will transcend tablet and smartphone devices and will be integrated to the environment around us or projected on surfaces".

The Internet of Things is a system of systems, interconnected at different scales, with the information architecture conceptualization of cross-channel ecosystems [3]. As the interoperability between systems gets more accurate, the user experience journey detach from isolated devices and becomes a pervasive interaction, crossing many interactive devices and environments. From the architecture information's point of view, Benyon and Resmini [4] bring forward that the cross-channel experience understands that actors (users) interact with information across a multitude of devices, platforms, locations, and contexts, resulting in complex, activity-based ecosystems that are experienced as a continuous flow or movement across digital and physical space, and as an "ecosystem resulting from actor-driven choice, use, and coupling of touchpoints, either belonging to the same or to different systems, within the context of the strategic goals and desired future states actors intend to explicitly or implicitly achieve" [3]. Bødker and Kolomose [5] have presented the idea of an increasingly interaction with multiple interactive artifacts with overlapping capabilities during users daily activities. The authors, at first, called this integrated interaction an artifact ecology, as the use of an interactive artifact cannot be understood in isolation and each artifact influence the use of others. Each device used, affects users' perception, as defined by Kirsh's Theory of embodied cognition [6], provides a first clue of things to come. When people heft a tool the neural representation of their body schema changes as they recalibrate their body perimeter to absorb the end point of the tool [7]. As mastery develops, the tool reshapes the users' perception, altering how they see and act, revising their concepts and changing how they think about things.

The embodied cognition applies the same way to digital devices and cross-channel interactions, as the interoperability between systems brings users to feel devices, features and the whole experience as an extension of their bodies and abilities. The understanding, adapting and mastery of the interactions within devices and the experience journey change users perceptions, mental model, expectations [8] and their sense of things [9].

The more and more present pervasive cross-channel interaction, bringing together different artifacts to help users fulfill an experience journey, the more difficult is to analyze the experience focusing on isolated devices. The development of new digital apparatuses, with new interactions and new display sizes, brings the necessity of new heuristics to better guide developers and analyze usability focused on each new proposed device.

2 Heuristics Evolution

In 1990, Nielsen and Molich developed a set of usability principles to be considered when planning a system with a visual interaction interface [10]. The ten principles were well known as the ten usability heuristics and became a base for a heuristic evaluation method, to evaluate a system's usability and point out problems. The heuristics are consecution to users' needs in a time where the world was moving to the second wave of computing (one computer to one user). As Nielsen and Molich, other authors developed similar principles to help evaluate usability and interaction with interfaces: Bastien and Scapin [11] developed eight ergonomic criteria in 1993 for systems' interface evaluation and Schneiderman [12], in 1986, based on his research regarding human-computer interaction presented eight golden rules for user interface.

When new necessities derived from the third wave of computing (many computers to one user), researchers adapted the 10 usability heuristics of Nielsen and Molich to new contexts of interaction and devices. Apted *et al.* [13], analyzing tabletop's possibilities of interactions, focused on the difference of display size, its proposal of collaborative interaction and reorganization of elements, proposed 7 heuristics that takes human reach, physical ergonomics and multiple interaction points into consideration. Inostroza *et al*. [14] in 2012 presents twelve heuristics for smartphone interaction. Based on Nielsen's ten heuristics, Inostroza *et al*. [14] express the heuristic seven (flexibility of use) as "customizations and shortcuts", specify a new eighth heuristic regarding efficiency and performance of use, and add a twelfth heuristic regarding the smaller display size of smartphones and the context of interaction with fingers: physical and ergonomic interaction. When Neto and Campos [15], two years after Inostroza, present their 12 heuristic principles for multi-modal interactive ambiences (a combination of inputs and outputs of several sensory modalities – hearing, smell, taste, touch, sight – as part of a more natural computer communication), it is perceptive the strong relation with the ten heuristics from Nielsen and Molich, added by important points on physical ergonomics and vocal commands (organized content, direct manipulation, human range).

However, these proposed principles focus on direct use with specific devices, making it hard to use on pervasive cross-channel scenarios. The five principles of Resmini and Rosatti [1] encompass the cross-channel concept from the information architecture perspective and are the perfect start, due to its pervasive cross-channel characteristics, to the idea of dynamic interaction ecology. But when relating to usability and UX evaluation, the five concepts have to expand, considering user experience concepts and its relation to visual cognitive recognition, physical and vocal interactions, as perceived in a previous research evaluating the GloboPlay system interaction [16].

In order to analyze the whole experience journey and the cross-channel interaction within one integrated ecology system, a set of 9 UX heuristics [17] was proposed in 2016, encompassing interface cognitive perception, 5-sense interaction, pervasive information and interoperability between systems (Table 1).

Table 1. Comparison of the three heuristic sets used in this research project – Nielsen and Molich [10], Inostrozza [14] and Renzi [17]

Nielsen e Molich (1990)	Inostroza (2012)	Renzi (2016)
Desktop	Smartphone	Cross-channel
Usability of graphic interfaces	Usability of graphic interfaces	Usability of graphic interfaces, pervasive information and 5-sense interactions
1. Visibility of system status	1. Visibility of system status	1. Place-making
2. Match between system and the real world	2. Match between system and the real world	2. Consistency
3. User control and freedom	3. User control and freedom	3. Resilience
4. Consistency and standards	4. Consistency and standards	4.Reduction
5. Error prevention	5. Error prevention	5. Correlation
6. Recognition rather than recall	6. Minimize the user's memory load	6. Equivalency to cultural conventions
7. Flexibility and efficiency of use	7. Customization and shortcuts	7. Visual intuitive content
8. Aesthetic and minimalist design	8. Efficiency of use and performance	8. Natural, intuitive and direct interactions
9. Help users recognize, diagnose and recover from errors	9. Aesthetic and minimalist design	9. Contextual ergonomics
10. Help and documentation	10. Help users recognize, diagnose and recover from errors	
	11. Help and documentation	
	12. Physical interaction and ergonomics	

This paper intends to present the results of simultaneous heuristic evaluations of the Airbnb ecology system using different apparatuses of access. The company has over 3 million active places for renting in more than 191 countries and 65 thousand cities. Over the years, more than 100 million guests have used at least one of Airbnb's affiliated hosts. During the established period of research, the Airbnb system could be reached through desktop/laptop (website), Ipad (app) and smatphone (app). Due to these possibilities, three heuristic evaluations were conducted for comparison: (1) heuristic evaluation of Airbnb website using Nielsen's 10 heuristics [10], (2) heuristic evaluation of smartphone app using Inostroza's 12 heuristics [14], and (3) UX heuristic evaluation [18] using Renzi's 9 heuristics [17] permeating three devices (desktop, tablet and smartphone).

3 Heuristic Evaluations: Nielsen, Inostrozza and Renzi

The three heuristic evaluations of the Airbnb system were conducted following Nielsen's proposal, in order to keep the factors involved to a minimal of influence. Each type of heuristic evaluation was executed isolated, with no exchange of data before the final documentation. Three main researchers conducted the evaluations, but each one was responsible for just one heuristic evaluation type (desktop, smartphone, cross-channel). One research coordinator managed all three researchers and put together all data for analysis.

According to Nielsen's experiment in 1992 [19], with 19 usability experts, the use of heuristic evaluation showed that using 3–5 experts could bring up around 75% of the total usability errors (Fig. 1). The heuristic evaluations using desktop [10] and smartphone [14] encompassed 5 usability specialists, each one. For the UX cross-channel evaluation [17], 3 UX experts with diverse background agreed to participate. For the UX evaluation, the invitee had to be UX experts with previous experiences in evaluating systems using the traditional heuristic evaluation, as well as understanding the pervasive cross-channel concept.

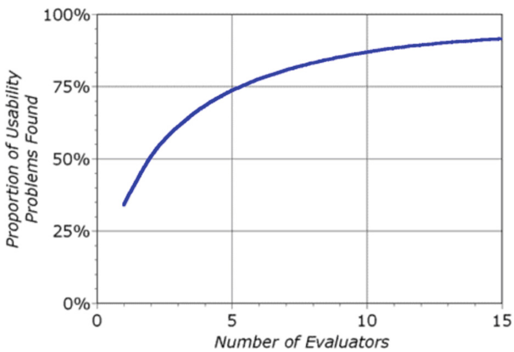


Fig. 1. Nielsen's relation between number of experts and percentage of errors found [19].

Following Nielsen's proposal, the heuristic evaluation consists of experts going through a system to analyze and evaluate its interactions based on a set of heuristics. However, since the UX heuristic evaluation has a broader analysis (encompassing all channels involved), an order for interaction evaluation was proposed to help the experts test the whole experience journey. The sequence of interaction followed previous findings from interviews [20], confirming Google's insights on travel costumer journey 4 micro-moments: dreaming moment (I want to get away), planning moment (time to make a plan), booking moment (let's book it) and experience moment (can't wait to explore). The evaluation tried to encompass the planning, booking and experience moments [18], with each specialist starting the analysis journey with a smartphone, continuing with a desktop/laptop, going through a tablet and finishing with a smartphone (Fig. 2). All pointed heuristic problems were categorized by the experts in degrees of gravity, from 0 to 4:

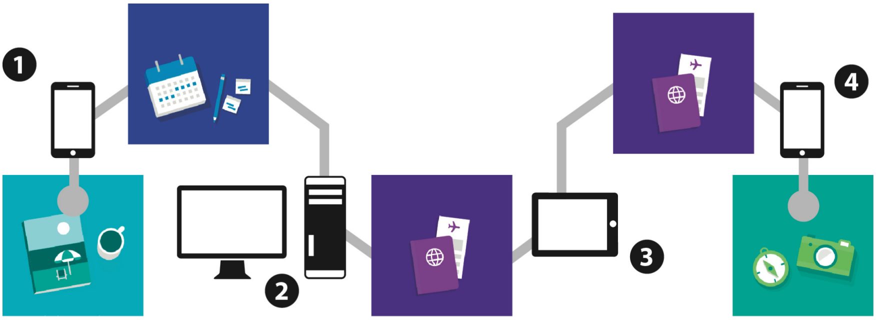


Fig. 2. Planned pervasive actions for evaluation, following users' travel journey [18].

- Gravity 0 (zero): It is not a problem. Usually pointed to parts of the system or interface that could be considered awkward, but does not get in the way of users.
- Gravity 1 (one): It is an aesthetic problem. Pointed mostly to parts of the system related interface layout that may not follow users expectations or are considered an aesthetic flaw, but do not affect users interactions or interoperability of information
- Gravity 2 (two): It is a minor problem. These are problems that cause discomfort to users, may bring misunderstandings of information, result in interaction or action confusion, oblige users to re-do or take longer actions, or distort the notion of place-making. These problems do not impede procedures nor may cause serious interaction problems, but certainly are problems that bring discomfort and irritability to users.
- Gravity 3 (three): It is a major problem. These are problems that may diverge users from their objectives. They can take forms such as misleading information or interactions, actions requirements that are not obvious or are completely different from other channels, lack of interoperability between channels, information or action break, functionalities that are far from users expectations or do not follow a set pattern or have low affordance. These problems can make users give up using the system if there is any other option out of it.
- Gravity 4 (four): It is a catastrophic problem. These are problems that may impede users to reach objectives or compromise the whole system structure or interfere seriously in users' sense of Place-making. Mostly, these are problems instated in the system's concept, structure, taxonomy, interoperability and information system that affects the whole idea of cross-channel interaction, or even just isolated interactions. Serious problem can bring companies to re-think the whole idea of a product.

4 Results from Simultaneous Evaluations

Since evaluations were applied separately, their respective results are presented apart for better comparison and discussion: desktop, smartphones and cross-channel experience.

The heuristic evaluation on desktops pointed a total of 51 problems. The heuristic 2 (match between system and the real world) was related to 10 problems and the most prominent among all heuristics, followed closely by heuristic 4 (consistency and standards), related to 9 usability problems. Heuristics 3 (user control and freedom), 5 (error prevention), 6 (recognition rather than recall) and 7 (flexibility and efficiency of use) had similar number of usability problems: 6 and 7.

Heuristics 1 (visibility of system status), 8 (aesthetic and minimalist design) and 10 (help and documentation) had a lower incidence, related to 3–4 usability problems. Heuristic 9 (help users recognize, diagnose and recover from errors) was related to only one problem.

As expected from the heuristic evaluation method, the majority of problems were considered of minor (27) and major (13) gravities. Catastrophic severity was related to 5 problems and 7 problems were classified as aesthetic issues (Fig. 3).

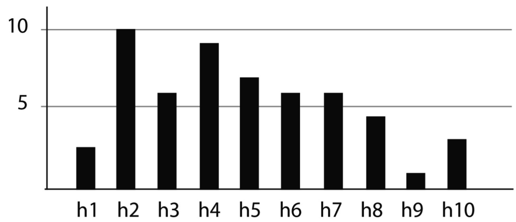


Fig. 3. Airbnb website usability problems reported by experts using desktop/laptop and relation with each heuristic.

Although the invited specialists were able to surface diverse usability problems on every step of the common search-compare-reserve process [18], the reservation phase presented problems that could lead users to give up their purchase. Some of these reported issues were classified as catastrophic, such as the payment slip generator not working and the information disappearance between steps of the reservation process. Evaluators also pointed as catastrophic problem the disappearance of user information whenever a process is interrupted or users try to go back to a precedent phase or homepage. The search results presented places not related to keywords was as well classified as catastrophic and some of the experts reported difficulty with filters.

Major and Minor problems were mostly related to icons significance, the use of map and its lack of location precision, the apt/room information, design inconsistencies and the half-translated texts.

The heuristic evaluation of the Airbnb smartphone app pointed a total of 59 usability problems. The heuristic 4 (consistency and standards) from Inostrozza heuristics was the most cited (17) related usability issues, followed closely by the heuristics 2 (match between system and the real world) and 8 (efficiency of use and performance), each related to 13 different pointed problems – Fig. 4.

Heuristics 1, 5 and 6 are similarly related to 7–8 usability problems and heuristics 3, 7, 9 and 11 to 4–5 problems. The 10th heuristic was related to just one problem and heuristic 12 was not cited by any of the experts (physical interaction and ergonomics).

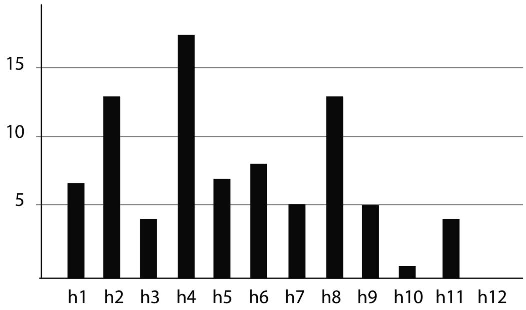


Fig. 4. Airbnb website usability problems reported by experts using smartphones and relation to heuristics.

It is important to notice that the heuristic 8, an adaptation for smartphone interaction evaluation, played an important role to help categorize issues with the device. The same occurs for the heuristic 7 – also adapted from Nielsen's proposal. Even if these pointed problems can be related to desktop heuristics descriptions, Inostroza's adaptation seems more appropriate for smartphone specifics. No usability problem was related with heuristic 12 (physical interaction and ergonomics).

Among all presented errors, a total of 25 problems have been considered to be minor by the experts, followed closely by 20 pointed major problems. Just 4 usability issues were classified as an aesthetic matter, but a higher incidence of catastrophic problems was reported with the smartphone interaction: 8.

The 8 catastrophic problems are related to various simultaneous heuristics, but the frequent relation to h5 (error prevention) and h8 (efficiency of use and performance), are connected to disastrous results of information disappearance and impediment of interaction to fulfill tasks.

The UX cross-channel evaluation pointed a total of 81 problems, distributed through smartphone, desktop and tablet. More than half (53%) of reported UX errors were connected to smartphone, while 35% was related to desktop and just 12% to tablet.

The most part of problems found was related to heuristic 7 (visual intuitive content), occurring 9 times with the Airbnb app for smartphones, 6 times for the desktop and 4 times while using the company's app for tablet. The gravity of problems related to heuristic 7 (visual intuitive content) was rated mostly as minor and major problems, as expected from heuristic evaluations. But 1 catastrophic problem related to this heuristic is reported while using desktop.

The heuristic 6 (equivalency to cultural conventions) is the second most reported heuristic, more present during the use of smartphones with 9 noted reports. Desktop use shows 5 problems and tablet presents just 1 issue. Although not as many errors as the visual intuitive content (h7), the experts considered 5 errors connected to this heuristic as catastrophic (gravity 4), with 3 errors using the smartphone and 2 for the desktop. These are mostly related to information inconsistency during pervasive interaction with different channels. Evaluators point as an example, the price of the saved apartment changing from desktop to smartphone. This was considered a catastrophic problem as it may directly infuse doubts about the security of the system.

Heuristics 1 (place-making) and 2 (consistency) follow closely, related to 12 and 10 UX errors, respectively. Both of them having more occurrences while using the smartphone: place-making (h1) appears 4 times on desktop, 7 times on smartphone and just once while using tablet; Consistency (h2) appears 3 times on desktop, 5 times on smartphone and twice on tablet. Heuristics 3, 4, 5 and 8 fall behind with similar number of errors (6–7). Except for heuristic 5 and 9, the rest of the heuristics follows the pattern of having the majority of errors related to smartphone use (Fig. 5).

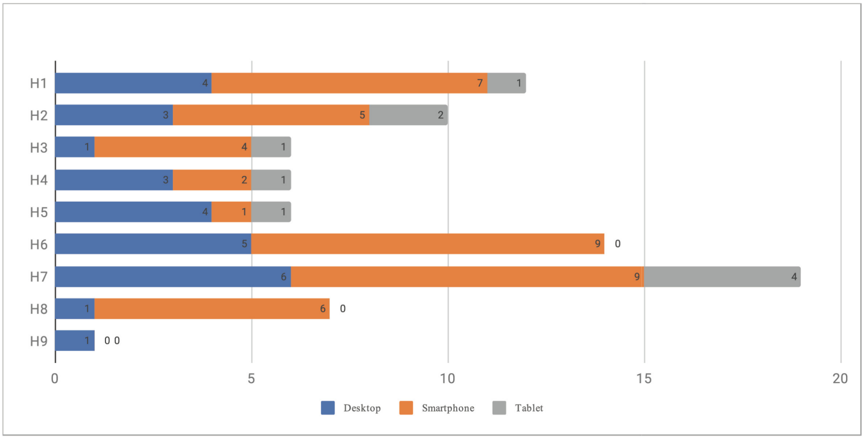


Fig. 5. UX heuristics related to each interaction device: desktop, smartphone and tablet.

Analyzing the pointed problems by their gravity, catastrophic problems (gravity 4) are connected to the use of smartphone app. Major problems (gravity 3) are frequently related to the use of the smartphone app and desktop. The tablet app only shows presence related to minor problems (gravity 2) – except for one (h5), but still is surpassed by both smartphone app and desktop experiences.

Part of the notified problems became more evident during the cross-channel interaction evaluation. As for example, while using the search tool on desktop it was not possible to find a place for rent, previously found using the smartphone Airbnb app. Not even rewriting the same words. The problem got worse when evaluators proceeded to step 3 (from desktop to tablet) and realized that it was not possible to retrieve information of the place, previously saved on smartphone and resaved on desktop. It was necessary to search once again the desired place and save it again. Evaluators considered this error through channels related to heuristics 1 (place-making) and 3 (resilience).

Another cross-channel problem example regards taxonomy and functionality recognition that affects users' sense of self-location within the system: different names are used in different devices, confusing users. Related to heuristic 1 (place-making) due to the loss of self-location, heuristic 4 (reduction), due to the system's unnecessary complexity of interaction, and heuristic 5 (correlation), due to taxonomy differences and lack of relation between representations of same functionality.

Noticed by evaluators in all 3 heuristic evaluations processes and present in all channels, while users visit apartment options, there is no information indication of neighborhood nor precision of locality, unless the renter insert in the title of the place. This was considered to affect users' experience journey and related to heuristics 5 (correlation) and 3 (resilience). Considering the importance detailed information about a place's location and its relation to mobility easiness, safety and distance to tourist points, from previous interviews with users [20], the lack of information about location was marked as a major problem and could completely invalidate places for renting.

5 Conclusions

The simultaneous heuristic evaluations helped to show that some errors are solely related to branched systems of specific devices and do not repeat throughout the UX journey. But also shows that part of noted problems appears in more than one apparatus (sometimes in all of them), integrated to the whole system ecology. For instance, the lack of location precision while analyzing places for rent was considered a severe problem by all evaluators. The catastrophic problem occurs in all system ecology, not to a specific device, for example.

The use of UX heuristics to evaluate the user experience full journey within the Airbnb system was able to pinpoint more errors (81) than the usability evaluation of each specific device (51 and 59 errors). Although the summation of both smartphone and desktop errors surpass the UX evaluation and the analysis of 3 integrated apparatuses naturally brings up more problems, many of these pointed errors repeat on all devices and were noted as one occurrence by the UX evaluation.

As shown in Nielsen's graphics [19], correlating "number of evaluators" and "number of errors found", having only 3 evaluators participating in the UX evaluation – as oppose to the 5 experts for each of the other devices (desktop and smartphone) – naturally surfaces a less number of problems. Therefore, part of the isolated device reported problems were not equally found during the UX evaluation.

Comparing the reported issues from the 3 groups of evaluators, it is perceptive that usability problems based on Nielsen's and Inostroza's respective heuristics can be re-categorized by the 9 UX heuristics. But not always the contrary. If all errors registered to solely the smartphone and desktop would be translated to equivalent UX heuristics, the total number of UX errors would increase at least 30%. As an example, the noted problem regarding the Airbnb logo being part of the fixed footer, near action icons in the smartphone app was related to Inostroza's heuristic 9 (aesthetic and minimalist design). The UX heuristics related to this problem would naturally be 7 (visual intuitive content). Similarly, the error "too many steps to reach the starting page", related to Inostroza's heuristic 8 (efficiency of use and performance), is equally represented by UX heuristic 4: reduction.

However, it is important to acknowledge that the heuristic 5 (error prevention), present in both Nielsen and Inostroza's set of heuristics, could be hard to have an equivalent in UX heuristics, depending on the context. For instance, the "payment slip generator not working", noticed on the desktop evaluation, is a system error hard to translate to a UX heuristic. The closest translation would be place-making (h1), as the malfunction of the system could leave users confused with the status of their UX journey and impeded of moving forward.

Commonly, systems and evaluations efficiency are compared based on number of errors, primarily during the second wave of computing, but this experiment surfaces important comparative information beyond numeric metrics. Analyzing the results of the 3 different groups, it becomes evident that:

- The adaption of Nielsen's heuristics into 12 heuristics to evaluate smartphone systems seems effective, as some pointed problems using the Airbnb app were more properly categorized. The utterly presence of heuristic 8 related to various problems is a good example of this.
- The 3 different evaluations brought up similar problems, confirming common issues of the system ecology, despite the variety of experts and the differences of heuristics, the majority of problems are related to similar concepts (Fig. 6).

Desktop (51 errors)	Smartphone (59 errors)	UX heuristics (81 errors)
h2 - Match between system and the real world h4 - Consistency and standards	h2 - Match between system and the real world h4 - Consistency and standards	h6 - Equivalency to cultural conventions h7 - Visual intuitive content
h3 - User control and freedom h6 - Recognition rather than recall	h1 - Visibility of system status h6 - Minimize the user's memory load	h1 - Place-making h2 - Consistency

Fig. 6. Airbnb website usability problems reported by experts using smartphones and relation to heuristics.

- Noted errors based on both isolated devices (desktop and smartphone) can be categorized to equivalent UX heuristics, but not necessarily the contrary. Since UX heuristics investigates experiences and information that goes between devices and the whole system ecology, some cross-channel concepts don't have equivalency in evaluations focused on isolated devices [17].
- Evaluating the Airbnb system ecology using UX heuristics, crossing different channels, surfaced problems that could not be traced by traditional heuristic evaluations, confirming previous research [17].
- All 3 sets of heuristics are effective to analyze systems and can utterly bring complementary results. Nevertheless, objectives, apparatuses involved and the context of research/project are imperative to choose the appropriate evaluation set of heuristics.
- It seems that (h5) error prevention could be translated to different UX heuristics, depending on the context of the problem. Nevertheless, it is important to further investigate the subject.

References

1. Resmini, A., Rosatti, L.: Pervasive Information Architecture – Designing Cross-Channel User Experiences. Morgan Kaufmann, Burlington (2011)
2. Renzi, A.B., Freitas, S.: Delphi method to explore future scenario possibilities on technology and HCI. In: Human-Computer Interactions International. DUXU, pp. 644–653. Springer. Los Angeles (2015)
3. Resmini, A., Lacerda, F.: The architecture of cross-channel ecosystems – from convergence to experience. In: Proceedings of the 8th International ACM Conference on the Management of Digital Ecosystems (MEDES 2016), Hendayne (2016)
4. Benyon, D., Resmini, A.: User experience in cross-channel ecosystems. In: 31st British Human Computer Interaction Conference, Sunderland, UK (2017)
5. Bodker, S., Klokmose, C.N.: Dynamics in artifact ecologies. In: Proceedings of NordiCHI 2012, pp. 448–457 (2012)
6. Kirsh, D.: Embodied cognition and the magical future of interaction design. ACM Trans. Comput.-Hum. Interact., IN (2011)
7. Làdavas, E.: Functional and dynamic properties of visual peripersonal space. Trends Cogn. Sci. **6**(1), 17–22 (2002)
8. Renzi, A.B., Freitas, S.: Affordances and gestural interaction on multi-touch interface systems: building new mental models. In: HCII 2014, Crete (2014)
9. Sande, A., Renzi, A. B., Schnaider, S.: Experience, usability and the sense of things. In: Design, User Experience, and Usability: Designing Pleasurable Experiences. HCI International 2017 Proceedings, Part II, Vancouver, BC, pp. 77–86 (2017)
10. Nielsen, J.: 10 usability heuristics for interface design (1995). www.nngroup.com/articles/ten-usability-heuristics
11. Bastien, C., Scapin, D.L.; Critères Ergonomiques pour l'Évaluation d'Interfaces Utilisateurs (version 2.1). INRIE. Relatório Técnico No. 156, Paris (1993)
12. Schneiderman, B., Plaisant, C.: Designing the User Interface. Addison Wesley, Boston (1986)
13. Apted, T., Collins, A., Kay, J.: Heuristics to support design of new software for interaction at tabletops. In: 2009 Computer-Human Interactions, Boston, MA (2009)
14. Inostroza, R., Rusu, C., Roncaglioso, R., Rusu, V.: Usability heuristics for touchscreen-based mobile devices. In: 9th International Conference on Information Technology. IEEE Computer Society, Temuco (2013)
15. Neto, E.V., Campos, F.F.C.: Evaluating the usability on multimodal interfaces: a case study on tablets applications. In: Design, User Experience, and Usability: Theories, Methods, and Tools for Designing the User Experience. DUXU, Part 1, Crete, Greece, pp. 484–495 (2014)
16. Agner, L., Necyk, B.J., Renzi, A.B.: Pervasive information architecture and media ecosystem: a Brazilian video on demand user experience. In: Design, User Experience, and Usability: Designing Pleasurable Experiences. HCI International 2017 Proceedings, Part III, Vancouver, BC, pp. 570–580 (2017)
17. Renzi, A.B.: UX heuristics for cross-channel interactive scenarios. In: Design, User Experience, and Usability: Theory, Methodology, and Management, Vancouver, BC, pp. 481–491 (2017)
18. Renzi, A.B., Almeida, O.: UX heuristic evaluation of cross-channel interaction systems: first experimentations. In: Interaction Latin America 2018, Rio de Janeiro, RJ (2018)
19. Nielsen, J.: How to Conduct a Heuristic Evaluation (1995). https://www.nngroup.com/articles/how-to-conduct-aheuristic-evaluation/
20. Renzi, A.B., Muniz, J.V.A., Fiúza, F.; In: Experiência do usuário pervasiva no planejamento de viagens: mapeando modelo mental e criando personas. In: SPGD 2018 Proceedings, Rio de Janeiro (2018)

Banking Cross-Channel System UX Evaluation

Olnei Fonseca de Almeida[1(✉)] and Adriano B. Renzi[2]

[1] Banco Original, São Paulo, Brazil
nei.fda@gmail.com
[2] UFF – Universidade Federal Fluminense, Niterói, Brazil
adrianorenzi@gmail.com

Abstract. As interaction evolves from the second wave of computing to the third wave of computing, where one user connects to many computers, the physical and digital world becomes more integrated and the user's experience journey encompass diverse apparatuses to build one whole story. The human-computer interaction turns to human-information interaction, systems become ecosystems, users become intermediaries, static becomes dynamic, dynamic becomes hybrid, horizontality overcomes verticality, product design gradually turns into experience design and experiences become cross-media. As the cross-channel interaction brings the necessity of new concepts and new ideas, a new set of 9 UX heuristics was proposed in 2016 [1] to map cross-channel experiences with new technology and interaction possibilities. This paper presents the evaluation of the system using cooperative evaluation based on the UX heuristics to discuss the system role as part of the UX journey and effectiveness of the UX heuristics.

Keywords: User experience · Cross-channel · Ecosystems · Human-computer interaction

1 Introduction

As interaction evolves from the second wave of computing to the third wave of computing, where one user connects to many computers, the physical and digital world becomes more integrated and the user's experience journey encompass diverse apparatuses to build one whole story. The human-computer interaction turns to human-information interaction, systems become ecosystems, users become intermediaries, static becomes dynamic, dynamic becomes hybrid, horizontality overcomes verticality, product design gradually turns into experience design and experiences become cross-media [2].

Cross-channel interactions regard the experience of multiple users moving through individual ecosystems of connected touchpoints to achieve a desired objective. Resmini and Lacerda [3] formally define a cross-channel ecosystem as "the ecosystem resulting from actor-driven choice, use, and coupling of touchpoints, either belonging to the same or to different systems, within the context of the strategic goals and desired future states actors intend to explicitly or implicitly achieve". John C. Thomas, PhD, from the IBM J.T. Watson Research Center sees the process of user experience as

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 559–570, 2020.
https://doi.org/10.1007/978-3-030-20040-4_50

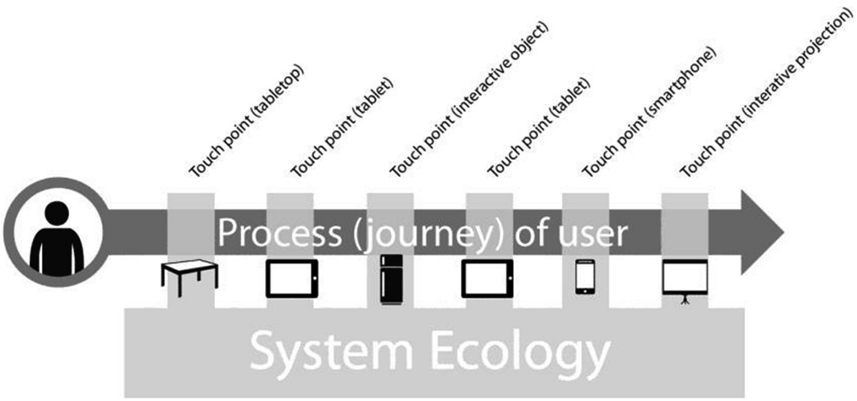


Fig. 1. Construction of an UX narrative through pervasive cross-channel sequence of interactions. Touchpoints with each apparatus can have diverse temporality and contexts.

narratives [1], and he structures it as several short stories with connections and touchpoints (Fig. 1).

Each touchpoint is connected to a system, part of a ecology of systems, integrated as one, to give users the sense of one whole integrated experience. Interactions with digital systems are not isolated interactions, as systems are connected through a systems' ecology [4], within physical ambiances, involving actors and other users, to create a pervasive cross-channel experience [5] with blended spaces [6]. Even when not properly planned to blend in, users tend to connect channels by themselves, using different systems on complementary actions and even going out of the systems, if necessary, to fulfill the gaps of interaction [7, 8]. As observed with people going through the 4 micro-moments to fulfill a trip [7], users utilize different devices and services to complete the UX journey and the context of use outside the apparatuses can be factors of influence to choices. The UX designer has to understand the whole journey in order to plan possible touchpoints and create a service or product experience integrated by different channels: a pervasive cross-channel experience.

New technology and interaction possibilities, take users to gradually change their mental models, their behavior and expectations towards new affordances and experiences. As the cross-channel interaction brings the necessity of new concepts and new ideas, a new set of 9 UX heuristics was proposed in 2016 [1] to map cross-channel experiences with new technology and interaction possibilities that are expected to surface and improve in the next 10 years [9]: touch, vocal, gestural, vision, though etc.:

1. Place-making – refers to the self localization of users in the system and within the experience journey. Visual interaction, hierarchical layout and structure, as well as physical environment should facilitate the user's understanding of where he is. Since users most likely will use different devices and the ambience itself1 to fulfill objectives, it is important that physical environments also be part of the user experience strategies to create an integrated journey;

2. Consistency – the system has to present visual, typographic, information, actions and interaction consistency. If a user utilizes different devices to execute partial actions of the whole experience, each touchpoint access has to present the same rules and responses to actions, independent of the artifact that is being used;
3. Resilience – flexibility of the interaction flow and touch points in order to adequate to different users, different journey strategies and different contexts of use. The interactive ambiance and the system structure should be prepared to search, interaction and journey diverse strategies by different users, sometimes with distinct roles in the same journey;
4. Reduction – even if the back-end of the system is complex in its structure, the options and the contents have to be presented to the users in objective way and with simple usage, providing reduced interactive actions and minimum cognitive workload in their journey. From the users' point of view, the path of possible actions has to be obvious to their necessities;
5. Correlation – the system has to go beyond the semantic correlation to help users find information and content naturally. The principle expands to a correlation of data between distinct points of interaction and apparatuses, as well as correlation of actions from different users within the same experience journey;
6. Equivalency to cultural conventions – it is important to understand users' references regarding technology, processes, functionality comprehension and interactions, to use as a base in the development of a new system. To create structures and interactions that users are not familiar with can surface doubts and misunderstanding about the system;
7. Visual intuitive content – users must recognize functionalities, hierarchy, pathways and information with minimal memory load, by making objects, actions and options easy to recognize and understand;
8. Natural, intuitive and direct interactions – any touchpoint of interaction with the system should be as intuitive as possible, by direct gestural manipulation or objective simple vocal commands;
9. Contextual ergonomics – physical environments, contexts of use within the journey of experience and human physical limitations should be considered while projecting touch points of interaction with the system.

First experimentations with these heuristics were conducted through heuristic evaluation of the Airbnb ecosystem, using 3 digital artifacts (desktop computer, table and smartphone) in a cross-channel pre-set path [8]. The collected results clarified important questions regarding the evaluation process and the use of the heuristics in a cross-channel context. This paper presents sequential test with the UX heuristics, but using Think-aloud Protocol with final users [10, 11].

2 Cross-Channel Digital Banking

The Original Bank was founded in 2015, in a growing wave of pervasive interactions, and is recognized as an innovative entrepreneurship for being the first Brazilian bank institution to provide the possibility of opening a checking account through online channels and have a virtual assistant through its social networks.

It is a Brazilian 100% digital bank – considered in this research as banks that do not have physical branches and all offered transactions to clients are through digital artifacts. Its ecosystem encompass different channels of interaction with clients: the Original smartphone app, the official website, the internet banking platform and many conversational interfaces to fill in any possible gap, such as Messenger, WhatsApp and URA (audible response unit), through which users can execute transactions, money transfers, credit card tracking, activate the security system, check their account balance and read information about procedures in the bank.

Since the primal connection with users occurs exclusively through digital channels, the company continually invests in UX teams to make the whole experience journey smooth and easy to users. Constant usability and performance tests are run in the system and service offers evolve according to technology possibilities and new channels that may surface to common use.

The system is constantly evaluated with the help of users and mostly based on usability heuristics focused on specific apparatuses. But in order to properly evaluate the ecology system, it is important to have a cross-channel approach.

Using the 9 UX heuristics, a first experimentation was conducted on the Airbnb ecosystem [8] using the Heuristic Evaluation method, for comparison with isolated one-artifact usability evaluations. Collected results from the first experiment presented interesting outcomes and possibilities in systems development and evaluation. Although the UX heuristics are the foundation to evaluate the system in this research, the method is shifted from Heuristic Evaluation to Cooperative Evaluation with final users. As exposed in previous research [8], the UX Heuristic Evaluation can take a long time to fully execute, depending on how many channels and digital artifacts are involved in the UX journey. The Airbnb system evaluation had to go through smartphone, desktop/laptop, tablet and back to smartphone to fully test the system. The experimental execution of the evaluation also required UX experts very familiar and experienced with Heuristic Evaluation, which showed to be a recruiting challenge.

Since the Original Bank has a high security system with no open access to new users or outside experts. Any evaluation has to be conducted with internal experts and this could be problem to use the UX Heuristic Evaluation, on aspects of experts' availability and time availability.

As new approach to further experiment the heuristics and fully evaluate the ecosystem, this research proposes the use of Think-aloud Protocol as the interactive testing basis and all collected data to be categorized and discussed on the 9 UX heuristics.

3 Think-Aloud Protocol

Think-aloud Protocol was applied to better understand users' interactions with the ecosystem. According to Villanueva [10], the method consists of a researcher observing users doing specific tasks within a controlled environment. The users' actions and thoughts are to be described verbally aloud by themselves on real time. The researcher records the users' actions by written notifications, video shooting or voice recorder.

Filming and voice recording have the advantage of capturing the exact steps and descriptions of users, while written notifications depends on the researcher experience with observing reactions and quickness in writing down relevant actions of the experiment [11]. However, using writing notes have the advantage of creating an informal observing environment, resulting in more flowing sessions with the participant. While the choice of filming captures every movement for future analysis, but could be considered intimidating. For this research the chosen direction was to use both voice recording and written notifications in order to keep the users in an informal environment and record all their spoken actions.

When noticing some reluctance from the users in verbalizing actions and thoughts during the Think-aloud Protocol, questions related to the users' actions can placed to keep the flow of verbalization of their thoughts, based on Xiao's [12] experiments: the researcher used general questions as users advanced on the proposed tasks for the session: "what do you think about the tool?", "did you understand the steps?", "any doubts regarding the tour?", "do you have suggestions?". According to Xiao, the insert of questions during the session helped to identify problems from the perspective of users of varying degrees of skill with the virtual system, as well as get constructive ideas for improving the tool.

Villanueva [10] indicates the use of Think-aloud Protocol in 4 steps

1. Organize a small number of users (around 4)
2. The researcher meets each user separately
3. The researcher provides a prototype for users to execute a list of tasks
4. Users verbalize thoughts and actions while executing tasks
5. Researcher makes notes on improvements to be considered for the system when analyzing results.

The Think-aloud Protocol was conducted with the help of 4 users, between the ages of 25 and 35 years old, familiar with technology. Half of the invitee are recent Original Bank clients and half are not, in order to minimize the familiarity with the system as an influence factor. The method follows a pre-set cross-channel path to fulfill specific tasks in order to simulate a real context, based on interviews and task observation with the help of 3 users, with their actions written and audio recorded. The context of use was presented to all participants and all doubts clarified before starting each session.

In order to evaluate the ecosystem and its information architecture pervasively, the proposed task obliged the users to go through 2 different channels and 4 different displays, based on most frequent actions by real users: cash deposit. The task goes through 4 sequential steps (Fig. 2):

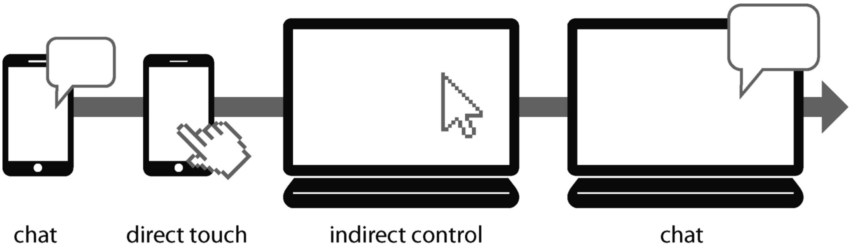


Fig. 2. Sequence of touchpoints with the ecosystem through each apparatus with diverse types of interaction: textual chat, direct touch, indirect control and textual chat.

1. Add the number 11 4004 0800 in the smartphone list of contacts and name it Virtual Assistant Original Bank. Initiate a conversation with the Virtual Assistant using WhatsApp to understand how to deposit in cash.
2. Access the Original Bank App through smartphone to make the proposed deposit on a pre-set checking account.
3. Access the website (www.original.com.br) and the Original Internet Banking using a desktop or laptop. Search for the account statement to identify the deposit from step 2.
4. Back to the Original Internet Banking, contact the Virtual Assistant to check information regarding the deposit status.

The collected results from all 4 users were further analyzed using the 9 UX heuristics. Each reported problem and its referenced discussion was related to one or more correspondent heuristic for further discussion and analysis.

4 Results

The first phase of tasks had users insert the bank official contact number in their contact agenda and star a conversation through WhatsApp and obtain information on how to do a deposit in cash. All participating users showed easiness in accomplishing this first step and interacted naturally with the Virtual Assistant and none have flagged any difficulties: “I liked that the important words are in bold and help for a faster reading”.

However, a discomfort was evident with the quantity of information given from the start of the conversation. The problem reported is related to *heuristic 4 (reduction)* as an unnecessary amount of information is exposed to users – even if the back-end of the system is complex in its structure, the options and the contents have to be presented to the users in objective way and with simple usage. Two participants noted the lack of objectivity from the Virtual Assistant answers, as one user expressed during the Cooperative Evaluation: “It doesn’t have to be so extensive. I wish the assistant answered direct to the point. This huge text is unnecessary”.

After receiving the instructions from the Virtual Assistant, users should access the Bank App to make the deposit. As explained by the chat bot, digital banks use a bank invoice/slip (printed by the client) in order to make the payment and the money is

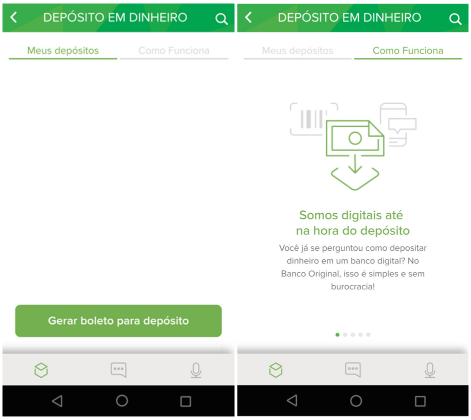


Fig. 3. Deposit in cash section – accessing the How it works page.

credited to the client's checking account. When accessing the bank app, all participating users had difficulties in finding where to reach Deposit in the app's menu. At this point, all users minimized the app and went back to the Virtual Assistant to check again the instructions. It was confirmed that one step was missing in the instructions. After a few minutes searching through the app's starting page to find the Deposit link inside the menu. As users interacted with the Virtual Assistant to get answers on steps to fulfill the task, the missing information was an impact factor to mislead users and keep them from understanding what would be the next natural step. These reported problems are related *to heuristics 1 (place-making) and 5 (correlation)*, as users get confused on where they are in the experience, the lack of consistency in the instructions with actions – correlation of data between distinct points of interaction and apparatuses, as well as correlation of actions.

Two users expressed their frustration: "Oh! But this was not mentioned in the instructions", "I found it because I have experience with other bank apps, but my father would be looking around looking for the deposit icon" and "The answer (Virtual Assistant) has to specify the need to enter the menu to reach deposit. One step was missing" (Fig. 3).

Users follow the sequence of actions after Deposits is found and none showed difficulties in identifying the option deposits in cash. But when entering the deposit area, two users stopped the task as presenting difficulties to understand the relation between deposits in cash the bank invoice generator. When instigated by the evaluator to search for instructions, these users reached the "how it works" button to obtain informative instructions and returned to fulfill the task.

The problems registered for this phase of the experience journey are related to *heuristics 2 (consistency), 5 (correlation) 6 (equivalency to cultural conventions) and 7 (Intuitive visual content).* Comparing the app experience with known daily concepts, users found it difficult to finish the task. The instruction to generate a bank invoice with no detailed explanation and no relation to physical cash deposit convention resulted in doubts, misunderstanding of the correct and insecurity by users.

After the bank invoice session is entered, users have to fulfill the slip information using the app digital form. One user showed problems to select the payment date for the same day, as the test was executed on Saturday and no banks are open for deposit. Nevertheless, the problem was related to heuristics *6 (equivalency to cultural conventions) and 7 (intuitive visual content)* as the app's calendar does not the rules clear enough, making the user confused: "I can't choose today, only after tomorrow. How can I make the deposit?"

The exclusion of weekends for deposit was based on physical bank availability, however, for physical banks it is possible to deposit using the bank's ATM machine (a cultural convention), making it natural to users to be surprised by this unavailability. In the sequence, three users chose immediately the copy the barcode as sharing option, while the last user tried to manually copy the barcode numbers before realizing the copy function.

When accessing the website (Fig. 4), the visual design was celebrated, but one user had problems entering his account through the Internet Banking, as the link did not respond accordingly and the access menu did not appear. The user had to reopen the browser and access the Internet Banking once again, related *to heuristic 1 (place-making)*. The problem could be from the browser, but the user blamed the bank system: "It's not working. It doesn't click. That's bad!".

Fig. 4. Original bank homepage – access your account signalized in red on upper right area.

When entering the Internet Banking to check the account statement, user spent a few seconds to familiarize with the design and structure of the page. All passed the mouse over the shortcut to account statement, but only two identified it as a shortcut and clicked on it for a quick access. The other two users went through the option account, on the lateral menu. But the system did not have the recent deposit statement for them. For one user, the lack of statement caused confusion for a while, but relieved when finding the information in future transactions. To reach future transactions, it is necessary to click the futures transactions on the lateral menu. The problem is related to heuristics *1 (pace-making), 4 (reduction) and 7 (intuitive visual content),* as users got confused when not finding the statement and had to do an extra action to reach the information. The action was not obvious.

Although all users were able to fulfill the task, they expressed confusion and questioning based on the lack of information about their recent deposit. The user who used the future transaction option showed even more confusion: "I don't get it. I am at the account statement and there is no info on the money I've just deposited. Was it accepted? I'm not sure where to check if I did it right or not" (Fig. 5).

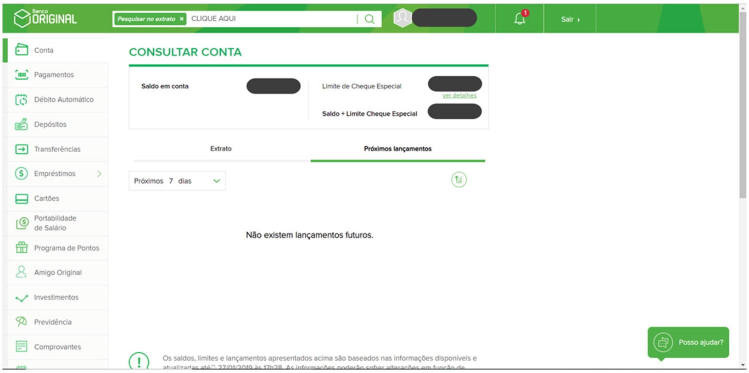


Fig. 5. Future transactions page – message: there is no planned future transactions.

The last part of the task would be to go back to the Virtual Assistant using the Internet Banking and check information regarding the deposit status. For 2 users, going back to the assistant using the channel showed no problems, but to one user it was not obvious to find the option in the Internet Banking system. He did not understand how to proceed and use the interaction with the assistant: "Can I use WhatsApp? I am more familiar with it". Another user ignored the proposed task and went back to WhatsApp to fulfill the task. The problem is related to *heuristics 7 (intuitive visual content) and 8 (natural, intuitive and direct interactions),* as the action was not considered obvious and the user could not understand the procedure – any touchpoint of interaction with the system should be as intuitive as possible.

When using again the Virtual Assistant to get information about the deposit statement, users did not find the answers conclusive and would contact an human attendant of the bank to ask about it (Fig. 6). All users were confused on how to proceed and how to properly interact with the chat bot, causing an extra effort and frustration: "I give up, I 'm calling the bank and ask a human", "I feel very ignorant when talking with the assistant (chat bot)". The interaction problem with the bot is related to *heuristics 4 (reduction), 5 (correlation) and 8 (natural, intuitive and direct interactions),* as the responses are redundant, don't relate to the other channels and the interaction is not natural, nor direct.

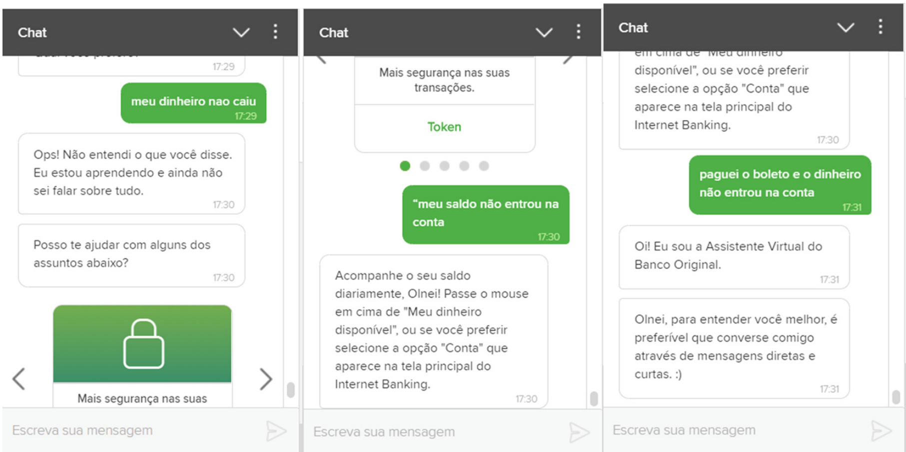


Fig. 6. Virtual Assistant chat bot conversation – user: my money has not got in | bot: ops! I didn't understand what you said. I am learning and do not know about this. Can I help you with one of the subjects bellow? | user: my statement account didn't get in | bot: Check out our account statement daily. Mouse over "my available money", or if you prefer, use the option "account" that shows at the main screen of Internet Banking | user: I paid a bank invoice and the money did not get in | bot: Hi! I am the Virtual Assistant. To better understand you, it is preferable to talk with me using short direct messages.

5 Conclusion

The use of Think-aloud Protocol has shown to be more a agile method to help analyze the system, if compared to the previous experience with the cross-channel UX evaluation [8]. Both methods have to go through all channels related to the whole UX journey to fully evaluate the ecosystem. But using Think-aloud could be faster, as the process includes users and all problems are real time verbalized, instead of inviting UX experts and having formal notifications.

The previous experimentation with UX heuristic evaluation presented difficulties in reaching experts with proper experience to participate, as it happens with all heuristic evaluations. The Think-aloud protocol, or any method that works directly with users, can be faster as there is more availability of participants and no particular expertise is necessary. Since the ecosystem would be hard to connect by outside experts, the Think-aloud Protocol was an assertive choice to have the system evaluated in less time.

However, the use of the UX Heuristic Evaluation could surface more problems and more depth of discussion, as the system would be evaluated with experienced specialists, as stated by 83 problems in its first experiment. The use of Think-aloud Protocol needs a lower budget and faster accepted by stakeholders, and due to its results it could be a first step to further investigation with a higher budget.

The collected results from the Think-aloud Protocol could be easily translated to one or more UX heuristics. This experiment in particular was interesting because provided the possibility of having conversation interaction part of the evaluation and properly translated to the heuristics, as it was primarily proposed to encompass visual, touch, voice, gestural and though actions.

The insertion of chat bot interaction as part of the cross-channel journey in a planned pervasive interaction is important to keep the flow [13, 14] and vividness [15], but if inserted isolated and with bad correlation to the journey, could bring bigger frustration than if it didn't exist in the system, as presented by users facing problems with the Virtual Assistant. It is perceived that many pointed problems by users could be prevented if information and documentation were offered complete and in context of use. Many situations showed that comprehension of functionalities and help were crucial to forward users actions, but were absent or creating misunderstanding about the system.

Users showed hesitation when interacting with conversational interfaces, difficulties in communicating their doubts in condensed phrases for better recognition by the Virtual Assistant. As a consequence, the whole journey was affected by conversation interaction problems and users' unfamiliarity with this type of interaction. The system has to adapt to users; expectations and mental model, therefore, be more simple to use and have a broader vocabulary reach to understand poorly constructed phrases and contexts of use – heuristic 4 (reduction), heuristic 5 (correlation) and heuristic 8 (natural, intuitive and direct interactions).

Interactions and information architecture have to be planned more pervasively in order to accommodate close future technology advancements and users' expectations, as everything will be more integrated and connected [9]. The near future brings our five-senses together in one flow of interaction to build experience journeys. Entrepreneurs, designers and developers need to strategically plan possibilities of interaction to accommodate the flexible strategies of users and minimize glitches in the UX process.

References

1. Renzi, A.B.: Experiência do ususário: a jornada de Designers nos processos de gestão de suas empresas de pequeno porte utilizando sistema fantasiado em ecossistema de interação cross-channel. Doctorate thesis. 239p. Escola Superior de Desenho Industrial. Rio de Janeiro, Brazil (2016)
2. Resmini, A., Rosatti, L.: Pervasive Information Architecture – Designing Cross-Channel User Experiences. Morgan Kaufmann, Burlington (2011)
3. Resmini, A., Lacerda, F.: The architecture of cross-channel ecosystems – from convergence to experience. In: Proceedings of the 8th International ACM Conference on the Management of Digital Ecosystems (MEDES 2016), Hendayne (2016)
4. Unger, R., Chandler, C.: A Project Guide to UX Design – For User Experience Designers in the Field or in the Making. Peachpit Press, Berkeley (2009). 288f
5. Renzi, A.B.: Experiência do usuário: construção da jornada pervasiva em um ecossistema. In: SPGD2017 Proceedings, Rio de Janeiro, Brazil, 2017
6. Benyon, D., Resmini, A.: User experience in cross-channel ecosystems. In: 31st British Human Computer Interaction Conference, Sunderland, UK (2017)

7. Renzi, A.B., Muniz, J.V.A., Fiúza, F., Experiência do usuário pervasiva no planejamento de viagens: mapeando modelo mental e criando personas. In: SPGD 2018 Proceedings, Rio de Janeiro, RJ (2018)
8. Renzi, A.B., Almeida, O.: UX heuristic evaluation of cross-channel interaction systems: first experimentations. In: Interaction Latin America 2018, Rio de Janeiro, RJ (2018)
9. Renzi, A.B., Freitas, S.: Delphi method to explore future scenario possibilities on technology and HCI. In: Human-Computer Interactions International. DUXU, pp. 644–653. Springer, Los Angeles (2015)
10. Villanueva, R.D.A.: Think-aloud protocol aril heuristic evaluation of non-immersive, desktop photo-realistic virtual environments. Dissertation (Master of Science) – University of Otago, Dunedin – New Zealand (2004)
11. Renzi, A.B. Freitas, S.: Aplicação de think-aloud protocol em teste de usabilidade na procura de livros em livrarias online: recomendações de leitores. In: 10º USIHC – Congresso Internacional de ergonomia e usabilidade. Conference Proceedings, Rio de Janeiro (2010)
12. Xiao, D.Y.: Experiencing the library in a panorama virtual reality environment. Library Hi Tech **18**(2), 177–184 (2000). National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov
13. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper and Row, New York (1990). 303 p.
14. Agarwal, R., Karahanna, E.: Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Q. **24**(4), 665–694 (2000)
15. Coyle, J.R., Thorson, E.: The effects of progressive levels of interactivity and vividness in web marketing sites. J. Advert. **30**(3), 65–77 (2001)

Desate Nó: Management System App for Buildings

Evelin Daiane Barreto dos Santos[1](✉) and Adriano Bernardo Renzi[2]

[1] Serviço Nacional de Aprendizagem Comercial, Rio de Janeiro, Brazil
evelindbsantos@gmail.com
[2] UFF – Universidade Federal Fluminense, Niterói, Brazil
adrianorenzi@gmail.com

Abstract. Apartment complexes in Brazil are usually managed by its residents, as each one represents an equal share in administrative decisions regarding the building they live. According to the housing syndicate survey, applied to apartment complexes administrations, the most common pointed problems are related to communication, behavior, security and payment debts. This paper is part of an extensive research, part of the Interaction Design program project, with building managers and residents, involving observation, interviews, storytelling, cooperative evaluation and think-aloud protocol to help understand and map users' mental model, their expectations and affordances to integrate the use of digital devices into building managing processes and improve the user experience.

Keywords: User experience · Human factors · Human-computer interaction · Buildings management

1 Managing Apartment Complexes

In Brazil, apartment complexes are usually managed by its residents and each one represents an equal share in administrative decisions regarding the building they live in. According to Brazilian legislation, the residents, part of the same apartment complex, must elect one central administrator to manage the building. The elected manager can be either one of the residents or an outside professional administrator, who becomes the spokesperson for all administrative matters, who deals with daily dilemmas of residents and ensure that the building rules are applied. The manager also takes care of the financial management, manages employees and suppliers, as well as, has other responsibilities.

According to the housing syndicate survey (Rio Grande do Sul), applied to apartment complexes administrations, the most common pointed problems are related to communication, behavior, security and payment debts. The survey shows that 91% prefers to contract administrative companies to act in partnership with the elected central administrator (either a resident or an outside professional administrator). Independently of the source of administration, 73% of the interviewed wish to have more agility in issues resolve.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 571–581, 2020.
https://doi.org/10.1007/978-3-030-20040-4_51

Using digital apparatuses to help accomplish tasks and do daily chores have been gradually integrated to people's everyday lives. Future interaction scenario prediction surfaces expectations of an even greater integration of digital and physical worlds, with wearable technology more accessible and interactive surfaces everywhere [1]. But even nowadays, the usage of smartphones, smart watches, tablets etc. are already part of cross-channel experiences to accomplish tasks and build experience journeys [2, 3].

This paper is part of an extensive research, part of the Interaction Design program project at Senac, with building managers and residents, involving observation, interviews, storytelling, cooperative evaluation and think-aloud protocol to understand and map users' mental model, their expectations and affordances to integrate the use of digital devices into building managing processes and improve the user experience.

Based on users' mental model, the digital platform Desate Nó was proposed to easier the issue reports process and connect residents directly to the responsible administrator, minimizing behavior problems, bringing transparency of the administrative processes and making the communication better.

Although a research with many phases, this paper focus mostly on the high fidelity prototyping and testing phase, using Think-aloud Protocol [4] and Cooperative Evaluation [5]. From precedent research on users' mental model and context of use, a digital prototype was built to test the system usability, as well as users' expectations, to better simulate the connection of users with the system through smartphones. The results surfaced important adjustments to the app and better understanding of users' needs.

Attempting to map residents' most urgent necessities and their mental model, a qualitative semi-structured interview was conducted with inhabitants of a pre-selected apartment complex in Rio de Janeiro. The interview structure followed concepts of Muchielli's [6] psychosocial surveys. All interviewed participants are members of the same WhatsApp apartment complex group, as the use of WhatsApp is common practice in Brazil and considered the easiest channel to communicate and exchange information within a preset group of people. Observing the group using Netnographic research method [7] has also brought important information regarding the major discussed issues and residents' social relation. Although an informal channel, the WhatsApp group brings people together to discuss common problems, exchange experiences and share suggestions for improvement. However, it can become a nightmare to the building manager, as it doesn't control of people behavior nor prevents out of topic subjects, which frequently could vary from a burnout bulb to the latest Internet meme to neighbors arguing about personal issues.

Important information was collected during the users' mental model mapping phase that helped establish the basic structure for a high fidelity prototype development. Among these collected results, we highlight a few:

- Some of the residents acknowledge that probably their messages reporting a problem can sometimes be missed by the manager, as the amount of messages is huge and it is hard to reach older posts.
- The physical book of occurrences, located near the building entrance, is the official channel of communication, but residents prefer to use the smartphone as is more convenient and gives the notion of more agility.

- All interviewed residents would like to receive an immediate reply from their message and have the possibility of following the status of proceedings regarding their reported problem.
- Residents wish for more agile procedures to solve reported problems.
- Some of the interviewed residents show dissatisfaction for the lack of reply and lack of transparency from the building manager. For these reasons, they stopped reporting issues.
- Part of the interviewed users do not trust the management.
- Residents wish to be heard by the building manager, to have a closer and transparent administration, and to live in a building with cleanness and order.

In order to better understand the manager's point of view, the main researcher played the role of manager-for-a-day and became responsible for following the WhatsApp group, checking the book of occurrences, solving reported issues (with the help of the official manager) and feeding back the status of activities. The experiment helped confirm a few noted data from the interviews and a few factors were acknowledged to better understand the manager's mental model and develop the structure of the management system:

- The manager can get confused with so many messages from the WhatsApp group. It is hard to read all messages through the fast passed daily routine with many tasks to accomplish along the day.
- Too many demands to be concluded, as it is hard to keep up with employees, tasks and messages to respond.
- It is constantly asserted for transparency in his administration.
- Doesn't have the global understanding of correlated reported issues and preventive maintenance, as the incoming information are decentralized and not organized.
- Part of the manager's day is consumed by phone calls, many invoices requests and follow licenses expiration dates and needs.
- The task and employee management is sometimes ineffective, due to the lack of accomplished tasks and to do list overview. The difficulty leads the manager to personally go many different places within the building just to have an overview of the general "on-goings".

Based on collected information from interviews, the netnographic observation [7] and the experience of being the manager for one day, a structure of information (Fig. 1) was established to develop a high fidelity prototype for further testing. The structure and system was planned for smartphone use, as it became clear this type of device was the easiest channel for users to access and communicate.

2 Usability Testing: Think-Aloud Protocol

The collected data and noted factors of influence helped configure the basics of information architecture structure for a refinement based on Resmini and Rosatti [8] and Agner [9] concepts of dynamic architecture. From insights concerning smartphone's heuristics [10] a high fidelity prototype was developed for testing with users.

Fig. 1. Basic structure of Desate Nó app and flow of actions

Before developing the prototype, an additional card sorting test [11, 12] was applied to 4 users who lived at the building.

The prototype was built with the help of InVision. The system enables an easier online access by users through their own smartphones and presents building features to better simulate the final product. Each user accessed the prototype individually, using his/her own smartphone, accompanied by the researcher. In order to evaluate the prototype two methods were used simultaneously for comparison: Think-aloud Protocol and Cooperative evaluation.

The main objective of using Think-aloud Protocol and Cooperative evaluation for the prototype phase is to check if the system can be easily understood by users, the steps to fulfill tasks are obvious, the interaction flow is smooth and if there are any usability problems along the experience. Both methods are very similar in their proceeding, but with a few specific differences and results. It was expected that results from each method could surface similar as well as complimentary problems.

For each method, 4 users between 25 and 75 years old, representatives of male and female genders and diverse education formation and professions were selected to participate. All invitee had to live in the apartment complex previously part of the research and use often the smartphone on daily basis.

According to Villanueva [4], the method consists of a researcher observing users doing specific tasks within a controlled environment. The users' actions and thoughts are to be described verbally aloud on real time. The researcher records the users' actions by written notifications, video or voice recording.

Filming and voice recording have the advantage of capturing the exact steps and descriptions of users, while written notifications depends on the researcher experience with observing reactions and quickness in writing down relevant actions of the experiment [13].

However, the use of writing notes has the advantage of creating an informal observing environment, resulting in more flowing sessions with the participant, while the choice of could be considered intimidating. For this research the chosen direction was to use both voice and screen touch recording as well as written notifications in order to keep the users in an informal environment and record all their spoken actions. The App DU recorder was used for this phase, where the participants could be recorded while doing the required tasks at the same time.

When noticing some reluctance from the users in verbalizing actions and thoughts during the Think-aloud Protocol, questions related to the users' actions should be placed to keep the flow of verbalization of their thoughts [11]: "what do you think about the tool?", "did you understand the steps?", "any doubts regarding the tour?", "do you have suggestions?".

Villanueva [4] indicates the use of Think-aloud Protocol in 4 steps

1. Organize a small number of users (around 4)
2. The researcher meets each user separately
3. The researcher provides a prototype for users to execute a list of tasks
4. Users verbalize thoughts and actions while executing tasks
5. Researcher makes notes on improvements to be considered for the system when analyzing results.

In Cooperative evaluation, the researcher also selects users to analyze a system. Each user, separately, fulfills pre-defined tasks by the researcher and verbalizes issues and doubts during the process [13]. The surfaced issues are discussed with the researcher, who usually sums questions with the purpose of deepening further discussed subjects and gather data for the analysis. It is important to be full aware of users' mistaken actions and misunderstanding perception in order to pinpoint problems and elucidate solutions from users' perspective. User and researcher evaluate the system together and the participants are encouraged to ask questions about the purpose and interaction of the product.

Miranda and Moraes [14] recommend the use of Cooperative evaluation in products that need improvement and in prototypes while in intermediary and final phases. According to the authors, the evaluation follows 4 basic steps:

1. Recruit participants who are representative of final users
2. Plan and prepare tasks for the participants to execute in order to let them explore the whole system or at least the parts considered most relevant
3. Record and make notes of all actions, questions and discussions during the execution of tasks by users
4. Analyze the collected results.

The selected 4 tasks, to be executed by users, simulated the whole user experience based on results from previous collected data (interviews, netnography, manager-for-a-day) and intended to reach all major pages of interaction:

– Create a new report of a problem in the building
– Search and follow the problem report (previously created) status
– Verify if the report was sent and checked by the management
– Access non read notices from other users and open at least one to read.

Although both of the chosen methods are considered as qualitative research and do not require too many participants to help test the system, the research schedule the sessions according to users' availability to avoid any discomfort or rush from participants while executing tasks. Each session lasted for 10–30 min and some selected locations for the tests were difficult for audio recording, due to playground noise or just people walking/chatting nearby.

3 Results

Before the high fidelity prototype development, a prior design was established for testing and discussion with the design team and interaction design mentors from national and international companies, as part of the educational program. The design team and mentoring evaluated the first proposal graphic language and pointed a few changes in the project. The graphic layout used bright warm colors and elements related to infancy, which could bring fun and childish characteristics that could distort the seriousness impact of the app. After a few sessions to discuss the direction of the project, a new proposal was presented with a more seriousness graphic impact. The prototype for usability testing was developed following the new graphic proposal.

Users easily accomplished the first task (to create a new report of a problem in the building). The icons were considered obvious as well as the order of actions to complete it (Fig. 2): open app > click on report > add an issue > write specifics > send.

Fig. 2. Sequence of actions to report a problem using the Desate Nó app.

While fulfilling the second task (search and follow the problem report status), a few usability problems were highlighted by the participating users. As they tried to find and select the proper reported issue, three users did not recognize the significance of the proposed icons to filter reports nor their relation to filtering feature (Fig. 3). The icons were related to reported issues. The first one (clock icon) refers to reported problems in progress, which helped filter the issues in read and non-read messages. The checked icon referred to solved problems, the flag icon referred to saved problems for a closer monitoring and the heart icon referred to problems supported by other residents, with further comments and different points of view.

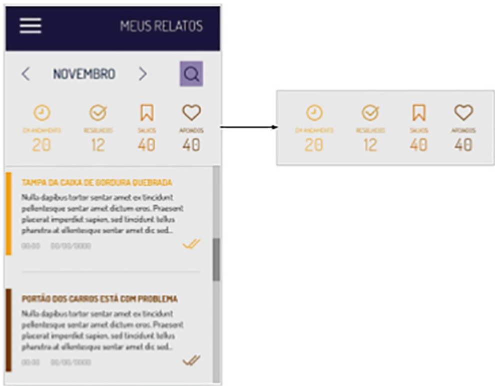


Fig. 3. Icons are hard to comprehend due to lack of contrast, legibility and choice of iconographic representation.

Still related to the perception of the icons, 3 of 4 users reported lack of contrast, making it difficult to notice and “read” them. Although the colors were considered matching choices, the gradient visual brings up a non-clickable feeling: “The icons are too subtle and the colors don’t have enough contrast”, “These do not look like buttons. They look like indicators of quantity or something like that”.

The third proposed task (verify if the report was sent and checked by the management) showed many issues, as only 2 users were able to recognize (Fig. 4) the similarity of checked items with Whatsapp checked messages. While browsing the list of problem reports, users could not understand which messages were already read by the management. The icons and difference of colors used for this feature were not considered obvious.

Even the 2 users who recognized the icons, took considerable time to understand the significances and what to do. For an older user, not very experienced with messaging apps, to conclude this task was impossible and reported to not understand the differences of messages nor what to do.

Although the forth task (access non read notices from other users and open at least one to read) was accomplished by all users, the action was considered completely compromised by its relation to task 2. As visual elements and position of icons in these two tasks have many similarities, the pointed problems in task 2 were perpetrated on task 4.

Fig. 4. The difference of sent messages and messages that are already read was not obvious, as it didn't mimic WhatsApp cultural conventions.

The same way reported on task 2, users mention similar problems during the execution of the 4th task. Part of participants could accomplish the task from the experience executing task 2, but not all. Users took a long time to associate icons with same functionality from the previous phase because were presented with different designs. The colors are noticed as a problem once more: "These colors look like a gradient, but too subtle. It confuses too much".

The Cooperative Evaluation, applied after the Think-loud Protocol, validated all problems noted intuitively by users fulfilling the tasks. Although the app was considered with clean and enjoyable design and with easy first steps, when going through more complex actions participants concluded that as progressing through the tasks, the system turned out to be hard to understand and accomplish.

Users wished to have more explanatory documentation with easy access and more obvious graphical elements to help the understanding of features and actions to take. It was unanimous the feedback on how the chosen colors palette affected negatively the fulfilling of tasks, as the used colors showed lack of variety, significance and contrast.

During the Cooperative Evaluation, discussions have enabled a deeper understanding of pointed problems and many suggestions for improvements were given. Some of the most relevant suggestion from users:

- Some participants suggested that all filters of informational content should have a direct brief written explanation above or bellow of the icons in sections "my story" and "comments". All participants agreed that this could help the understanding and use of the functionalities.
- Users propose the removal of quantity indicators, below the filtering icons in "my story" and "comments". They were considered useless information and cause distraction from important things.
- All emphasize the need to change completely the color palette. They are too lightened and need to have more contrast and legibility. The colors are too subtle.
- A check $\surd$ icon shows when messages are already read. But users suggested the check marks should mimic the ones used on WhatsApp for sent message ($\surd$), for message received by manager ($\surd\surd$) and message read ($\surd\surd$). This way, users would automatically relate to symbols well know for them.

4 Conclusion

The Think-aloud Protocol helped highlight problems through users natural interactions with the app, grasping glitches and usability problems from a first time usage and many outside references to influence perception and actions. The Cooperative Evaluation enabled users to be part of the process and discuss problems further and suggest solutions for specific parts, as well as, the whole experience. The merge of both methods helped compare the natural bumps of the system with descriptive and discussed issues with users. If only Think-aloud Protocol was used, the collected data would not have the deepness needed to fully understand the needs of residents.

It is important to notice that usually researchers use one of these two methods, integrated to a usability test with experts, such as Heuristic Evaluation. Although Cooperative Evaluation and Think-aloud Protocol have very similar procedures and the participation of users is the main detail to tell them apart, the complementary results showed an interesting facet to be further tested in the future.

The whole previous integrated process with interviews, Card Sorting and manager for one day were crucial to properly direct the prototype development and create a trusting relation with users (residents of the apartment complex) and a closer partnership between users and researcher. The division of the project in many phases having the direct participation of users was very relevant to users develop a higher perception of factors involved, a better sense of critique and more detailed contributions.

After collecting the data from the prototype evaluations, it was possible to set a strategy to fix problems and enhance important features to the user experience. The adjustments were organized in two phases: (1) revise all visual language of the app, setting visual standards for easy comprehension and legibility; (2) re-think the visual design and interaction for the sections “my story” and “comments”. Throughout the 2 phases of adjustments it is important to make visual elements more cognitively obvious, include explanatory labels, exclude unnecessary information, make sequence of actions more optimized and flexible to different users and make interaction closer to residents’ cultural conventions (messaging, WhatsApp, problems process rituals, users with different roles within the same UX journey).

The cultural conventions that should be considered are related to messaging rituals inherited from Whatsapp use. The recognition of icons and functionalities should also consider graphical elements from this system, to make cognitive comprehension more obvious.

The cultural conventions are built outside digital apparatuses as well and residents are already familiarized with the apartment complex procedures involved in the problems solving processes. From discovering a problem and reporting it to checking results, the sequential phases should be the main line of the UX experience journey [2, 3] to understand the possible touchpoints, the sense of things [2] and contexts of use for the app be part of it. The journey (Fig. 5) has many different actors (users) involved, each with his/her own objective, contributing to a better coexistence and usufruct of the building area, including their own short stories to the whole and having complementary roles at some points.

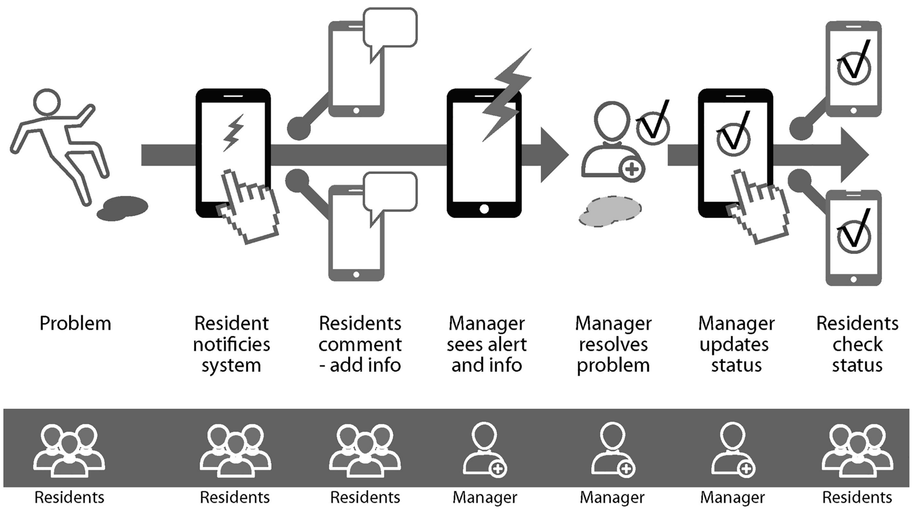


Fig. 5. UX experience journey through a pervasive sequence of interactions, integrating physical areas, contexts of use and digital artifacts (smartphones). The system has many actors through one channel.

The UX journey occurs pervasively, as it starts with a resident noticing (outside the app) a problem in the building area, using the app (touch interaction + text) to notify the problem and interact with other residents (text) – many actors. The manager (different actor) gets alert (touch interaction) and solves problem on the marked area. Manager updates information (touch interaction) and residents follow the update (touch interaction).

Although the UX journey takes place on both physical areas of the building and smartphones, the Desate Nó project would concentrate the whole communication on just one channel, minimizing problems of communication, lack of focus from users and loss of information related to problems. It would optimize the process through organized data, bring transparency to management and make easier to check the status of problems.

References

1. Renzi, A.B., Freitas, S.: Delphi method to explore future scenario possibilities on technology and HCI. In: Human-Computer Interactions International, DUXU, pp. 644–653. Springer, Los Angeles (2015)
2. Renzi, A.B.: Experiência do ususário: a jornada de Designers nos processos de gestão de suas empresas de pequeno porte utilizando sistema fantasiado em ecossistema de interação cross-channel. These de doutorado. 239 p. Escola Superior de Desenho Industrial, Rio de Janeiro (2016)

3. Sande, A., Renzi, A.B., Schnaider, S.: Experience, usability and the sense of things. In: Design, User Experience, and Usability: Designing Pleasurable Experiences. HCI International 2017 Proceedings, Part II, Vancouver, BC, pp. 77–86 (2017)
4. de Villanueva, R.A.: Think-aloud protocol aril heuristic evaluation of non-immersive, desktop photo-realistic virtual environments 2004. Dissertation (Master of Science) – University of Otago, Dunedin – New Zealand (2004)
5. Teixeira, E., de Moraes, A.: Avaliação cooperativa da interface de sites hipermídias focados na "banda larga". In: Congresso Internacional de Ergonomia e Usabilidade, Design de Interfaces e Interação Humano-Computador, Rio de Janeiro, vol. 3 (2004)
6. Mucchielli, R.: O questionário na pesquisa psicossocial. Martins Fontes Publishing, São Paulo (2004)
7. Bowler, G.M.: Netnography: a method specifically designed to study cultures and communities online. Qual. Rep. **15**(5), 1270–1275 (2010). Accessed https://nsuworks.nova.edu/tqr/vol15/iss5/13
8. Resmini, A., Rosatti, L.: Pervasive Information Architecture – Designing Cross-Channel User Experiences. Morgan Kaufmann – Inprint of Elsevier, Burlington (2011)
9. Agner, L.: Ergodesign e arquitetura de Informação – trabalho com usuário. Ed. Quartet, Rio de Janeiro (2012)
10. Inostroza, R., Rusu, C., Roncaglioso, R., Rusu, V.: Usability heuristics for touchscreen-based mobile devices. In: 9th International Conference on Information Technology. IEEE Computer Society, Temuco (2013)
11. Usability.gov: Card Sorting. https://www.usability.gov/how-to-and-tools/methods/card-sorting.html
12. Padovani, S., Ribeiro, M.A.: Card sorting: adaptation of the technique for application to non-digital information system design. Revista Brasileira de Design da Informação/Braz. J. Inf. Des. **10**(3), 293–312 (2013)
13. Renzi, A.B., Freitas, S.: Aplicação de Think-aloud Protocol em teste de usabilidade na procura de livros em livrarias online: recomendações de leitores. In: 10º USIHC – Congresso Internacional de ergonomia e usabilidade. Conference Proceedings, Rio de Janeiro (2010)
14. Miranda, F., Moraes, A.: Avaliação da interface de um site de comércio eletrônico através da técnica avaliação cooperativa. In: Anais 2º USIHC, Rio de Janeiro (2003)

Towards Machine Translation of Chinese Complex Structures

Pingfang Yu, Jiali Du(✉), and Xinguang Li

Guangdong University of Foreign Studies, Guangzhou 510420, China
dujiali68@126.com

Abstract. The machine parsing and translation of Chinese complex structures need the involvement of syntactic and semantic knowledge. The application of computational linguistics in parsing and translation helps bilingual system construct the special knowledge database used to revise and post-edit the strings. It is concluded that with the help of advanced bilingual dictionaries and domain-specific corpora, system can effectively analyze the lexical information and thematic relationship, parse the syntactic structure and translate the original sentence to the target one.

Keywords: Machine translation · Computational linguistics · Natural language processing

1 Introduction

Machine translation offers an attractive prospect for computational linguistics. In 2013, 143 machine translation systems including translation tasks, run-time estimation, and unofficial metrics tasks were successfully evaluated both automatically and manually [1]. MT system focuses on the challenging problems faced by designers of systems and on explanation of hope of attaining human-level performance [2]. It seeks to dispense with the human translator, and the computerized aids seek to support the human translator and increase productivity by automating those tasks that can be automated reliably [3]. For example, the application of hardware bytecode translation is effective for the improvement of Java virtual machine [4], and systems based on edge weight instrumentation during the translation from executed code to native code can converge more quickly [5].

Statistical machine translation (SMT) system is efficient for natural language processing. Quality was the goal of a shared task at the workshop on SMT, the dissemination is limited by the lack of reliability of outputs [6]. The system can bring a significant increase in efficiency for the translation of morphologically rich and complex languages [7]. The statistics-based translation of long and complex sentences needs enormous computational time and space. The method to segment the long sentences into a few segments to analyze separately is proved to be helpful for the problem resolution [8]. With the help of standard phrase-based and alternative N-gram-based systems, both morpho-syntactical and statistical information can be effectively used to capture short- and long-distance word distortion dependencies [9]. The

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 582–593, 2020.
https://doi.org/10.1007/978-3-030-20040-4_52

alignments, defined as a mapping between the words in the source and in the target sentence, are useful to train the statistical models and to link the words in the source sentence to the words in the partial hypotheses [10].

Machine translation involves the interaction between linguistic and computational knowledge. On the basis of computational model of tutorial feedback and extensive linguistic analysis, linguistic ability is proved to be practical to enhance the naturalness of interactions [11]. The ontological semantic technology with an explicit understanding of what meaning is can be helpful for natural language understanding and information processing [12].

Semantic information is necessary for MT system. The improvement of efficiency of natural language processing needs the semantic involvement and the explicit planning of a semantic process from cognitive linguistic perspective at the memory level and the justification of the functionality [13]. For example, lexicalization by which people lexicalize new referring expressions in different situations involves the choice of appropriate vocabulary to transform the conceptual content of a referring expression into the corresponding text in natural language [14]. The nonlinguistic factors can affect the result of processing. For example, during the experiments of binding information in the mixed-effects models, old adults with higher working memory span rather than readers with average span are proved to be able to allocate extra time to resolve global ambiguities [15]. The parsing is sensitive to the conceptual representation of a plural constituent. With the existence of conceptual plurality, sentences with reciprocal verbs sometimes can bring the temporarily ambiguous effect, i.e. garden path effect [16].

This paper will discuss the machine parsing and translation of complex syntactic structures of Chinese garden path sentence from the perspective of computational linguistics, and show how the parsing and translation is affected by syntactic and semantic knowledge.

2 Machine Parsing and Translation Model

Translation Corresponding Tree (TCT) Structure is an example-based machine translation system [17]. It considers bilingual knowledge base as the basic annotation structure to facilitate the representation of examples which are usually annotated using a pair of analyzed structures. The construction of the TCT structures is involved in the automatically processing and generating a preliminary TCT representation structure for a translation example. The TCT editing program can be used to edit the annotation tree to meet the requirement of any changes of the representation structure. See Fig. 1.

Machine Translation system needs grammar-based analysis of original and target sentences. Having finished the establishment of Bilingual Texts, system begins the process of translation. With the help of Parser and Lexical Analyzer, TCT generation starts from the grammatical analysis. Correspondences Mapping, Correspondences Tree and Phrase Alignment are the result of syntactic analysis of Parser, which is efficient to find the correspondences between the surface sub-strings and the inter levels of the structure. Word Alignment is the result of involvement of lexical analysis of Lexical Parser and Bilingual Dictionary whose linguistic knowledge is helpful for segmenting and tagging the words with a part of speech. The system relies on the

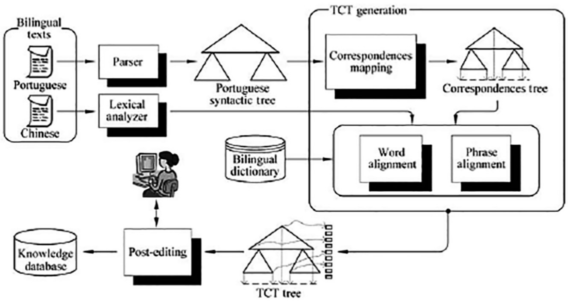


Fig. 1. TCT structure by Wong et al. [17]

grammatical information of the analyzed structure and a given bilingual dictionary to search for the corresponding translation sub-strings to establish the translation correspondences for the structural constituents. TCT tree is the comprehensive involvement of phrasal and lexical information. The processing above is basically automatic and the systematic result needs the manually post-editing process to obtain the final representation, which is considered the basic element of the Knowledge Database.

TCT system shows the importance of structural analysis. Garden path sentence has a complex structure which is difficult for system to parse. During the translation from Chinese to English, the syntactic analysis and parsing of Chinese garden path sentence has to be efficient and effective. The structural analysis of the complex syntactic structure will be shown below.

3 Structural Analysis of the Chinese Complex Structure

The analysis of the complex syntactic structure needs the support of skills of computational linguistics and computer techniques. Context-Free Grammar is an effective programming presentation for the processing of complex structure. Well-formed substring table is another useful application of computational linguistics for sentence parsing. Recursive Transition Network and CYK algorithm originating from the computer science are proved to be helpful for sentence processing in MT system.

3.1 The Application of CFG in Parsing

A context-free grammar belongs to the domain of formal language theory. In the grammar, every production rule is the form of "V → w". "V" means the nonterminal symbol. The symbol of "w" mainly refers to the terminals and/or non-terminals ("w" can be empty). Regardless of the context of non-terminals, the right terminals can be used to replace the left non-terminals, and production rules of the grammar are considered to be effective and efficient. Please see the segmentation and parsing of Example 1.

The rules of CFG in Fig. 2 show that the sentence can be clearly segmented and parsed. If the rule of "IP → NP VP" and the structure of [[dongshi zhang]NP+[chuchai shanghai gongsi]VP] IP are accepted by the system firstly, the effect of garden path comes into being. Please see Fig. 3.

Example 1: *Dongshizhang chuchai Shanghai gongsi de kuaiji tuo ta gei Fudan dushu de erzi shao tai diannao.*
G={Vn, Vt, S, P}
Vn={S, WHP, IP, PP, NP, VP, V, Quan, N, Prep, Pron, De}
Vt={*dongshizhang, chuchai, Shanghai, gongsi, de, kuaiji, tuo, ta, gei, Fudan, dushu, de, erzi, shao, tai, diannao*}
S=S
P:
a. S→WHP IP
b. WHP→(WH) IP
c. PP→(Prep)NP
d. IP→NP VP
e. VP→V PP
f. VP→PP V
g. VP→V NP
h. VP→VP PP VP
i. NP→Quan N
j. PP→ Prep NP
k. NP→Pron
l. NP→N N
m. NP→N
n. NP→N De N
o. NP→VP De N
p. Pron→{*ta*}
q. N→{*dongshizhang, Shanghai, gongsi, kuaiji,Fudan, erzi, diannao*}
r. V→{*chuchai, tuo, dushu, shao*}
s. Prep→{*gei*}
t. De→{*de*}
u. Quan→{*tai*}

Fig. 2. Rules of CFG of Example 1

Dongshizhang chuchai Shanghai gongsi…	
N *chuchai Shanghai gongsi…*	q
NP *chuchai Shanghai gongsi…*	m
NP V *Shanghai gongsi…*	r
NP V N *gongsi…*	q
NP V N N …	q
NP V NP …	l
NP V PP …	c
NP VP …	e
IP …	d
?	
BREAKDOWN AND BACKTRACKING	

Fig. 3. Breakdown and backtracking of Example 1

Having experienced the processing breakdown, system returns to the crossroad where the rule of "WHP (IP) → NP VP" and the structure of [[dongshi zhang]NP +[chuchai shanghai]VP] WHP are parsed. Then we can obtain the successful result.

Dongshizhang chuchai Shanghai gongsi…	
N *chuchai Shanghai gongsi…*	q
NP *chuchai Shanghai gongsi…*	m
NP V NP *gongsi…*	m
NP V PP *gongsi…*	c
NP VP *gongsi…*	e
IP *gongsi…*	d
WHP *gongsi…*	b
WHP N *de…*	q
WHP N De *kuaiji…*	t
WHP N De N *tuo…*	q
WHP NP *tuo…*	n
WHP NP V *ta…*	r
WHP NP V Pron *gei…*	p
WHP NP V NP *gei…*	k
WHP NP VP *gei..*	g
WHP NP VP Prep *Fudan…*	s
WHP NP VP Prep N *dushu…*	q
WHP NP VP Prep NP *dushu…*	m
WHP NP VP Prep PP *dushu…*	c
WHP NP VP Prep PP V *de…*	r
WHP NP VP Prep VP *de…*	f
WHP NP VP Prep VP De *erzi…*	t
WHP NP VP Prep VP De N *shao…*	q
WHP NP VP Prep NP *shao…*	o
WHP NP VP PP *shao…*	j
WHP NP VP PP V *tai…*	r
WHP NP VP PP V Quan *diannao*	u
WHP NP VP PP V Quan N	q
WHP NP VP PP V NP	i
WHP NP VP PP VP	g
WHP NP VP	h
WHP IP	d
S	a
SUCCESS	

Fig. 4. Successful processing of CFG for Example 1

The parsing in Fig. 4 can be shown in the tree diagram in which the rule of "S → WHP IP" becomes the top rule. Thus the complex syntactic structure is parsed successfully. Please see the Fig. 5.

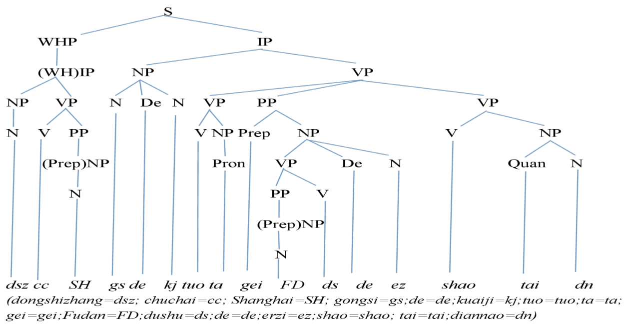


Fig. 5. Tree diagram of Chinese GP sentence

Besides CFG and tree diagram, the Recursive Transition Network is another formal language helpful to analyze the complex structure of Chinese garden path sentence. Please see the application of RTN in the processing.

3.2 The Application of RTN in Parsing

A recursive transition network is a graph-related theory. It can present the rules of a context-free grammar and show the full construction of the formed sub-string table. If the processing is involved in the backtracking and breakdown, the table will be the non-well-formed. However, if the processing is perfect, the well-formed sub-string table takes effect. According to the CFG-based analysis above, we can obtain the particular RTN used to describe the procedure of sentence parsing. Please see Fig. 6.

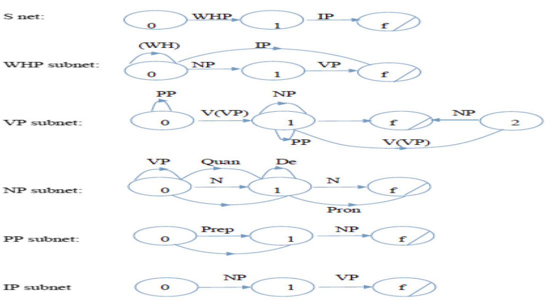


Fig. 6. The RTN of Chinese GP sentence

In Fig. 6, RTN comprises one net and five subnets. Nearly all the elements in CFG are included in the network. By means of the RTN, the steps of processing breakdown and backtracking can be described step by step.

In the initial processing in Fig. 7, system suffers the breakdown now that the right elements, i.e. “de kuaiji tuo ta gei fudan dushu de erzi shao tai diannao”, fail to be included in the parsing if the structure of “[[dongshizhang]NP+[chuchai shanhai gongsi]VP] IP” is chosen as the prototype. The result is as same as the one analyzed in CFG.

Dongshizhang chuchai Shanghai gongsi...
<S/0, *dongshizhang chuchai Shanghai gongsi...*, >
<WHP/0,*dongshizhang chuchai Shanghai gongsi...*, S/1: >
<NP/0,*dongshizhang chuchai Shanghai gongsi...*,WHP/1: S/1: >
<NP/1,*dongshizhang chuchai Shanghai gongsi...*, WHP/1: S/1: >
<NP/f, *chuchai Shanghai gongsi...*, WHP/1: S/1: >
<VP/0, *chuchai Shanghai gongsi...*, WHP/f: S/1: >
<VP/1, *Shanghai gongsi...*, WHP/f: S/1: >
<PP/0, *Shanghai gongsi...*, VP/f: WHP/f: S/1: >
<PP/1, *Shanghai gongsi...*, VP/f: WHP/f: S/1: >
<NP/0, *Shanghai gongsi...*, PP/f: VP/f: WHP/f: S/1: >
<NP/1, *gongsi...*, PP/f: VP/f: WHP/f: S/1: >
<NP/f, *de...*, PP/f: VP/f: WHP/f: S/1: >
<PP/f, *de...*, VP/f: WHP/f: S/1: >
<VP/f, *de...*, WHP/f: S/1: >
<WHP/f, *de...*, S/1: >
<IP/0, *de...*, S/f: >
?
BREAKDOWN AND BACKTRACKING

Fig. 7. The breakdown and backtracking of RTN

When the processing shifts from the prototypical structure to the low frequency one, the garden path phenomenon appears. Please see the correct processing in which all the elements are parsed successfully in Fig. 8.

In the procedures shown below in Fig. 8, we find the fact that the parsing is well-formed, and the structure of “[[dongshizhang]NP+[chuchai shanghai]VP] WHP” is proved to be the ultimate processing which can lead to the success.

<NP/1, *gongsi*..., PP/f: VP/f: WHP/f: S/1: >
<NP/f, *gongsi*..., PP/f: VP/f: WHP/f: S/1: >
<PP/f, *gongsi*..., VP/f: WHP/f: S/1: >
<VP/f, *gongsi*..., WHP/f: S/1: >
<WHP/f, *gongsi*..., S/1: >
<IP/0, *gongsi*..., S/f: >
<NP/0, *gongsi*..., IP/1: S/f: >
<NP/1, *de*..., IP/1: S/f: >
<NP/1, *kuaiji*..., IP/1: S/f: >
<NP/f, *tuo*..., IP/1: S/f: >
<VP/0, *tuo*..., IP/f: S/f: >
<VP/1, *ta*..., IP/f: S/f: >
<NP/0, *ta*..., VP/f: IP/f: S/f: >
<NP/1, *ta*..., VP/f: IP/f: S/f: >
<NP/f, *gei*..., VP/f: IP/f: S/f: >
<PP/0, *gei*..., VP/f: IP/f: S/f: >
<PP/1, *Fudan*..., VP/f: IP/f: S/f: >
<NP/0, *Fudan*..., PP/f: VP/f: IP/f: S/f: >
<VP/0, *Fudan*..., NP/1: PP/f: VP/f: IP/f: S/f: >
<PP/0, *Fudan*..., VP/1: NP/1: PP/f: VP/f: IP/f: S/f: >
<PP/1, *Fudan*..., VP/1: NP/1: PP/f: VP/f: IP/f: S/f: >
<NP/0, *Fudan*..., PP/f: VP/1: NP/1: PP/f: VP/f: IP/f: S/f: >
<NP/1, *Fudan*..., PP/f: VP/1: NP/1: PP/f: VP/f: IP/f: S/f: >
<NP/f, *dushu*..., PP/f: VP/1: NP/1: PP/f: VP/f: IP/f: S/f: >
<PP/f, *dushu*..., VP/1: NP/1: PP/f: VP/f: IP/f: S/f: >
<VP/1, *dushu*..., NP/1: PP/f: VP/f: IP/f: S/f: >
<VP/f, *de*..., NP/1: PP/f: VP/f: IP/f: S/f: >
<NP/2, *erzi*..., PP/f: VP/f: IP/f: S/f: >
<NP/f, *shao tai diannao*, PP/f: VP/f: IP/f: S/f: >
<PP/f, *shao tai diannao*, VP/f: IP/f: S/f: >
<VP/1, *shao tai diannao*, VP/f: IP/f: S/f: >
<NP/0, *tai diannao*, VP/2: VP/f: IP/f: S/f: >
<NP/1, *diannao*, VP/2: VP/f: IP/f: S/f: >
<NP/f, , VP/2: VP/f: IP/f: S/f: >
<VP/2, ,VP/f: IP/f: S/f: >
<VP/f, , IP/f: S/f: >
<IP/f, , S/f: >
<S/f, , >
<, , >
SUCCESS

Fig. 8. The successful processing of RTN

3.3 The Application of CYK in Parsing

CYK algorithm is a parsing algorithm used for CFG. CYK is the abbreviation of first letters of the inventors' family names, namely J. Cocke– D. Younger–T. Kasami. Bottom-up parsing and dynamic programming are the basic strategies employed by CYK algorithm which has high efficiency in the parsing of garden path sentence. According to the algorithm, we can construct a matrix describing the processing. Please see the Table 1.

Table 1. CYK matrix for breakdown of Example 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	{N}	{}	{}	{S}	{}	{}	{}	{}	{}	{}	{}	{?}	{}	{}	{}	{?}
1		{V}	{}	{VP}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
2			{N}	{NP}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
3				{N}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
4					{D}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
5						{N}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
6							{V}	{VP}	{}	{}	{}	{}	{}	{}	{}	{VP}
7								{P}	{}	{}	{}	{}	{}	{}	{}	{}
8									{P}	{}	{}	{}	{PP}	{}	{}	{}
9										{N}	{VP}	{}	{NP}	{}	{}	{}
10											{V}	{}	{}	{}	{}	{}
11												{D}	{}	{}	{}	{}
12													{N}	{}	{}	{}
13														{V}	{}	{NP}
14															{Q}	{NP}
15																{N}

In Table 1, the symbol “S” can be obtained at the chart(0, 4), which means not all the elements are included in the processing. Since no rule of CFG is available for all the strings’ processing, the created structure of “[[dongshizhang]NP+[chuchai shanghai gongsi]VP]” is proved to be a failure, and the sub-string table is not well-formed. Please see the presentation of processing breakdown in Fig. 9.

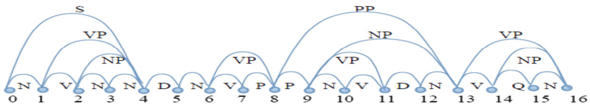


Fig. 9. Non-well-formed sub-string table of Example 1

The construction of Fig. 9 originates from the steps of Table 1 in which no more CFG rules used to replace the terminals of “S, D, N, VP, PP, VP”, which means the failure of processing. However, if the symbol “WHP” can be obtained at the chart (0, 3) and the ultimate symbol “S” is accepted at the final chart (0, 16), system succeeds in parsing the complex structure. Please see the matrix of successful parsing in Table 2.

Table 2. CYK matrix for parsing of Example 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	{N}	{}	{W}	{}	{}	{}	{}	{}	{}	{}	{}	{?}	{}	{}	{}	{S}
1		{V}	{VP}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
2			{N}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
3				{N}	{}	{NP}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{IP}
4					{D}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
5						{N}	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
6							{V}	{VP}	{}	{}	{}	{}	{}	{}	{}	{VP}
7								{P}	{}	{}	{}	{}	{}	{}	{}	{}
8									{P}	{}	{}	{}	{PP}	{}	{}	{}
9										{N}	{VP}	{}	{NP}	{}	{}	{}
10											{V}	{}	{}	{}	{}	{}
11												{D}	{}	{}	{}	{}
12													{N}	{}	{}	{}
13														{V}	{}	{VP}
14															{Q}	{NP}
15																{N}

According to the procedures in Table 2, we can create the well-formed sub-string table in Fig. 10 in which each element in Example 1 will be included.

According to Fig. 10, the procedure of how the complex structure is parsed is shown step by step in Fig. 11. The created CYK algorithm is long and complex, and we select the last step in which the string of “j = 16, chart(15, 16) =: {diannao}” is parsed. Please see the steps below:

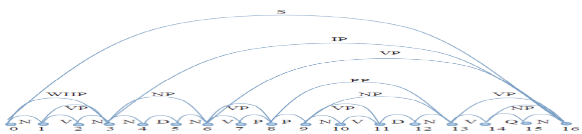


Fig. 10. Well-formed sub-string table of Example 1

```
j=16, i=2, k=8,chart(2, 8) ∪ chart(8, 16):={} ∪ {}={}
j=16, i=2, k=9,chart(2, 9) ∪ chart(9, 16):={} ∪ {}={}
j=16, i=2, k=10,chart(2, 10) ∪ chart(10,16 ):={} ∪ {}={}
j=16, i=2, k=11,chart(2, 11) ∪ chart(11, 16):={} ∪ {}={}
j=16, i=2, k=12,chart(2, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=2, k=13,chart(2, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=2, k=14,chart(2,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=2, k=15,chart(2, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=3, k=4,chart(3, 4) ∪ chart(4, 16):={N} ∪ {}={}
j=16, i=3, k=5,chart(3, 5) ∪ chart(5,16 ):={} ∪ {}={}
j=16, i=3, k=6,chart(3, 6) ∪ chart(6, 16):={NP} ∪ {VP}={IP}
j=16, i=3, k=7,chart(3, 7) ∪ chart(7, 16):={} ∪ {}={}
j=16, i=3, k=8,chart(3, 8) ∪ chart(8, 16):={} ∪ {}={}
j=16, i=3, k=9,chart(3,9 ) ∪ chart(9, 16):={} ∪ {}={}
j=16, i=3, k=10,chart(3, 10) ∪ chart(10, 16):={} ∪ {}={}
j=16, i=3, k=11,chart(3, 11) ∪ chart(11, 16):={} ∪ {}={}
j=16, i=3, k=12,chart(3, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=3, k=13,chart(3, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=3, k=14,chart(3,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=3, k=15,chart(3, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=4, k=5,chart(4, 5) ∪ chart(5, 16):={D} ∪ {}={}
j=16, i=4, k=6,chart(4, 6) ∪ chart(6,16 ):={} ∪ {VP}={}
j=16, i=4, k=7,chart(4,7 ) ∪ chart(7, 16):={} ∪ {}={}
j=16, i=4, k=8,chart(4, 8) ∪ chart(8, 16):={} ∪ {}={}
j=16, i=4, k=9,chart(4, 9) ∪ chart(9, 16):={} ∪ {}={}
j=16, i=4, k=10,chart(4, 10) ∪ chart(10, 16):={} ∪ {}={}
j=16, i=4, k=11,chart(4, 11) ∪ chart(11, 16):={} ∪ {}={}
j=16, i=4, k=12,chart(4, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=4, k=13,chart(4, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=4, k=14,chart(4,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=4, k=15,chart(4, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=5, k=6,chart(5, 6) ∪ chart(6, 16):={N} ∪ {VP}={}
j=16, i=5, k=7,chart(5, 7) ∪ chart(7, 16):={} ∪ {}={}
j=16, i=5, k=8,chart(5, 8) ∪ chart(8, 16):={} ∪ {}={}
j=16, i=5, k=9,chart(5, 9) ∪ chart(9, 16):={} ∪ {}={}
j=16, i=5, k=10,chart(5,10 ) ∪ chart(10, 16):={} ∪ {}={}
j=16, i=5, k=11,chart(5, 11) ∪ chart(11, 16):={} ∪ {}={}
j=16, i=5, k=12,chart(5, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=5, k=13,chart(5, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=5, k=14,chart(5,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=5, k=15,chart(5, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=6, k=7,chart(6, 7) ∪ chart(7, 16):={V} ∪ {}={}
j=16, i=6, k=8,chart(6, 8) ∪ chart(8,16 ):={VP} ∪ {VP}={VP}
j=16, i=6, k=9,chart(6, 9) ∪ chart(9, 16):={} ∪ {}={}
j=16, i=1, k=2,chart(1, 2) ∪ chart(2, 16):={V} ∪ {}={}
j=16, i=1, k=3,chart(1,3 ) ∪ chart(3, 16):={VP} ∪ {IP}={}
j=16, i=1, k=4, chart(1, 4) ∪ chart(4, 16):={} ∪ {}={}
j=16, i=1, k=5,chart(1, 5) ∪ chart(5, 16):={} ∪ {}={}
j=16, i=1, k=6,chart(1, 6) ∪ chart(6, 16):={} ∪ {VP}={}
j=16, i=1, k=7,chart(1, 7) ∪ chart(7,16 ):={} ∪ {}={}
j=16, i=1, k=8,chart(1, 8) ∪ chart(8, 16):={} ∪ {}={}
j=16, i=1, k=9,chart(1,9 ) ∪ chart(9, 16):={} ∪ {}={}
j=16, i=1, k=10,chart(1, 10) ∪ chart(10, 16):={} ∪ {}={}
j=16, i=1, k=11,chart(1, 11) ∪ chart(11, 16):={} ∪ {}={}
j=16, i=1, k=12,chart(1, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=1, k=13,chart(1, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=1, k=14,chart(1,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=1, k=15,chart(1, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=2, k=3,chart(2,3 ) ∪ chart(3, 16):={N} ∪ {IP}={}
j=16, i=2, k=4,chart(2,4 ) ∪ chart(4, 16):={} ∪ {}={}
j=16, i=2, k=5,chart(2,5 ) ∪ chart(5, 16):={} ∪ {}={}
j=16, i=2, k=6,chart(2, 6) ∪ chart(6, 16):={} ∪ {VP}={}
j=16, i=2, k=7,chart(2, 7) ∪ chart(7, 16):={} ∪ {}={}
j=16, i=6, k=10,chart(6, 10) ∪ chart(10, 16):={} ∪ {}={}
j=16, i=6, k=11,chart(6, 11) ∪ chart(11,16 ):={} ∪ {}={}
j=16, i=6, k=12,chart(6, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=6, k=13,chart(6, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=6, k=14,chart(6,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=6, k=15,chart(6, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=7, k=8,chart(7, 8) ∪ chart(8, 16):={P} ∪ {}={}
j=16, i=7, k=9,chart(7,9 ) ∪ chart(9, 16):={} ∪ {}={}
j=16, i=7, k=10,chart(7, 10) ∪ chart(10, 16):={} ∪ {}={}
j=16, i=7, k=11,chart(7, 11) ∪ chart(11, 16):={} ∪ {}={}
j=16, i=7, k=12,chart(7, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=7, k=13,chart(7, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=7, k=14,chart(7,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=7, k=15,chart(7, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=8, k=9,chart(8,9 ) ∪ chart(9, 16):={P} ∪ {}={}
j=16, i=8, k=10,chart(8, 10) ∪ chart(10,16 ):={} ∪ {}={}
j=16, i=8, k=11,chart(8, 11) ∪ chart(11, 16):={} ∪ {}={}
j=16, i=8, k=12,chart(8, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=8, k=13,chart(8, 13) ∪ chart(13,16 ):={PP} ∪ {VP}={VP}
j=16, i=8, k=14,chart(8,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=8, k=15,chart(8, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=9, k=10,chart(9, 10) ∪ chart(10,16 ):={N} ∪ {}={}
j=16, i=9, k=11,chart(9,11 ) ∪ chart(11, 16):={VP} ∪ {}={}
j=16, i=9, k=12,chart(9, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=9, k=13,chart(9, 13) ∪ chart(13,16 ):={NP} ∪ {VP}={}
j=16, i=9, k=14,chart(9,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=9, k=15,chart(9, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=10, k=11,chart(10, 11) ∪ chart(11, 16):={V} ∪ {}={}
j=16, i=10, k=12,chart(10, 12) ∪ chart(12, 16):={} ∪ {}={}
j=16, i=10, k=13,chart(10, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=10, k=14,chart(10,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=10, k=15,chart(10, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=11, k=12,chart(11, 12) ∪ chart(12, 16):={D} ∪ {}={}
j=16, i=11, k=13,chart(11, 13) ∪ chart(13,16 ):={} ∪ {VP}={}
j=16, i=11, k=14,chart(11,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=11, k=15,chart(11, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=12, k=13,chart(12, 13) ∪ chart(13,16 ):={N} ∪ {VP}={}
j=16, i=12, k=14,chart(12,14) ∪ chart(14, 16):={} ∪ {NP}={}
j=16, i=12, k=15,chart(12, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=13, k=14,chart(13,14) ∪ chart(14, 16):={V} ∪ {NP}={VP}
j=16, i=13, k=15,chart(13, 15) ∪ chart(15, 16):={} ∪ {N}={}
j=16, i=14, k=15,chart(14, 15) ∪ chart(15, 16):={Q} ∪ {N}={NP}
SUCCESS
```

Fig. 11. Processing algorithm of CYK for Example 1

The processing algorithm at the chart (0, 16) above reflects the fact that the structure is well-formed and all the elements are parsed successfully.

Based on the CFG, RTN and CYK algorithm, the structural analysis of Example 1 is proved to be effective and efficient. Generally speaking, the processing of Chinese garden path sentence comprises initial breakdown, in which the most likely interpretation is proved ungrammatically correct and turns out to be a dead end, and ultimate processing which replaces the original breakdown and is inconsistent with the path down which readers have been led.

The structural analysis paves the way for translation. Once the successful syntactic parsing completes, the efficiency of translation system is promoted considerably.

4 Machine Translation of Chinese Complex Structures

The bilingual lexical analysis in translation is necessary. In the mental lexicon or lexicon, system knows the lexical item and the distribution frame. The subcategorization frame, selection restriction and thematic relationship help system find the appropriate elements. With the help of bilingual dictionaries and corpora, system can analyze the lexical information and translate the original sentence to the target one [18–22]. In the online Chinese-English dictionary (http://www.iciba.com/) and English-English dictionary (http://www.ldoceonline.com/), we can obtain the necessary lexical information in Fig. 12.

dongshizhang: NP$_{[+Agent]}$
CEO: NP$_{[+Agent]}$
chuchai: V; [NP$_{[+Agent]}$PP$_{[+Location]}$;NP$_{[+Agent]}$(NP$_{[+Theme]}$)]
travel (on official business): V;
[NP$_{[+Agent]}$PP$_{[+Location]}$;NP$_{[+Agent]}$(NP$_{[+Theme]}$)]
Shanghai: NP$_{[+Location]}$
Shanghai: NP$_{[+Location]}$
gongsi: NP$_{[+Theme]}$
company: NP$_{[+Theme]}$
de: [NP NP]
of: [NP NP]
kuaiji: NP$_{[+Agent]}$
accountant: NP$_{[+Agent]}$
tuo: V; [NP$_{[+Agent]}$NP$_{[+Theme]}$NP$_{[+Theme]}$; NP$_{[+Agent]}$NP$_{[+Theme]}$PP$_{[+Location]}$]
entrust: V;
[NP$_{[+Agent]}$NP$_{[+Theme]}$NP$_{[+Theme]}$; NP$_{[+Agent]}$NP$_{[+Theme]}$PP$_{[+Location]}$]
ta: NP$_{[+Agent]}$
him: NP$_{[+Agent]}$
gei: V:
[NP$_{[+Agent]}$NP$_{[+Theme]}$NP$_{[+Theme]}$; NP$_{[+Agent]}$NP$_{[+Theme]}$PP$_{[+Goal]}$]
give: V;
[NP$_{[+Agent]}$NP$_{[+Theme]}$NP$_{[+Theme]}$; NP$_{[+Agent]}$NP$_{[+Theme]}$PP$_{[+Goal]}$]
Fudan:NP$_{[+Location]}$
Fudan: NP$_{[+Location]}$
dushu: V; [NP$_{[+Theme]}$;NP$_{[+Theme]}$PP$_{[+Location]}$]
study: V; [NP$_{[+Theme]}$;NP$_{[+Theme]}$PP$_{[+Location]}$]
de: [NP NP]
of: [NP NP]
erzi: NP$_{[+Theme]}$
son: NP$_{[+Theme]}$
shao: V; [NP$_{[+Agent]}$NP$_{[+Theme]}$NP$_{[+Theme]}$; NP$_{[+Agent]}$NP$_{[+Theme]}$PP$_{[+Goal]}$]
take (on one's way): V;
[NP$_{[+Agent]}$NP$_{[+Theme]}$NP$_{[+Theme]}$; NP$_{[+Agent]}$NP$_{[+Theme]}$PP$_{[+Goal]}$]
tai: Q
a: Det
diannao: NP$_{[+Theme]}$
computer: NP$_{[+Theme]}$

Fig. 12. The bilingual lexical analysis of Example 1

According to the lexical semantic analysis above, we can construct a basic framework of the translation in Fig. 13 by means of the corpora.

dongshizhang	(while)CEO
chuchai	travel(s) on official business
Shanghai	(to) Shanghai
gongsi de kuaiji	an account of company
tuo ta gei	(sb) be entrusted with the task of doing (sth)
fudan dushu de erzi	the son who has been studying in Fudan University
shao tai diannao	take a computer to (sb)

Fig. 13. The framework of translation for Example 1

Thus, after the word and phrase alignments, system can post-edit the frame from the structural and semantic perspectives. *"(while)CEO travel(s) on official business (to) Shanghai, (sb__NP[+Agent]) be entrusted with the task of (taking__NP[+Agent] NP[+Theme]PP[+Goal]] a computer to sb__NP[+Theme] the son who has been studying at Fudan University)(by) ___NP[+Agent] an accountant of company".*

According to the knowledge database with cognitive and linguistic information, system can obtain the relatively perfect translation shown below. *"While CEO traveled on official business to Shanghai, he was entrusted by an accountant of company with the task of taking a computer to her son who had been studying at Fudan University.*"

The result of translation above shows the importance of pre-analysis of syntactic and semantic information. Knowledge database is helpful for the post-editing and post-revising, by which the relatively perfect result can be achieved.

5 Conclusion

Syntactic analysis and semantic construction are the basic strategies of machine translation. During the natural language processing of Chinese complex structures, the methods of computational linguistics are effective and efficient. Lexical subcategorization frame, selection restriction and thematic relationship help system complete the translation with a high quality.

Acknowledgments. This article is sponsored by "Thirteen-Five Major Project" for National Science and Technology Terminology Validation Committee (ZD2019001), China Social Science Fund (2018FYY006), China Natural Science Fund (61877013), and Guangdong Science and Technology Department Project (2018A070712032).

References

1. Bojar, O., et al.: Findings of the 2013 workshop on statistical machine translation. In: Proceedings of the Eighth Workshop on Statistical Machine Translation, pp. 1–44 (2013)
2. Isabelle, P., Foster, G.: Machine translation: overview. In: Encyclopedia of Language & Linguistics, pp. 404–422 (2006)
3. Macklovitch, E.: Machine-aided translation: methods. In: Encyclopedia of Language & Linguistics, pp. 394–404 (2006)
4. Oi, H.: Local variable access behavior of a hardware-translation based Java virtual machine. J. Syst. Softw. **81**, 2059–2068 (2008)

5. Ung, D., Cifuentes, C.: Dynamic binary translation using run-time feedbacks. Sci. Comput. Program. **60**, 189–204 (2006)
6. Wisniewski, G., Singh, A.K., Yvon, F.: Quality estimation for machine translation: some lessons learned. Mach. Transl. **27**, 213–238 (2013)
7. Sadat, F., Mohamed, E.: Improved Arabic-French machine translation through preprocessing schemes and language analysis. In: Advances in Artificial Intelligence, pp. 308–314. Springer, Heidelberg (2013)
8. Kim, Y.S., Oh, Y.J.: Intra-sentence segmentation based on support vector machines in English-Korean machine translation systems. Expert Syst. Appl. **34**, 2673–2682 (2008)
9. Khalilov, M., Fonollosa, J.A.R.: Syntax-based reordering for statistical machine translation. Comput. Speech Lang. **25**, 761–788 (2011)
10. Rodríguez, L., García-Varea, I., Gámez, J.A.: On the application of different evolutionary algorithms to the alignment problem in statistical machine translation. Neurocomputing **71**, 755–765 (2008)
11. Porayska-Pomsta, K., Mellish, C.: Modelling human tutors' feedback to inform natural language interfaces for learning. Int. J. Hum.-Comput. Stud. **71**, 703–724 (2013)
12. Raskin, V., Taylor, J.M., Hempelmann, C.F.: Meaning-and ontology-based technologies for high-precision language an information-processing computational systems. Adv. Eng. Inform. **27**, 4–12 (2013)
13. Duan, Y., et al.: Modeling value evaluation of semantics aided secondary language acquisition as model driven knowledge management. In: Computer and Information Science, pp. 267–278 (2013)
14. Hervás, R., Francisco, V., Gervás, P.: Assessing the influence of personal preferences on the choice of vocabulary for natural language generation. Inf. Process. Manag. **49**, 817–832 (2013)
15. Payne, B.R., et al.: Aging and individual differences in binding during sentence understanding: evidence from temporary and global syntactic attachment ambiguities. Cognition **130**, 157–173 (2014)
16. Patson, N.D., Ferreira, F.: Conceptual plural information is used to guide early parsing decisions: evidence from garden-path sentences with reciprocal verbs. J. Mem. Lang. **60**, 464–486 (2009)
17. Wong, F., Dong, M., Hu, D.: Machine translation based on translation corresponding tree structure. Tsinghua Sci. Technol. **11**, 25–31 (2006)
18. Goto, I., Utiyama, M., Sumita, E., et al.: Preordering using a target-language parser via cross-language syntactic projection for statistical machine translation. ACM Trans. Asian Low-Resour. Lang. Inf. Process. **14**(3), 1–23 (2015)
19. Klubička, F., Toral, A., Sánchez-Cartagena, V.M.: Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian. Mach. Transl. **1**, 1–21 (2018)
20. Domingo, M., Álvaro, P., Casacuberta, F.: Segment-based interactive-predictive machine translation. Mach. Transl. **3**, 1–23 (2018)
21. Du, J.L., Yu, P.F., Zong, C.Q.: Towards computing technologies on machine parsing of English and Chinese garden path sentences. In: FTC 2018, pp. 806–827. Springer, Switzerland (2018)
22. Du, J.L., Alexandris, C., et al.: Controlling interaction in multilingual conversation revisited: a perspective for services and interviews in mandarin Chinese. In: HCI 2017. LNCS, pp. 573–583. Springer (2017)

Challenge of Implanting Educational Management Systems in Brazilian Schools

Eline H. C. Guimaraes(✉), Angelica F. S. Dias(✉), and Juliana B. S. França(✉)

PPGI – UFRJ – Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
eline@el4us.com, angelica@nce.ufrj.br, julibsf@gmail.com

Abstract. The development, implementation, and contracts involving interactive systems in Brazil have direct influence over educators and managers regarding and the digital technology potential. And these factors facilitate to choose, among many possible tools, the ones that have elements to better support learning in school environments. EMIS (Educational Management Integrated System) is a fundamental piece to consolidate quality education. Any ERP is usually capable of integrating diverse data in one system: finance, sales, accounting, stock, debts and income balance, human resources, production processes and logistic activities. However, education related systems surface factors related to classroom dynamics, to students and to teachers. These factors must be contemplated and evaluated as part of the whole context, part of a country with such cultural diversity and lack of resources, such as Brazil. This paper presents the unfolding of the project developed between IDB (Inter-American Development Bank) and FVG (Fundação Getúlio Vargas) to analyze EMIS systems in two public schools for basic education in Brazil.

Keywords: Human-computer interaction · Interoperability · Human factors · BID · Human-systems integration

1 Introduction: Brazil's Context

The Education Ministry in Brazil is the regulator above everything that concerns about educational national politics and highly motivated to care about every kind of systems and political articulation with twenty-seven states and Federal District. Every state or municipal has responsible to develop or contract their educational management systems, that we called EMIS (Educational Management Integrated System)[1]. States and municipal governments can develop even their properly educational curriculum,

[1] In Portuguese is SIGED.

Head of Management Consulting for Educational Business at Elfus. Education researcher focused digital platforms, product manager, project/products developer and Distance Education. Marketing Digital Expert to attract new customers and/or maintenance. Elfus founder. Master in NCE-UFRJ.

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 594–602, 2020.
https://doi.org/10.1007/978-3-030-20040-4_53

teachers KPIs and training and choose who will manage every school unity, but just need respect national Education Ministry guidelines. The municipal and state education systems and the federal system comprise a National Education System not constituted by law, but by responsibilities and attributions defined in the Federal Constitution - FC and in the Law of Guidelines and Bases of National Education - called LDB. However, this national system is characterized as not organic and functional, decentralized and asymmetrical.

The federal constitution determines that federal, state and municipal governments must have integrated learning systems, to ensure mutual collaboration and better education access by the population in Brazil. The federative union is responsible for the authorization, recognition, accreditation, supervision and evaluation of institutions of higher education. The states are responsible for middle school and emphasis in high school. The municipal governments are responsible for kindergarten and part of elementary school. The federal district has state's and municipal's responsibilities. The normative, supplementary and redistributive function of the Union to the States [4], Federal District and Municipalities, whose basic education is the responsibility of these federated entities, has contributed to the marked reduction of the inequality of access to education and the universalization of education in all segments. However, despite advances such as legal programs and mechanisms, the quality of education remains a challenge for the Brazilian government. Actions to improve basic education require access to educational data. According to LDB, Brazil must have to collect, analyze and disseminate information on education and to ensure the national process of evaluation of school performance in elementary, middle and high schools in collaboration with other federated entities.

The National Institute of Educational Studies and Research Anísio Teixeira - INEP [5] is the federal agency, linked to the Ministry of Education, whose mission is to promote the formulation of educational policies through the collection and statistical treatment of data. INEP is responsible for calculating the Basic Education Development Index - IDEB. Created in 2007, the IDEB is an indicator (KPI) that gathers two concepts relevant to the quality of education: (1) School census, which provides data on school approval, and (2) Prova Brasil, a census evaluation conducted every two years, to analyze the performance in Portuguese and Mathematics of elementary and high school students of the public school system [1]. The IDEB is published every two years and works as a tool for monitoring education goals, as set out in the National Education Plan (in portuguese: PNE).

INEP indicators contribute to monitoring the performance of basic education. However, the administration of education systems and their school institutions requires the management of other elements and educational processes. In Brazil, the characteristic of basic education education systems is that few have a culture of academic follow-up and, mainly, a management aimed at learning. Therefore, educational processes in all dimensions require improvements so that managers can increasingly focus on the issue of learning.

Understanding that, we can easily identify that the efficient and effective management and improvement of the educational system are intrinsically related to the improvement of Educational Management and Information Systems, covering all processes and resources. This makes research timely and justifies the study of EMISs available in the Brazilian market.

Specially when talking about education related to public schools. Precedent research pointed that among the twenty-seven Brazilian states and the federal district, five do not have (or didn't identified) the EMIS (Educational Management Integrated System) implemented. Only two states have their systems developed by private companies or consulting and thirteen systems are from either self-development initiatives or from a partnership with CAED – UFJF (Universidade Federal de Juiz de Fora). According to Marins [6], the adoption of these kinds of systems is a result from (1) disappointments with former incompatible system, (2) impossibility to integrate multiple systems or (3) the centralization of indicators.

2 EMIS Project

To understand every characteristic to have an effective EMIS, is very important that we can release about every kind of school management that we can find although kindergarten, elementary, middle of school and high school. In Brazil, K12 schools needs to report to local entities. They visit the schools to check if everything works as low. So, EMIS has an important contribution in these visits, because they can view reports and student information in real time (Fig. 1).

Fig. 1. Information about routine. Font: by the authors

Elementary and Middle School:
The follow-up of the students of the elementary and middle school segment, in systemic and process view, has great impact when compared to the routine of the High School segment. At this point in the child's life, the concern is how to report to parents about the food routine, issues related to hygiene, medications and development itself. In addition, issues such as disciplinary occurrences, classroom support, and the family-school relationship are critical. Therefore, the keywords in EMIS access interfaces need to reflect what, in fact, corresponds to the parent environment, as well as facilitating access to essential information. Here, we have an example of a functional interface for parents (Fig. 2):

Fig. 2. Mobile EMIS. Font: sistema AGENDAEDU

That EMIS called AGENDAEDU has functions and keywords suitable to parents on mobile phone. We can see the day-resume of students' activities, all academics activities, everything about events in school, school-parents communicate by chat, photos, health and notifications. We have a very large contrast If we compare with what we found in the interface of most EMISs in Brazilian public schools was that there was a great lack of detailing of keywords, even though there were the same functionalities. Clusters of data are confusing or pointless to the parents' environment. Here an example (Fig. 3):

Fig. 3. Interface with differents functionalities. Font: SEDU PE

3 Discovering EMIS in Brazil

These are all systems found in the Brazilian states. In general, they were verified from the following functionalities:

Management of Physical Infrastructure, Furniture and Equipment
The physical infrastructure management, furniture and equipment management process refers to the management of school buildings, inventory of assets, basic services available as well as physical capacity for educational attendance of buildings. It covers interventions and routine management of physical infrastructure and equipment, allowing the insertion of programmed supervision and maintenance routine. It also refers to the management of information in real time, with visions and access to the different managers of the educational system.

Management of School Institutions
This process concerns the management of school institutions. It covers the management of lesson plans, faculty and grouping of students in shifts, classes, segments and school grades. It addresses the fulfillment of the minimum space per student for the proper occupation in the classroom. Issues related to the transfer of resources directly to schools for small repairs and emergency assistance, as well as access to real-time information, according to access profile, are also addressed in this item.

Management of Human, Budgetary and Financial Resources
In the process of human, budgetary and financial management, the various aspects related to the administration of personnel, financial resources and budgetary resources for schools are analyzed, allowing a global view of the Department of Education. It seeks to evaluate the single registration in the personal data system, academic and professional trajectory, registration of training courses, assiduity, complaints, adjustments, assignments, functions and positions held. It also addresses issues such as the number of teaching positions or not, related to student demand, salary settlement based on professional registration, and assessment of the school environment, such as views and privileges associated with different managers of the education system.

Student and Learning Management
Student and learning management includes aspects related to student management through the implementation of the student's unique register. The register is composed of personal data, academic, socio-educational, school performance, promotion, repetition, dependencies, transfer and dynamics of student mobility between classes and school institutions. It includes attendance, performance, conduct and individual monitoring with online access for parents and guardians. In addition, the student management and learning system verifies the issuance of certificates and declarations via EMIS and allows the real time access to the managers of the educational system.

Management of Digital Technologies and Content for Teaching Learning and Training

The topic of digital content management refers to the management of digital tools for student learning and teacher training, based on new curricular requirements and the learning challenges of students, with guaranteed access, regardless of the location of the teacher. In addition, it incorporates tools to support and monitor students' learning and development, such as learning and assessment platforms, distance education and Learning Management System (LMS) and transversal tutorials (asynchronous and/or synchronous). It also considers the training of teachers and the encouragement of publication and sharing of practices based on digital resources, as well as access to information in real time.

Tools for Strategic Management

By tools for strategic management is meant the use of business intelligence tools for the delivery of updated and consolidated management indicators individually or in combination, displayed on control panels. The combination of data allows the visualization of information such as total cost versus academic performance, investment in assistance (transport, food, scholarship), student-teacher ratio, paid and unpaid leave, vacant and occupied positions and charges by number of students. In addition, the panels allow the interpretation of the performance of the different processes, programs and projects at each level of EMIS, such as real-time information for system managers according to the access profile.

Technological Infrastructure

Refers to aspects related to technological infrastructure and redundancy in the production environment. It covers information protection and restoration, automatic data auditing, and technical and operational documentation of the implemented systems. It also considers the arrangements for connectivity, identity protection, cyber security, interoperability with external systems, and EMIS's maintenance and support policy.

Governance and Institutional Arrangement

This process concerns the clear and strategic view of the information system by the highest authority of the educational system as a tool to support strategic planning and optimize institutional processes. It is related to the availability of human resources and financial resources to improve the EMIS in the short, medium and long term and to establish norms that regulate the processes and their improvement, as well as the dissemination of information linked to the results of the current educational system.

Below, we will see that Brazil is very fragmented, with different systems. In this research, we bring a survey of systems used only in the Brazilian states, understanding that the number can be around 10 times higher, since the municipalities of the Brazilian states also use different EMIS (Fig. 4):

Fig. 4. Brasilian states

ACRE – AC: Not found.
ALAGOAS – AL: SAGEAL – Developed by CAED UFJF.
AMAPÁ – AP: SIGEDUC – Developed by ESIG Software and Consulting is a company specialized in the construction and implementation of management systems. The company was born in 2011 from a UFRN spin-off where its founders were responsible for the development and implementation of GIS systems in the institution. Since then, the company has worked to contribute with other organizations in order to reach maturity in computerization in the areas of education, e-government and science and technology.
AMAZONAS – AM: SIGEAM - Developed by PRODAM - Data Processing of Amazonas S/A is a privately held mixed-capital company with state control of the state. It was created by Law No. 941 of July 10, 1970, and began operations in September 1972. Currently the company is administratively linked to the Secretariat of Planning and Economic Development (SEPLAN), in accordance with Law 2,783/2,003.
BAHIA – BA: SEC BA – Own development: SEC/CMO/Desenvolvimento
CEARÁ – CE: – SIGEESCOLA – Own development: SEDUC CE

DISTRITO FEDERAL – DF: SEI – Electronic Information System, developed by the Federal Regional Court of the 4th Region (TRF4) and provided free of charge to the public administration.
ESPÍRITO SANTO – ES: SEGES – Developed by CAED UFJF
GOIÁS – GO: SIGE – Own development.
MARANHÃO – MA: SIAEP Móvel
MATO GROSSO – MT: SIGEDUCA - Own development
MATO GROSSO DO SUL – MS: SGDE – Own development
MINAS GERAIS – MG: SIMADE – Developed by CAED UFJF
PARÁ – PA: Not found.
PARAÍBA – PB: Not found.
PARANÁ – PR: SERE – Sistema Educacional de Registro Escolar – Developed by CELEPAR, Companhia de Tecnologia da Informação e Comunicação do Estado do Paraná.
PERNANBUCO – PE: SIEPE – Developed by AUGE.
PIAUÍ – PI: Sistema Acadêmico – Own development.
RIO DE JANEIRO – RJ: Conexão Educação – Own development
RIO GRANDE DO NORTE – RN: SIGEduc – Developed by SEEC RN.
RIO GRANDE DO SUL – RS: Not found.
RONDÔNIA – RO: Not found.
RORAIMA– RR: Not found.
SANTA CATARINA – SC: SISGESC – Desenvolvimento próprio.
SÃO PAULO – SP: Not found.
SERGIPE – SE: SIGA e SIAE – Developed by CODIN – Coordenadoria de Informática do Estado.
TOCANTINS – TO: SGE – Own development: SEDU.

4 Conclusions

We can verify, with the above survey that of the 26 states and Federal District, 7 do not have or could not identify the EMIS. Only 2 are developed by outside companies or consulting. 11 of the systems are own development initiative and the rest in partnership with CAED - UFJF.

When carrying out the proposed survey, it was observed a great difference in the process of contracting the functionalities and understanding of the interface of an EMIS to Brazilian public schools, mainly regarding the process of choice, contracting and the use of modules such as food and school transportation, being optional in some segments and lasting in all public schooling. As a suggestion of improvements, it would be important to check and standardize the keywords of the systems, establishing a pattern of interactivity between schools and parents.

References

1. INEP. Conheça o INEP. http://portal.inep.gov.br/conheca-o-inep. Acessoem 02 out 2018
2. BRASIL: Congresso Nacional. Lei de Diretrizes e Bases da Educação Nacional – LDB no. 9.394/1996. Brasília (1996)
3. Zednik, C.: Are systems neuroscience explanations mechanistic? Preprint volume for Philosophy Science Association 24th Biennial Meeting, pp. 954–975. Philosophy of Science Association, Chicago (2014)
4. BRASIL: Constituição. Constituição da República Federativa do Brasil. Brasília, DF: Senado Federal (1988). http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm
5. BRASIL: Lei n. 11.494, de 20 de junho de 2007. Regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação – FUNDEB. Site oficial da República Federativa do Brasil. http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/lei/l11494.htm
6. Marins, F.A.S.: Sistemas ERP: características, custos e tendências. Prod. **15**(1), 102–113 (2005)

Author Index

I. L. Nunes (Ed.): AHFE 2019, AISC 959, pp. 603–605, 2020.
https://doi.org/10.1007/978-3-030-20040-4